Low-voltage, Low-power Feedforward-compensated Active Mixers with Improved Linearity

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
on gezag van de Rector Magnificus Prof. dr. K.C.A.M. Luyben
voorzitter van het College voor Promoties,
in het openbaar te verdedigen

op dinsdag 27 April 2010 te 15:30uur

doors

Su-Tam Lim
geboren te Maleisië
The work described in this thesis was supported by NXP (previously Philips) Semiconductors and the Delft University of Technology.

Keywords: distortion, linearity, distortion cancellation, feedforward cancellation, silicon radion frequency integrated circuit design, third order intercept point (IP3).


Copyright © 2010 by Su-Tam Lim

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the prior written permission of the copyright owner.
Table of Contents

1 Introduction ........................................................................................................... 1
  1.1 Motivation ....................................................................................................... 1
  1.2 Design Challenges .......................................................................................... 3
    1.2.1 Low Power and Low Voltage Operation ................................................. 3
    1.2.2 Multi-Standard Operation ...................................................................... 3
    1.2.3 Production Volume .................................................................................. 5
  1.3 Objectives of This Thesis ............................................................................... 6
  1.4 Organization of This Thesis ........................................................................... 6

2 Background Information ...................................................................................... 9
  2.1 Spurious-Free Dynamic Range .................................................................... 9
  2.2 Noise ............................................................................................................. 10
  2.3 Distortion......................................................................................................... 12
    2.3.1 Weakly Nonlinear Behaviour ................................................................ 12
      2.3.1.1 Distortion in the absence of interference ........................................ 14
      2.3.1.2 Distortion induced by interference .................................................. 15
    2.3.2 Strongly Nonlinear Behaviour ................................................................. 17
    2.3.3 Examples of Distortion in Circuits ........................................................... 18
      2.3.3.1 BJT amplifiers .............................................................................. 19
      2.3.3.2 CMOS amplifiers ...................................................................... 23
  2.4 Receiver Architecture ..................................................................................... 26
    2.4.1 Heterodyne Receiver .............................................................................. 26
    2.4.2 Homodyne Receiver .............................................................................. 27
    2.4.3 Image-Reject Downconverter ................................................................. 29
    2.4.4 Building and Comparing Receivers ......................................................... 30
  2.5 Mixer .............................................................................................................. 32
    2.5.1 Active Mixer ........................................................................................... 33
    2.5.2 Passive Mixer ......................................................................................... 35
  2.6 Summary ......................................................................................................... 37

3 Distortion in Active Mixers ............................................................................... 39
  3.1 Switching Stage .............................................................................................. 39
  3.2 Transconductor (input) Stage ........................................................................ 44
    3.2.1 Smoothing the Transconductance ............................................................ 46
    3.2.2 Out-of-Band Termination, Pre-distortion, and Feedforward Cancellation 51
  3.3 Output Stage .................................................................................................... 56
  3.4 A Reference Mixer Example ........................................................................... 57
  3.5 Summary ......................................................................................................... 62

4 Gilbert Multiplier with Feedforward-Compensated Input Stage ................. 65
  4.1 BJT Transconductor ....................................................................................... 65
    4.1.1 Analysis of the Feedforward-compensated Differential Pair ............... 66
    4.1.2 Simulation Results for the Feedforward-compensated Differential Pair 68
    4.1.3 Discussion .............................................................................................. 72
  4.2 CMOS Transconductor ................................................................................... 72
    4.2.1 Design of the CMOS Version ................................................................ 73
    4.2.2 Measured Results for the Mixer with Linearized Transconductor ...... 77
Contents

4.2.3 Discussion .......................... 81
4.3 Summary .............................. 81
5 Feedforward-Linearized Switching-Quad .......................... 83
  5.1 Linearized Mixer Concept .......................... 83
  5.2 BJT Mixer Design ............................ 84
     5.2.1 Design and Simulation ................. 86
     5.2.2 Measurement ........................... 89
     5.2.3 Discussion ............................ 93
  5.3 Conclusions and Summary .......................... 96
6 Feedforward-Linearized Switching-Quad in CMOS .......................... 99
  6.1 CMOS Implementation .......................... 99
     6.1.1 Design and Simulation ................. 100
     6.2 Discussion .............................. 117
     6.3 Summary .............................. 119
7 Conclusion and Recommendations .......................... 121
  7.1 Future work .............................. 123

List of Figures

Fig. 1-1 A simplified interference scenario .......................... 2
Fig. 1-2 (a) Software radio and (b) multi-standards receiver .......................... 4
Fig. 2.1 Effects of noise and distortion on an amplifier output spectrum (a) and spurious-free dynamic range (b) ........................................ 10
Fig. 2.2 Total equivalent input noise versus frequency for a BJT or CMOS transistor .......................... 11
Fig. 2.3 Equivalent input voltage and current noise sources in (a) BJT and (b) FET [10] .......................... 12
Fig. 2.4 Modeling of weakly nonlinear distortion using nonlinear transfer function approach .......................... 13
Fig. 2.5 Effects of harmonic distortion in an amplifier with bandpass response .......................... 15
Fig. 2.6 Effects of intermodulation (IM) distortion in the frequency domain .......................... 16
Fig. 2.7 (a) Schematic of a common-source transconductance stage and (b) its DC transfer characteristic along with ac input and output waveforms .......................... 18
Fig. 2.8 Schematics of (a) a common-emitter and (b) emitter-coupled pair .......................... 19
Fig. 2.9 Equivalent model of a common-emitter for distortion analysis .......................... 19
Fig. 2.10 Equivalent model for an emitter-coupled pair .......................... 22
Fig. 2.11 Magnitude of IM3 generated at the collector current of a common-emitter stage versus frequency .......................... 23
Fig. 2.12 Schematic of (a) a common-source and (b) a source-coupled pair .......................... 24
Fig. 2.13 Normalized drain-source current and the first three derivatives versus gate-source voltage from simulation (VDS=1.2V) .......................... 25
Fig. 2.14 Principle building blocks of a heterodyne receiver .......................... 26
Fig. 2.15 Principle building blocks of a homodyne front-end .......................... 27
Fig. 2.16 Degradation of SNR by in (a) zero-IF and (b) low-IF architectures .......................... 28
Fig. 2.17 Image-reject architectures: (a) Hartley and (b) Weaver architectures .......................... 29
Fig. 2.18 An receiver with cascaded stages .......................... 30
Fig. 2.19 Receiver block specification using two different mixers .......................... 31
Fig. 2.20 Conceptual drawing of a mixer in operation .......................... 33
Fig. 2.21 Schematic of a conventional Gilbert multiplier .......................... 34
Fig. 2.22 Simulation testbench for a conventional passive mixer .......................... 35
Fig. 2.23 Simulated conversion gain and IP3 for a passive mixer versus LO swing .......................... 36
Fig. 3.1 Schematic of a Gilbert modulator with an ideal transconductance input stage for analyzing the distortion of the switching quad .......................... 40
Fig. 3.2 Effects of LO swing on the linearity of switching quad for fixed emitter length with varying bias currents .......................... 41
Fig. 3.3 Effects of LO swing on linearity of switching quad for fixed bias current and varying rout .......................... 42
Fig. 3.4 Effects of LO swing on linearity of switching quad for fixed bias current for various emitter length .......................... 43
Fig. 3.5 Mixer with an ideal switching stage to analyze distortion of the input stage .......................... 44
Fig. 3.6 Linearity of a Gilbert modulator with ideal switches at various bias current .......................... 45
Fig. 3.7 An emitter-coupled pair with series local feedback using REE: (a) schematic and (b) DC transfer characteristic for three different combinations of IEEERE .......................... 46
Fig. 3.8 First (a) and third (b) derivatives of the ECP DC transfer characteristic for three different combinations of IEEERE .......................... 47
Fig. 3.9 A two-stage multi-tanh transconductor using asymmetrical emitter .......................... 48
Chapter 1

Introduction

Although the leading communication systems deployed at various times in history differ in transmission media, transmitters, receivers, and level of complexity, they all serve the same essential function of delivering information between parties in a reliable manner. In the late 1980s, wireless telephony and the internet were introduced to the general public marking a revolution in how information is broadcasted and received. With the ability to make or receive a telephone call anywhere anytime, wireless telephony is so popular that it has become a ubiquitous and indispensable tool of our everyday life. On the other hand, the internet began as a local-area network (LAN) for information sharing. Since then, it has evolved into a medium for worldwide broadcasting, information dissemination, and interaction between individuals and their computer terminals. Although both wireless telephony and the internet were initially limited to voice and data transmission between two terminals, respectively, recent trends suggest that the two systems are converging to create a more unified communication infrastructure.

Applications of wireless communication have extended beyond cordless telephones, pagers, and cellular phones to facilitate the transmission of data content. Wireless communication is now viewed as the "last hop" in the information chain that connects a user to an information source. Often, new wireless applications are spurred by the desire for "wire-free" products by the general public. These wireless products allow users to access information irrespective of time and location. Table 1-1 shows a range of different wireless data standards that have been proposed and promoted by various industry groups. These standards ensure that products from different vendors can be interchanged seamlessly, thereby allowing wireless links with ubiquitous connectivity. In the short range (<10m), wireless personal area network (WPAN) standards such as the IEEE-802.15.1 (commonly known as Bluetooth) [1] are now ubiquitously embedded into devices. The most common usage of Bluetooth is connectivity between a headset and a personal digital assistant or cellular phone. For medium coverage (30-300m), IEEE-802.11x standards (commonly known as WiFi) [2] dominate the wireless local-area network (WLAN) market, providing internet access at rates up to 54Mbps, with even higher speeds promised in the future. A wider range of coverage (up to 50km) for wireless data communication is promised by the worldwide interoperability for microwave access (WiMax) where it is being standardized (IEEE-802.16) [3] and touted as a "last mile" connection for broadband network access. Although all these standards have various ranges, they all support higher data rates compared to existing cellular telephony standards (i.e., mega-bits-per second, compare to kilobits-per-second in 2G/2.5G/3G cellular standards).

1.1 Motivation

A radio-frequency (RF) front-end resides within all wireless devices. This RF front-end is responsible for the transmission and reception of radio signals. In most cases, it is the bottleneck to achieving maximum performance and will determine the overall quality of service (QoS) of a wireless link. In mobile wireless devices, performance of the transmission path is measured by the power delivered by a wireless device to an intended target (e.g., base station or other wireless device in a link). In the receiving path, the performance is judged by its ability to process the desired radio signal in the presence of noise and interferers.

Historically, high performance discrete transistors made of gallium arsenide (GaAs) were used in the RF front-end. Advances in semiconductor technology have increased the performance of both
frequency exceeds 100GHz \[4\]. This is comparable to what could be achieved by GaAs transistors, suggesting the realization of RF front-end with silicon-based processes. In addition to lower fabrication cost, the silicon-based processes also allow a higher level of integration by integrating the RF front-end with baseband circuitry. This integration will yield further cost savings as the number of required external components are reduced, hence a reduced overall form factor. The cost reduction has also made wireless communication affordable in consumer markets worldwide.

Although an increase in wireless activity may be desirable from an economical standpoint, this also increases the probability of many radios operating close to one another, leading to congestion and interference in data transfer. How the QoS of a wireless link can be compromised is highlighted by the simplified example shown in Fig. 1-1. The transceiver of the notebook computer is used to communicate with a wireless router, providing internet access to the notebook computer. However, the radio of the notebook computer also picks up transmissions from other wireless routers that are in the vicinity, as depicted by interfering routers A and B in this simplified scenario. Consequently, this decreases the signal-to-interferer ratio, resulting in a higher bit-error rate (BER) and lower data throughput. How signals from the two interfering wireless routers reduce the BER will be further described in Chapter 2.

### 1.2 Design Challenges

Ever increasing wireless activities deal with crowded radio spectrum and the resulting interference. As a result, radios are designed with a wider dynamic range - and in particular, higher linearity which is less susceptible to such interferences. Radio designs that meet this new challenge must strike a balance between conflicting requirements (e.g., power dissipation and linearity).

#### 1.2.1 Low Power and Low Voltage Operation

With wireless devices becoming ubiquitous and embraced by the masses, criteria other than an acceptable QoS will begin to influence consumer choice. Portable wireless devices are increasingly viewed as part of the wardrobe that projects a certain fashion statement and there is also a preference from consumers for portable wireless devices that are lighter, smaller, and have longer battery life. Since a major contributor to the overall weight and size of a portable device is the battery, it follows that reducing the number and/or size of the batteries could increase its consumer appeal. This reduction can be achieved by reducing the supply voltage and power dissipation of the circuit (e.g., using new circuit topologies) where lower capacity and/or smaller batteries can be used.

Wireless consumer devices with longer battery life are preferable as they will have longer standby times and operational times between charges. This battery life line can be prolonged by using a battery with higher capacity or reducing the amount of current drain from the battery. The former is generally avoided as a battery with higher capacity will lead to higher cost, more weight, and larger size. Reduction of current drain is often preferred for mobile devices, and this can be achieved at different levels in the design hierarchy. At a higher level, demand on current can be reduced via the proper selection of architectures and modulation schemes that have higher power efficiency. Power dissipation can also be reduced by having portable devices entering a “sleep” mode, where part or most of the circuits are shutdown and do not draw any power. At lower hierarchical levels, the demand on the battery can be relaxed by reducing the number of active circuit components that require a constant bias current, and choosing circuit topologies that allow the re-use of bias current.

#### 1.2.2 Multi-Standard Operation

Unlike their predecessors, portable wireless devices of today provide connectivity for various standards and functions. For example, the latest model of cellular handset is more than just a wireless telephone, it is also a music player, a global positioning system (GPS) device, a camera, a mobile television, and an e-mail client. Bundling of applications is motivated by service providers, who want...
to increase the average revenue per subscriber by increasing the offered services, and customers who want a device that differentiates from others (e.g., different features and services). Each portable device should therefore be capable of transmitting and receiving signals over a wide range of frequency spectrum accommodating the different standards.

One approach to achieving multi-standard support is the use of a software-defined radio (see Fig. 1-2a). In the receiver, the radio signal is digitized immediately after the antenna with an analog-to-digital converter (ADC) for processing (i.e., demodulation) by a digital backend. Similarly, a digital-to-analog converter (DAC) in the transmit path immediately before the antenna converts the desired digital stream into an analog signal for transmission. Since the digital back-end can be adjusted “on the fly”, a software radio can receive and transmit different standards. Although the software radio is attractive for its high level of re-configurability, it is impractical at the current time as it requires performance from both ADC and DAC that is beyond the capabilities of current technologies (e.g., gigahertz sampling rates at low power consumption). Furthermore, the requirement on front-end filtering and switching between bands to channel can be a formidable engineering design challenge. It should also be noted that although a software-defined radio can support multi-standards, it can only handle one wireless service at any given time. Therefore, devices that need to support multiple concurrent services (e.g., talking on the phone and web surfing) may still require several software-defined radios. Despite exhibiting tremendous advantages with its ability to configure “on the fly”, a receiver that is based on Fig. 1-2a has never been realized for consumer applications. The most relevant topology is a hybrid approach where a wide-band RF front-end is placed between the antenna and the ADC [5]. The wide-band RF front-end downconverts the desired signal to a lower frequency, alleviating the requirement on the baseband circuits, and allowing the receiver to receive signals from multiple wireless standards. However, the hybrid demonstrator still suffers from some drawbacks, especially in terms of filtering requirement. The absence of pre-select filters necessitated a series of sampler, baseband filter, and demodulation stages that were developed via numerous simulation iterations. As mentioned in [5], these filters might be inadequate or sub-optimal for other wireless standards.

Fig. 1-2b illustrates a multi-standard radio that supports IEEE-802.11a/b/g, W-CDMA, and Group A of the IEEE-802.15.3a ultra-wideband (UWB) standards. Each standard is provided with its own antenna interface, preselect bandpass filter, and low-noise amplifier (LNA) to simply antenna interfacing and to minimize RF signal loss between the antenna and the LNA input. All standards share the same mixer and ADC, allowing for a reduction in complexity, parts count, and current consumption. However, the mixer must be capable of operating over a wide range of frequencies and meeting the diverse requirements imposed by various standards. The linearity and susceptibility to interference of a radio receiver, such as the one shown in Fig. 1-2b, is typically limited by the intermodulation distortion (IMD) generated by the first downconverting mixer [6]. This is especially true when the RF input signal to the mixer has a wide bandwidth, as IMD can become acutely detrimental to the BER performance [5]. For the receiver in Fig. 1-2b, this means that the linearity requirement on the mixer is much higher than a typical narrowband receiver as the bandwidth of the RF signal to the mixer can as wide as 1.5GHz (when configured to receive the UWB standard). Therefore, the mixer must be designed such that its linearity is as high as possible while operating within the supply current/power consumption budget.

Similarly, the DAC and upconverting mixer could be shared in the transmit path for a savings in parts count and current consumption.

### 1.2.3 Production Volume

Wireless devices (such as mobile phones) are currently produced by the millions. Such high volume production poses manufacturing issues to sustain high yield, satisfactory performance, and minimal variation between units. Current silicon-based transceivers make extensive use of monolithic integration to reduce both the total number of external components and parameter variations. This decreases assembly cost and increases the overall yield of working units. Further cost savings can be achieved when a large production volume is required (i.e., leveraging economies of scale), which reduces the fabrication cost per circuit.

Complementary metal oxide semiconductor (CMOS) and bipolar/complementary metal oxide semiconductor (BiCMOS) are silicon-based process technologies that enable high levels of analog and digital circuit integration, providing further cost savings. This high level of integration was made possible when semiconductor manufacturers included additional passive components for RF and analog circuit design, such as metal-insulator-metal (MIM) capacitors, thick interconnect metal for inductors, and high quality varactors [4]. Furthermore, modern BiCMOS process typically offer silicon-germanium (SiGe) heterojunction bipolar transistors (HBT), which offer the highest current gain-bandwidth product available today in a silicon-based technology.

The choice between various CMOS and BiCMOS processes is often a trade-off among various factors such as performance, price, time to market, and first-time-right design. While CMOS offers a 20%-50% in cost saving and higher level of integration with digital circuits [7], the highest RF
performance transceivers with lowest power consumption are more easily realized using BiCMOS processes, as bipolar transistors have higher transconductance per unit current than CMOS transistors. Furthermore, BiCMOS designs are less susceptible to parasitic variations [8] (due to higher gain-bandwidth product transistors) and increasing the likelihood of first-time-right designs, thereby eliminating expensive and time-consuming fabrication iterations. The need to reduce the number of iterations is increasing necessary with the ever decreasing prices of new wireless consumer products. However, a design that will be sold in large quantity can allow higher engineering cost as the overall cost can be amortized over many more units, keeping the price per unit the same or acceptable. In such scenarios, a CMOS chip that might takes longer and have a higher initial cost (e.g., due to more fabrication iterations) can be less expensive than an equivalent BiCMOS chip.

1.3 Objectives of This Thesis

The main objective of this work is to investigate and demonstrate mixer topologies for CMOS transceivers that exhibit substantially lower IM and harmonic distortions than the present state of the art. These transceivers are needed because wide acceptance of wireless technology has increased the likelihood of sub-optimal performance due to interference caused by multiple radios operating in proximity to each other at the same time. This is specific to tightly-networked areas where radios are prone to receive more unwanted signals. The prototype will be designed for high linearity within the 2-6GHz unlicensed band (i.e., most commercially important bands including the WiFi standards). Moreover, the architecture of the prototype is similar to that shown in Fig. 1-2b, allowing support for multiple standards operating in multiple bands. Since the linearity of the receiver is normally limited by that of the first mixer, the majority of the work will concentrate on improving the IM performance of the mixer. Broadband (0.8-10GHz) mixer topologies are favoured, as they provide a robust design that will eventually accommodate existing and emerging standards.

For reasons of cost and integration concerns, the proposed topologies should achieve higher linearity (reduce IM and harmonic distortions by 5 to 10dB) while satisfying several other constraints. The circuits should be low power and operate at supply voltages that are compatible with digital circuits (e.g., 1.2V at 0.13μm CMOS process). The circuit should also be robust to process and device variations (reducing trimming cost) and operating temperature shifts. The use of passive elements to increase linearity should also be minimized as they do not readily scale with lithographic scaling and can occupy much more silicon area than transistors (~100-1,000 transistors equivalent). Lastly, the proposed topologies should not require exotic interface connections with other building blocks within the transceiver chain, as these junctures might incur higher cost along with an increased power dissipation.

1.4 Organization of This Thesis

This thesis is organized as follows. Chapter 2 will provide the basic terminology and a background survey of RF receiver building blocks and systems. Detailed attention will be given to distortion mechanisms generated by BJT and CMOS transistors and how nonlinear processes affect the overall performance of a system. A brief review of different receiver architectures will also be presented. Chapter 3 surveys approaches to mitigating distortion effects in RF amplifiers. A discussion of the source of distortion in monolithic mixers will also be included in this chapter. The main focus of this work is presented in Chapter 4, Chapter 5, and Chapter 6 where feedforward compensation is applied to designs demonstrating better IM performance. In the proposed feedforward approach, two mixers are biased such that they generate anti-phase IM products that are summed resulting in "distortion-free" outputs. The results of the feedforward compensation applied to the input stage of a Gilbert multiplier is detailed in Chapter 4 while the results of the feedforward compensation applied to the switching stage

Organization of This Thesis

of the Gilbert multiplier are described in Chapter 5 and 6. Simulated and measured results for the two topologies will be detailed in these three chapters. Chapter 7 concludes this thesis with a summary of contributions and suggestions for future work.
Chapter 2

In a heavily networked environment, the achievable bit-error rate (BER) of a radio receiver depends not only on the sensitivity of the receiver but also its selectivity. Sensitivity refers to the smallest input signal level required by the receiver to achieve a data throughput at a particular BER, which directly determines how far the receiver can operate from the basestation. On the other hand, selectivity refers to the ability of a receiver to separate the desired signal from all undesired interfering signals that are received at other frequencies and amplify the desired signal. This chapter examines how different mechanisms influence sensitivity and selectivity at various levels in the design hierarchy (i.e., device, circuit, and system levels). As mentioned in the previous chapter, a modern receiver is required to have larger operating bandwidth to support various wireless standards that have different operating frequencies. With larger operating bandwidth, more interfering signals will be present at the input of the receiver and these interfering signals could potentially reduce the BER of the receiver. Hence, emphasis of this chapter is on how distortion affects the selectivity of a receiver.

This chapter first presents a brief description that provides an overview on how noise and interference affect the desired signal. This is followed by a discussion of the sources of noise and how it is quantified. The discussion then focuses on the distortion generation mechanisms, including the effect of distortion on channel selectivity. Next, trade-offs in cost, power dissipation, and dynamic range associated with commonly-used architectures will be detailed, focusing on the methods that have been proven to remove out-of-band signals. An example will be presented which illustrates calculation of the overall noise figure, linearity, and dynamic range of a receiver. Finally, basic operation of two common mixer topologies used in integrated design will also be presented and discussed in this chapter.

2.1. Spurious-Free Dynamic Range

Fig. 2.1a illustrates the output spectrum of a nonlinear amplifier or receiver where the desired output signal is compromised by both noise and intermodulation distortion (IMD). In the figure, 3rd-order IMD was assumed to be the dominant distortion and is generated by two in-band interferers. A quality of measure that accounts the effects of both noise and distortion is the signal to noise and distortion ratio (SINAD). This parameter compares the power level of the desired signal to the sum of noise and distortion. The sum of noise and distortion is used as they are uncorrelated. The power level of the noise floor and the 3rd-order IMD component can be determined by the equations indicated in the Fig. 2.1a, where $G_p$, $NF$, $IIP_3$, $P_{in, noise}$, and $P_{in, interferers}$ are the power gain, noise figure, input referred third-order intercept point, noise power at the input terminal, and power level of interferers at input terminal, respectively. Sources of noise and distortion will be presented and discussed in the Section 2.2 and Section 2.3 of this chapter where the discussions will provide an understanding of how noise and distortion are generated at transistor level.

As the maximum achievable bit error rate (BER) for a given modulation is inversely proportional to the SINAD, the limitation imposed by noise and distortion is often quantified by the spurious-free dynamic range (SFDR). This parameter, SFDR, quantifies the maximum level of interference that a receiver can tolerate with acceptable quality of reception when the desired signal is at or near its lowest power level (see Fig. 2.1b) [9]. This minimal detectable signal (MDS) is the lower bound on the SFDR, and assuming the input is impedance matched to the source, it is defined as

$$P_{in, min} = MDS = 10 \log(47) + NF + 10 \log(I_{out}) + SNR_{min}$$

(2-1)
Background information

SINAD = \frac{P_{desired \ signal}}{P_{output \ noise} + P_{IM}}

where \( k, T, \) and \( \text{SNR}_{\text{min}} \) are Boltzmann's constant, absolute temperature, bandwidth, and minimum signal to noise ratio (SNR), respectively. The unit for MDS is dBm or power ratio in decibels of the measured power referenced to 1 mW. The input power where dominant intermodulation (IM) distortions level generated in a two-tone test are equal in amplitude to the MDS defines the upper bound on the SFDR. Consequently, the SFDR can be expressed as

\[
\text{SFDR} = \frac{P_{\text{IP}}}{P_{\text{MDS}}} \quad (2-2)
\]

where \( n \) and \( \text{IP}_n \) are the order of the dominant IM and the dominant intercept point, respectively. Generally, either the 2nd or 3rd-order intercept point are used to determine upper bound. The output power for both fundamental and IM3 harmonics saturate when the input power is sufficiently large. This is due to voltage or current clipping and will be discussed in Section 2-3-2.

2.2. Noise

Random motions of charge and carriers create current and voltage fluctuations in circuits that are called noise. Ultimately, noise determines the smallest input signal level (i.e., sensitivity) that a communication system can process at an acceptable error rate. In RF IC design, resistors and transistors are the major contributors of noise, with different noise sources within a circuit dominating at different frequencies. For resistors, random thermal motion of electrons generates a noise that is proportional to the resistance value and absolute temperature. This thermal noise is (ideally) constant with frequency.

On the other hand, noise produced by active devices (e.g., BJTs and MOSFETs) is frequency dependent, and it typically increases at low frequencies.

Fig. 2.2 shows a generic (input referred) noise spectral density versus frequency plot for a BJT or MOSFET. The spectral density can be separated into two regions: one where the flicker noise, and another where thermal noise is the dominant contributor. These two regions are separated by the flicker noise corner frequency \( f_c \). Flicker noise is caused by the random trapping and detrapping of charge carriers when a bias current flows through a semiconductor. It is inversely proportional to frequency and directly proportional to the bias current. For a MOSFET, flicker noise is also inversely proportional to the product of the width and length of the transistor (i.e., the gate area). On the other hand, thermal noise is caused by a combination of intrinsic base/gate resistance, base and collector-current shot noise, and gate leakage current. These noise sources can be modelled by a series noise (voltage) source and a shunt noise (current) source (see Fig. 2.3) [10]. Numerical values for the equivalent noise sources are determined by evaluating the root-mean-square equations in the figure with \( k, T, K_1, K_2, q, \text{g}^m, r_{gs}, r_{gb} \), and \( I_B \) or \( I_G \) being the Boltzmann's constant, absolute temperature in Kelvin, two process dependent constants, electron charge, transconductance, base or gate extrinsic resistance, and base current or gate leakage current, respectively. The noise model for the MOSFET shown in Fig. 2.3 does not included induced gate noise. This induced gate noise arises due to capacitive coupling of the thermal drain noise via the gate-source parasitic capacitor. The capacitive coupling induces a gate current that is proportional to frequency [11]. As the operating frequency of the RF circuits to be described in this dissertation is below 100 GHz and is much lower than the maximum unity frequency of the CMOS process, induced gate noise is ignored.

Similarly, the noise source in RF building blocks and systems can be represented by two equivalent input noise generators. The effects of noise is often quantified by taking the ratio of the desired signal power to the total noise power, or the signal-to-noise ratio (SNR). In RF design, the ratio of the SNR at the input to SNR at the output (i.e., the noise factor) is used as a figure of merit to compare the performance of different circuit topologies, as it quantifies how much noise is added to the signal by the circuit. Assuming the source impedance is 50Ω, the noise factor at room temperature is defined as,
2.3. Distortion

Aside from noise, distortion also affects the dynamic range of a radio receiver. The inherently nonlinear nature of all active devices generates distortion due to weakly and strongly nonlinear behaviours. The two classifications differ in the magnitude of the input signal driving the system. An input signal with sufficiently small magnitude driving a weakly nonlinear system generates a predominantly linear response with distortion components that have a relatively small magnitude (e.g., typically lower than the linear response by at least 40-60dB). Strongly nonlinear behaviour refers to the undesirable components generated by circuit boundary conditions (e.g., waveform clipping), these nonlinearities typically occur when a sufficiently large input signal is used to drive a nonlinear system. This causes gain saturation [12].

2.3.1. Weakly Nonlinear Behaviour

In most cases, active devices operate at a bias or quiescent operating point with an ac input signal that creates a small perturbation on the bias/quiescent voltage and currents (e.g., $V_{in} \ll V_T$ in a common-emitter BJT stage). Linear circuit theory may then be used to predict the circuit behaviour around the bias/quiescent operating point. However, transistors generate distortions that are not predicted by linear circuit theory. Distortion may be generated from the desired input signal, or by the desired signal in combination with other (unwanted) signals that are present at the input. Weakly nonlinear distortion is usually assumed to be time-invariant and may be modelled mathematically by the Volterra series [13] or Taylor series expansion. This type of distortion usually has a negligible effect on the linear response, but has the potential to affect the overall system performance (selectivity).

Volterra series analysis is an extension to linear theory which was first proposed to describe the output of a nonlinear system [14]. Fig. 2.4 depicts a graphical representation of the Volterra series analysis. Central to Volterra series analysis is the $n$th-order filter blocks ($H_n$ in Fig. 2.4) that represent all $n$th-order (nonlinear) transfer functions. These filter blocks are also known as the $n$th-order Volterra kernel [12][15]. All filter blocks are excited by the same input $x(t)$, representing a collection of multiple sinusoidal tones of different amplitudes that generate the different $n$th-order responses, $y_n(t)$. For example, the output from the 1st-order and 2nd-order filters are the linear and 2nd-order nonlinear responses, respectively. The $y_n(t)$ are then summed to form the output $y(t)$, that contains all linear responses and the weakly nonlinear distortion components. Although Volterra analysis was not originally applied to the analysis of distortion in circuits, it can and is used to predict nonlinear distortion generated by transistor amplifiers [16].

Analysis of a nonlinear system using the Volterra series is carried out by first expressing the input signal as the sum of different frequency sinusoids,

$$x(t) = \sum_{k=1}^{N} A_k \cos(\omega_k t)$$

where $A_k$ and $\omega_k$ are the amplitude and frequency of the $k$th tone, respectively. From Fig. 2.4, nonlinearity of the system generate m-responses and each response is shaped by an $m$-order Volterra kernel. Response of the $n$th-order output is defined as

$$y_n(t) = \sum_{k_1}^{N} \sum_{k_2}^{N} \ldots \sum_{k_n}^{N} A_{k_1} A_{k_2} \ldots A_{k_n} y_{m_k}(a_{k_1} a_{k_2} \ldots a_{k_n})$$

$$y(t) = \sum_{k_1}^{N} \sum_{k_2}^{N} \ldots \sum_{k_m}^{N} A_{k_1} A_{k_2} \ldots A_{k_m} y_{m_k}(a_{k_1} a_{k_2} \ldots a_{k_m})$$

where $y_n(t)$ is the output signal and $y(t)$ is the overall output signal. The $y_n(t)$ are then summed to form the output $y(t)$, that contains all linear responses and the weakly nonlinear distortion components. Although Volterra analysis was not originally applied to the analysis of distortion in circuits, it can and is used to predict nonlinear distortion generated by transistor amplifiers [16].

Analysis of a nonlinear system using the Volterra series is carried out by first expressing the input signal as the sum of different frequency sinusoids,

$$x(t) = \sum_{k=1}^{N} A_k \cos(\omega_k t)$$

where $A_k$ and $\omega_k$ are the amplitude and frequency of the $k$th tone, respectively. From Fig. 2.4, nonlinearity of the system generate m-responses and each response is shaped by an $m$-order Volterra kernel. Response of the $n$th-order output is defined as

$$y_n(t) = \sum_{k_1}^{N} \sum_{k_2}^{N} \ldots \sum_{k_n}^{N} A_{k_1} A_{k_2} \ldots A_{k_n} y_{m_k}(a_{k_1} a_{k_2} \ldots a_{k_n})$$

$$y(t) = \sum_{k_1}^{N} \sum_{k_2}^{N} \ldots \sum_{k_m}^{N} A_{k_1} A_{k_2} \ldots A_{k_m} y_{m_k}(a_{k_1} a_{k_2} \ldots a_{k_m})$$

where $y_n(t)$ is the output signal and $y(t)$ is the overall output signal. The $y_n(t)$ are then summed to form the output $y(t)$, that contains all linear responses and the weakly nonlinear distortion components. Although Volterra analysis was not originally applied to the analysis of distortion in circuits, it can and is used to predict nonlinear distortion generated by transistor amplifiers [16].
where \( a_i = -o_i \) and \( H(n_a, n_b, \ldots, n_k) \) represent the \( n \)-th order nonlinear transfer function or Volterra kernel. For example, \( H_3(n_a, n_b, n_c) \) represents the 3-th order nonlinear transfer function that responds to the \( (n_a, n_b, n_c) \) tone and has a frequency component \( (n_c) \). It should also be noted that all \( n \)-th order nonlinear Volterra kernels are equivalent despite the different permutations of their arguments (e.g., \( H_3(n_a, n_b, n_c) = H_3(-n_a, n_c, n_b) \)). The response observed at the output terminal of this nonlinear system consists of the sum of all these \( n \)-th order individual response or

\[
y(t) = \sum_{k=1}^{n} y_k(t).
\]

(2-6)

Equation 2-5 and Equation 2-6 show that all nonlinear systems will generate a large number of harmonics consisting of all possible combinations of the N-input tones. However, many of these distortion components have negligible effects on the overall response since their magnitudes are relatively small compared to the overall response.

Although a nonlinear system generates a large number of distortion components, its nonlinear behaviour is usually adequately captured by simply quantifying the effects of the dominant distortion components. These dominant distortion components are classified into two categories: distortion due to the desired signal, and distortion due to one or more interfering signals. The first category includes effects such as harmonic distortion and gain compression/expansion. The dominant distortion effects are characterized by driving the device under test (DUT) with either 1 or 2 input tones. Intermodulation distortion, gain desensitization, and cross modulation are generally of greater interest when the DUT is expected to operate in an environment with many interferers (e.g., a heavily networked environment).

In general, a "paper and pen" distortion analysis with the Volterra series is limited to just a few transistors as the mathematics becomes laborious [17][18], and computer-aided analysis is normally employed [15]. In cases where the reactive elements (i.e., inductors and capacitors) can be ignored, the circuits become memoryless (e.g., the output no longer depend on its past values) and a Taylor series expansion (a special case of the Volterra series) with less cumbersome mathematics can be used. In a Taylor series expansion, the \( n \)-th order Volterra kernel is simplified by making the substitution \( H_n(a_1, a_2, \ldots, a_n) = H(a_1)H(a_2)\ldots H(a_n) \) [12][15].

2.3-1-1. Distortion in the absence of interference

In an ideal environment, where only the desired signal is present at the input to a nonlinear system, the system will still generate distortions that could hamper performance. The effects of these distortions may be expected to operate using a single tone test. In the single tone test, the system is driven by a single tone input, that is \( M = 1 \) in Equation 2-4, and the distortion generated at the output is quantified by parameters such as total harmonic distortion (THD) and gain compression/expansion.

Harmonic distortion (HD) refers to outputs that are integer multiples of the input frequency (see Fig. 2.5). Effects of harmonic distortion are often quantified by the root-mean-square of the sum of all the harmonic distortions (magnitude) compared to the fundamental response, or THD,

\[
\text{THD} = \sqrt{\frac{H_3(a_1, a_2, a_3)^2 + \ldots + H_k(a_1, a_2, \ldots, a_k)^2 + \ldots}{H_1(a_1)}} \times 100.
\]

(2-7)

THD is often quoted at a specific input level. Although harmonic distortions are of great importance when characterizing analog circuits (e.g., an audio amplifiers), they are often not as critical for RF applications. Most RF amplifiers have a bandpass response where the harmonic distortions fall outside the passband (see Fig. 2.5). However, performance of wideband RF amplifiers could be hampered by harmonic distortions. For example, the output of ultrawideband (UWB) amplifiers [19][20] that are

\[
(\pm k_1 a_1, \pm k_2 a_2, \ldots, \pm k_n a_n)
\]

(2-8)

where \( k_n \) can be zero or a positive integer. Looking at Equation 2-8, some distortion components could potentially fall within the desired signal band, thereby causing bit errors. For example, Fig. 2.6 depicts a scenario where the input to an amplifier consists of the desired signal and two out-of-band interferers.
magnitude of the IMD at IP3 pcoi + cojl measures the magnitude of the IMD at and 3''^-order measures the where f,:,, is the input power delivered to the nonlinear amplifier in dBm and AIMN is the dB difference hypothetical input (or output) level where the undesired n"^-order IMD is equal in magnitude to the desired signal. The input referred n"^-order intercept points are calculated from the equation These intercept points represent a (IPn). Note that the order of the IM products is the sum of all \( i + j + \ldots + \ell \) in Equation 2-8. For high distortion in the frequency domain.

Although the interferers are outside of the passband of the amplifier, they can generate an IMD that does interfere with the desired signal (e.g., the \((2\omega_1 - \omega_2)\) 3rd-order IMD in Fig. 2.6). Similar to the interference-free case outlined in Section 2-3-1-1, the potential for sub-optimal performance due to frequency mixing can be quantified by monitoring under controlled conditions. The commonly used experiment for this is the "two-tone" test, where two signals of the same amplitude but at different frequencies, (i.e., \(N = 2\) and \(A_1 = A_2 = \delta\) in Equation 2-4), are applied to the nonlinear system.

In the two-tone test, the two input tones represent two interferers. The objective of the test is to determine the magnitude of all IMD (or IM3) at a particular input power, by measuring the ratio between the fundamental tones and the n-th-order IMD at the output, or AIMN (e.g., AIM3 in Fig. 2.6). Note that the order of the IM products is the sum of all \(i, (i + j, \ldots, i + j\) in Equation 2-8. For high linearity, it is desirable that the magnitudes of all IMD components should be minimized, which translates to a large value for AIMN. Since the fundamental tones at the output vary as the interferers are moved from the inside to the outside of the passband of an amplifier, the required (and achievable) AIMN will change with it being highest within the passband and decreases as the interferers are moved away from it. The dependence of AIMN on the input amplitude can be eliminated by quantifying the IMD by driving the system with an input

\[
x(t) = \frac{1}{2} [A_1 e^{j\omega_1 t} + e^{j\omega_2 t}] + A_2 [1 + m(t)] e^{j\omega_1 t} + e^{j\omega_2 t}
\]

where \(m(t)\) and \(A_2\) are the modulating waveform and amplitude for the input tones, respectively. For characterizing desensitization and cross-modulation, the two-tone tests are chosen such that \(|m(t)| = |m_2|\) and the interferer is unmodulated (i.e., \(m(t) = 0\)). On the other hand, cross-modulation distortion is analyzed by setting the amplitudes of the desired signal and the interferer to the same, and setting the modulating waveforms so that \(|m(t)| = 1\). Effects of the desensitization and cross-modulation can then be analyzed by monitoring the output response at \(\omega_1\), or solving the Volterra series for

\[
h_I(\omega_1, \omega_2) = \frac{1}{(\omega_1 - \omega_2)} [4H_2(\omega_1) + \frac{1}{2} A_1 A_2 (1 + m(t))] e^{j\omega_1 t} + e^{j\omega_2 t}
\]

\[(2-11)\]

From Equation 2-11, it is evident that distortion generated by \(H_2(\omega_1, \omega_2, \omega_3)\) will add or subtract to the desired output response depending on its phase relative to \(H_2(\omega_1)\). Using the desensitization case as an example, Equation 2-11 shows that the output is proportional to \(A_2^2\) instead of being proportional to \(|m(t)|^2\) when \(|m(t)| = 1\).

2.3-2. Weakly Nonlinear Behaviour

Unlike weakly nonlinear distortion, strongly nonlinear behaviour is caused by an abrupt change in the operating condition, such as, a common-source transconductance amplifier stage (see Fig. 2.7a) that is driven into clipping. The transistor is biased by a DC voltage \(V_{BH}\) across the gate-source terminals that will cause a DC bias current \(I_{B(H)}\) to flow between drain and source terminals. The desired input
Background information

Figure 2.7: (a) Schematic of a common-source transconductance stage and (b) its DC transfer characteristic along with ac input and output waveforms.

Signal (\(V_m\)) is superimposed onto \(V_{GS}\) to generate a corresponding ac current at the drain terminal (\(I_{ds}\)) that is proportional to transconductance of \(M_1\). As the DC current is \(I_{ds}^{\text{Max}}\), the maximal positive and negative excursions of the drain current are between zero and \(2I_{ds}^{\text{Max}}\). Consequently, if \(V_m\) is large, (i.e., \(V_m \gg V_GS\)), clipping will occur on \(I_{ds}\) (see Fig. 2.7b). When \(V_m\) is a single sinusoidal tone, as shown in Fig. 2.7b, the strongly nonlinear behaviour will generate harmonic distortion components. Likewise, when the \(V_m\) comprises of several tones with different fundamental frequencies, intermodulation distortion will also be generated [22]. These harmonic and intermodulation distortion components can manifest itself in the voltage and/or current domains, compromising the linearity of the circuit or system. As mentioned earlier in this section, strongly nonlinear behaviour cannot be modelled by Taylor or Volterra series and are typically analyzed using computer simulator.

The distortion generated by strongly nonlinear behaviour can be alleviated by increasing the bias currents and/or supply voltages, however, they are not practical for RF circuits that operate with a stringent current consumption budget and supply voltages that are compatible with deep submicron digital circuits (~1V). One possible solution to ensure that the highest possible compression point is achieved is selection of an optimal load where the maximum voltage and current swing occurs simultaneously [23].

2-3-3. Examples of Distortion in Circuits.

In this section, basic single-ended and differential BJT and CMOS circuits will be used to illustrate how distortion is generated and how it can be predicted. The discussion will focus on how the nonlinear behaviour of transistors affects the 3rd-order harmonic and IM distortions (important distortion components in a typical radio receiver). Insights gained into how the distortion can be reduced or eliminated will also be presented.

2-3-3.1. BJT amplifiers

Shown in Fig. 2.8 are the schematic diagrams for a common-emitter and an emitter-coupled pair. These stages are frequently used in BJT preamplifiers and mixers to convert RF input voltages into a current that is either amplified or frequency translated. The output currents are proportional to the voltage between the base-emitter terminals, however, the approximately linear relationship between output current and input voltage only exists close to the original bias point.

Similar to the small-signal analysis, distortion generated by the common-emitter stage can be analyzed using an equivalent model (shown in Fig. 2.9). In the equivalent model, the voltage-controlled current source (\(G_1\)) models the nonlinear behaviour of the transistor and defines the output collector current as a function of the base-emitter voltage. This output collector current contains both the desired small-signal current and distortion components. For high frequency operation (>~0.1fT), parasitic capacitances along with extrinsic base and emitter impedances are also included. The parasitic base-emitter capacitance is modelled by a bias-dependent and a bias-independent capacitors \(C_{be}\) and \(C_{je}\), respectively. The equivalent model also includes the base current, \(I_b = \beta(\alpha)\). Base-collector capacitance is ignored in order to simplify the analysis [24].
As mentioned earlier, the voltage-controlled current source (transconductor $G_1$ in Fig. 2.9) generates the desired output current, but also distortion. Ignoring all reactive elements and intrinsic impedances, this nonlinearity relationship can be expressed as a Taylor series expansion

$$i_c = g_{m}v_{be} + g_{m}v_{be}^2 + g_{m}v_{be}^3 + \cdots = \frac{g_{m}}{V_T^2}v_{be} + \frac{1}{2}g_{m}\frac{1}{2}v_{be}^2 + \frac{1}{6}g_{m}\frac{1}{6}v_{be}^3 + \cdots$$ (2-12)

where $g_{m}$, $V_T$, and $v_{be}$ are the $n$th-order coefficient of the Taylor series, collector bias current, and the BJT thermal voltage (i.e., 25mV at room temperature). The $n$th-order Taylor series coefficient is determined by taking the $n$th derivative of the collector current with respect to the base-emitter voltage. If the relationship between $i_c$ and $v_{be}$ is linear, all coefficients higher than $g_1$ are zero. However, non-zero coefficients in the series expansion may generate harmonic and IM distortions due to weakly nonlinear behavior.

Linearity, as measured by $3^{rd}$-order IM distortion ($IM_3$) and $ΔM_3$ of the common-emitter stage can be determined from a two-tone test. As outlined in Section 2-3-1-2, distortion analysis involves replacing $i_c$ in Equation 2-12 with $sin^2\phi + \sin^3\phi\phi$, expanding, and then collecting the terms that have the frequency components corresponding to IM3 (i.e., $2\omega_1 - \omega_2$). For input amplitudes $A$ less than $V_T$, the IM3 is dominated by the cubic term in Equation 2-12 and can be expressed as

$$ΔM_3 = \frac{A^3}{3(V_T)^3}$$ (2-13)

Since the linear term from equation Equation 2-12 is $(i_c/V_T)^2$, the $ΔM_3$ for the common-emitter stage is

$$ΔM_3 = \frac{A^3}{3(V_T)^3}$$ (2-14)

Replacing $A/V_T$ with $i_c/V_T$, Equation 2-14 can be rewritten as

$$i_c = \frac{(i_c/V_T)^2}{(V_T)^2}$$ (2-15)

Equation 2-14 and Equation 2-15 suggest that either decreasing the amplitude of the signals appearing across the base-emitter terminals or increasing the total bias current (while keeping $i_c$ constant) would increase $ΔM_3$ and the linearity of the amplifier.

With increasing operating frequency, the influence of parasitic reactive elements on the overall response cannot be ignored. Their influence on weakly generated distortion can be analyzed using the Volterra nonlinear model as discussed in Section 2-3-1. The analysis that is presented here uses the harmonic input method [15] that follows the development in [24]. Similar to the Taylor series analysis, two sinusoidal tones with amplitude $A$ are applied to the input terminal, generating the output current

$$i_c(t) = \frac{A}{\sqrt{2}}h_n i_{\omega_1} + i_{\omega_2} + \cdots + i_{\omega_n}\phi_{\omega_1} + i_{\omega_2}\phi_{\omega_2} + \cdots + i_{\omega_n}\phi_{\omega_n}$$ (2-16)

where $n \in (1, 2)$. Equation 2-16 has the same form as Equation 2-5. It describes the desired output current along with all harmonic and IM distortions that are generated by two sinusoidal input tones. Evaluating the IM3 requires determination of the $3^{rd}$-order nonlinear transfer function $H_3(\omega_1, \omega_2, \omega_3)$ in Equation 2-16. The $n^{th}$-order nonlinear transfer functions are functions of all $(n-1)^{th}$-order nonlinear transfer functions. These transfer functions are determined by applying the harmonic input method using Equation 2-12 and Equation 2-16 recursively. The closed-form solution for the three nonlinear transfer functions required to compute IM3 for the common-emitter transconductance stage are:

$$H_3(\omega_1, \omega_2, \omega_3) = \frac{g_{m}}{V_T^2}h_3 i_{\omega_1} + i_{\omega_2} + \cdots + i_{\omega_n}\phi_{\omega_1} + i_{\omega_2}\phi_{\omega_2} + \cdots + i_{\omega_n}\phi_{\omega_n}$$ (2-17)

where $g_{m}$, $V_T$, and $i_{\omega_n}$ are the $n$th-order coefficient of the Taylor series, collector bias current, and the BJT thermal voltage (i.e., 25mV at room temperature). The $n$th-order Taylor series coefficient is determined by taking the $n$th derivative of the collector current with respect to the base-emitter voltage. If the relationship between $i_c$ and $v_{be}$ is linear, all coefficients higher than $g_1$ are zero. However, non-zero coefficients in the series expansion may generate harmonic and IM distortions due to weakly nonlinear behavior.

Linearity, as measured by $3^{rd}$-order IM distortion ($IM_3$) and $ΔM_3$ of the common-emitter stage can be determined from a two-tone test. As outlined in Section 2-3-1-2, distortion analysis involves replacing $i_c$ in Equation 2-12 with $sin^2\phi + \sin^3\phi\phi$, expanding, and then collecting the terms that have the frequency components corresponding to IM3 (i.e., $2\omega_1 - \omega_2$). For input amplitudes $A$ less than $V_T$, the IM3 is dominated by the cubic term in Equation 2-12 and can be expressed as

$$ΔM_3 = \frac{A^3}{3(V_T)^3}$$ (2-14)

Replacing $A/V_T$ with $i_c/V_T$, Equation 2-14 can be rewritten as

$$i_c = \frac{(i_c/V_T)^2}{(V_T)^2}$$ (2-15)

Equation 2-14 and Equation 2-15 suggest that either decreasing the amplitude of the signals appearing across the base-emitter terminals or increasing the total bias current (while keeping $i_c$ constant) would increase $ΔM_3$ and the linearity of the amplifier.

With increasing operating frequency, the influence of parasitic reactive elements on the overall response cannot be ignored. Their influence on weakly generated distortion can be analyzed using the Volterra nonlinear model as discussed in Section 2-3-1. The analysis that is presented here uses the harmonic input method [15] that follows the development in [24]. Similar to the Taylor series analysis, two sinusoidal tones with amplitude $A$ are applied to the input terminal, generating the output current

$$i_c(t) = \frac{A}{\sqrt{2}}h_n i_{\omega_1} + i_{\omega_2} + \cdots + i_{\omega_n}\phi_{\omega_1} + i_{\omega_2}\phi_{\omega_2} + \cdots + i_{\omega_n}\phi_{\omega_n}$$ (2-16)

where $n \in (1, 2)$. Equation 2-16 has the same form as Equation 2-5. It describes the desired output current along with all harmonic and IM distortions that are generated by two sinusoidal input tones. Evaluating the IM3 requires determination of the $3^{rd}$-order nonlinear transfer function $H_3(\omega_1, \omega_2, \omega_3)$ in Equation 2-16. The $n^{th}$-order nonlinear transfer functions are functions of all $(n-1)^{th}$-order nonlinear transfer functions. These transfer functions are determined by applying the harmonic input method using Equation 2-12 and Equation 2-16 recursively. The closed-form solution for the three nonlinear transfer functions required to compute IM3 for the common-emitter transconductance stage are:

$$H_3(\omega_1, \omega_2, \omega_3) = \frac{g_{m}}{V_T^2}h_3 i_{\omega_1} + i_{\omega_2} + \cdots + i_{\omega_n}\phi_{\omega_1} + i_{\omega_2}\phi_{\omega_2} + \cdots + i_{\omega_n}\phi_{\omega_n}$$ (2-17)

where $g_{m}$, $V_T$, and $i_{\omega_n}$ are the $n$th-order coefficient of the Taylor series, collector bias current, and the BJT thermal voltage (i.e., 25mV at room temperature). The $n$th-order Taylor series coefficient is determined by taking the $n$th derivative of the collector current with respect to the base-emitter voltage. If the relationship between $i_c$ and $v_{be}$ is linear, all coefficients higher than $g_1$ are zero. However, non-zero coefficients in the series expansion may generate harmonic and IM distortions due to weakly nonlinear behavior.
for a two-tone input. Again, the harmonic input method can be applied to account for the effects of the reactive components and extrinsic impedances. In this case, the $|\text{IM}_3|$ determined from a two-tone test is (assuming that $A$ is the magnitude of the each tone)

$$|\text{IM}_3|_{ECP.V} = \frac{|I_{out}|^2}{A^2} = \left|H_1(s_1)H_1(s_2)|l + (s_1-s_2)\frac{Z_{d}}{Z_{s}}(2s_1-2s_2)\right|$$

where $g_m = I_m/(2V)$, $C_{BE} = C_{BE} + C_{CE}$, $Z_{s} = Z_{s} + Z_{d}$ and

$$H_1(s) = \frac{g_m}{sC_{BE}Z(s)} - \frac{sC_{BE}Z(s)}{\beta(s)A(s)} + \frac{sC_{BE}Z(s)}{\beta(s)A(s)}.$$

In the previous two examples, only the exponential relationship between the collector current and base-emitter voltage in the BJT was considered as it is the dominant source of distortion [15]. Although considered to be distortion-free, parasitic depletion capacitances of the base-emitter and base-collector junctions also contribute to the overall distortion. Fortunately, their contributions are only significant when the transistor is operating close to the unity current gain frequency [15]. Consequently, parasitic capacitances and intrinsic impedances are assumed to be distortion-free and only affect the distortion behaviour at high frequency, they are distortion-free. Linearity of subsequent circuit topologies can be analyzed and quantified using the Taylor series whenever applicable. This avoids laborious mathematics that is generally involved with Volterra series or the harmonic input method. Distortion in high frequency operation will be analyzed using computer simulation (i.e., numerical).

As an example, $I_{M3}$ generated by a common-emitter amplifier (Fig. 2.9) for a range of operating frequency was determined using Cadence Spectre™ simulator and Volterra series (Equation 2-21). The simulated and calculated results are plotted in Fig. 2.11. In the simulation, a BJT from a SiGe BiCMOS process with a maximum $f_T$ of 120GHz [26] was used, while the parameters used to evaluate Equation 2-21 were extracted from DC simulation. The values predicted by Taylor series (Equation 2-13) is also included in the figure.

As shown in the figure, both Spectre™ simulator and the Volterra series predict that the magnitude of IM3 component to be almost constant up to about 1GHz. The values predicted by Volterra and Taylor series are approximately 2.5 - 2.6 times larger. The main cause for this discrepancy is that the exact intrinsic base and emitter resistances were not available from the DC simulation and were assumed to be zero in calculation. However, when the operating frequency is between 1 and 20GHz, the Spectre™ simulator predicts that the magnitude of the IM3 increases with frequency, doubling in magnitude, while the Volterra series predict an opposite behaviour with a slight decline in magnitude. This deficiency is probably the result of factors such as ignoring the effect of base-collector capacitance and assuming a constant value for current gain, beta. It should be noted that the variation in IM3 as predicted by Spectre™ simulator will only cause an approximate change of 1.5dB in $I_{IP3}$. As the operating frequency increases further, both Spectre™ simulator and Volterra series predict that the magnitude of the IM3 generated by the common-emitter amplifier will decrease. The values obtained from both Spectre™ simulator and Volterra series agree well with each other.

2-3-3-4. CMOS amplifiers

The CMOS equivalent of the common-emitter and emitter-coupled pair are the common-source and source-coupled pair (see Fig. 2.12). The basic operation of the CMOS input stages are similar to the BJT where the output current (drain current) is voltage-controlled by the input (i.e., gate-source) voltage. Assuming a long-channel MOSFET, the relationship in saturation between drain-source current and input voltage is given as [27]
The nonlinearities present in the CMOS transistors but not predicted by Equation 2-28 are graphically illustrated in Fig. 2.12. The figure shows the normalized bias current ($I_{DQ}$) from simulation of a common-source stage for increasing gate-source voltage in a 0.13μm CMOS process. The first three derivatives ($g_1$, $g_2$, and $g_3$) of the bias current with respect to the gate-source voltage are also shown in the figure. The numerical values of $g_1$ to $g_3$ are proportional to the first three coefficients of the Taylor series expansion. As seen in the figure, the bias current is related to the square of the gate-source voltage (as predicted by Equation 2-28). The first derivative or the transconductance is (almost) linear with respect to $V_{GS}$ in the strong inversion region, but the second and third derivatives are complex functions of $V_{GS}$. The plots also show that the input gate-source voltage affects the distortion generated by a CMOS transistor in the output current. Furthermore, these plots differ from the predictions of the simple MOS model (Equation 2-28) where the second derivative is constant and the third derivative is zero.

Behaviour of the second- and third-order derivatives can be broken down into three regions of operation: weak inversion, moderate inversion, and strong inversion. The three regions differ in which of the two mechanisms, drift or diffusion, is the dominant contributor to the output drain current. In the weak inversion region, the drain current is primarily caused by diffusion current and the FET behaves like a bipolar transistor. As a result, the bias current and its successive derivatives are exponentially related to the gate-source voltage and this behaviour is reflected by the graphs shown in Fig. 2.13. Increasing the gate-source voltage further pushes the transistor into moderate inversion where both diffusion and drift currents are comparable and contribute to the drain current. Within this region, the gate-source voltage and threshold voltages are approximately equal to each other. Increasing the gate-source voltage increases the value of second-order derivative ($g_2$) but decreases the value of third-order derivative ($g_3$) until $g_2$ reaches a local maxima while the $g_3$ is zero. The gate-source voltage which this phenomena occurs is known as the threshold voltage [29]. At this gate-source voltage, the third-order distortion is minima ($g_3 = 0$) but the second-order distortion is highest. Increasing the gate-source voltage further leads to a decline in the value of $g_2$ but an increase in the magnitude of $g_3$. The transistor begins to operate in a strong inversion region and the drain current flows primarily due to diffusion. As shown in the figure, the magnitude of $g_2$ and $g_3$ decrease with increasing gate-source voltage, suggesting that both second- and third-order distortion at the drain current can be suppressed by biasing the transistor with a larger gate-source voltage. This trend is similar to the BJT (e.g., increasing bias current to increase linearity) suggesting that many linearization techniques developed for the BJT are also applicable to CMOS transistors.

Although Fig. 2.13 highlights that the distortion observed at the drain terminal is solely dependent on the gate-source voltage like the BJT, the distortion in CMOS transistor is also contributed from other secondary effects such as drain voltage modulation, mobility degradation, and series drain/ source resistance [28]. The contribution of drain voltage modulation can be become significant in modern CMOS processes where the drain-source bias voltage is often below 1.5V [30].

Models that replicate the behaviour shown in Fig. 2.13 exist but they require numerical analysis in order to evaluate the distortion components (even at low frequency) [31]. Consequently, it is often advantageous to analyze distortion of CMOS transistors with the aid of simulator.
2.4. Receiver Architecture

The main objective of the front-end in an RF receiver is to downconvert the received signal band while ensuring that the power level of the output signal is within the dynamic range of the demodulators and/or ADC. All receiver architectures use the main building blocks: preamplifier, mixer, synthesizer, and filters to receive and amplify the desired signal. Frequency downconversion is needed as high performance demodulators or ADCs that operate directly on the radio frequency signal would consume greater power and have poorer SFDR [32]. Each of the building blocks has its own gain, noise figure, and distortion performance that will influence the overall receiver. Architectures differ in how the main building blocks are connected to reject out-of-band interferers. In this section, architectures such as the heterodyne, homodyne, and the image-reject receiver will be described. A discussion on the effects of noise and distortion at this level in the system hierarchy on sensitivity and selectivity will also be highlighted.

2-4-1. Heterodyne Receiver

Early receivers adopted the heterodyne architecture, where the preamplifier, mixer, an local oscillator (LO), and filters are connected as shown in Fig. 2.14 [33]. After the RF signal is received by the antenna, it passes through a pre-select filter where signals outside the desired band are rejected. After the pre-select filter, a low noise amplifier is used to increase the RF signal power level while adding minimal noise. This signal is then passed through a mixer where it is downconverted (frequency multiplication) by the LO to an intermediate frequency (IF). The LO signal can be $\omega_{IF}$ higher/lower (i.e., high or low-side injection) than the center of the desired signal band, with low-side injection often preferred because it is easier to design a synthesizer with low phase noise at a lower oscillating frequency. Since the LO signal can be high or low side injected, the mixer will downconvert bands that are located at a frequency offset of $\omega_{LO}$ above and below the LO (i.e., desired and image bands).

While one of these bands contains the desired signal (for single sideband modulated signals), the other band (known as the image band) contributes only noise and interference that reduces the overall SINAD. Rejection of the interference from the image band is realized by putting an image-reject filter before the mixer. The desired RF signal is then selected by an IF filter after the mixer. The receiver can be tuned to receive different signal bands by varying the LO (called the super-heterodyne receiver [34]). Although the RF front-end (the components from antenna to IF filter in Fig. 2.14) provides gain and amplifies the desired signal, it is usually inadequate. Due to power consumption constraint and stability issue, it is often not possible to increase the gain along the RF front-end path. Instead, a high gain IF amplifier that operates at a much lower frequency that can be designed at a much lower power consumption is often used to provide necessary gain. This IF amplifier (shown as a variable gain amplifier or VGA in Fig. 2.14) typically provides a gain of 80dB or more [35]. The use of multiple building blocks in the heterodyne architecture allows the gain to be distributed between the stages (i.e., between the preamplifier and mixer in Fig. 2.14). This reduces the tendency of any one stage to oscillate and improves the overall stability. Furthermore, the use of multiple building blocks allows the heterodyne architecture to offer sensitivity and selectivity. Sensitivity is improved by utilizing the low-noise preamplifier where it help to suppress noise at the input. High selectivity is achieved by placing frequency selective bandpass filters in between the building blocks to filter out undesirable out-of-band components. These filters are usually external (i.e., off-chip) and implemented using surface acoustic wave or ceramic filters. Although these filters reject out-of-band interferers, in-band interferers via nonlinear behaviour of the individual building blocks can still affect the overall performance (e.g., reduction in SINAD). Some recent examples of super heterodyne receivers are found in [36] and [37], where external pre-select and image-reject filters are used.

2-4-2. Homodyne Receiver

The heterodyne architecture provides superior sensitivity and selectivity performance, but the drive for greater on-chip integration and lower power consumption has led to the development of the homodyne receiver [38]. The homodyne receiver architecture also uses a preamplifier, mixer, and LO to amplify and downconvert a modulated RF input signal for baseband processing. Fig. 2.15 shows the building blocks of the RF front-end of a homodyne receiver. In a homodyne receiver, the frequency of the LO is set at the centre of the RF signal band, therefore, it is also (commonly) known as a direct-conversion and zero-IF receiver. While there are similarities to the heterodyne, the homodyne architecture differs in a few major aspects. Firstly, with the LO and RF at the same frequency, the IF is at DC (0 Hz) so there is no image band. Therefore, an image-reject filter is not required, which reduces the overall component count. Secondly, to prevent any loss of information, quadrature (I and Q signals)
Background information

are required at baseband for phase and frequency modulation schemes. Thirdly, bandwidth requirements for the circuits after the mixers are relaxed as they are only required to operate up to the bandwidth of the desired signal (e.g., a digital filter can be used for the baseband filters).

The bandwidth of the desired signal extends to DC, thus any DC offset voltages (or currents) can corrupt the desired signal. DC offset can be generated by static sources such as imperfect matching in differential circuits, poor isolation between the transmitter and receiver paths and dynamic sources such as weakly nonlinear mechanism. Imperfect matching in differential circuits arises due to limitation in fabrication generating magnitude and phase mismatch between balanced signal. Imperfect LO isolation between transmit and receive paths along with capacitive and substrate coupling gives rise to the situation where a LO signal could “leak” to the input of the preamplifier or the mixers. This leakage will then mix with the LO at the mixer (self-mixing), thereby generating a DC component at the output. DC offsets are generated through weakly nonlinear mechanism in the form of even-order harmonics and IMD. In the preamplifier, interferers (such as generated by $H_2(\omega_L, -\omega_L)$) will produce low frequency even-order IMD close to DC. Similarly, even-order harmonics due to IMD generated by the mixers (such as generated by $H_2(\omega_L, -\omega_L)$ and $H_2(\omega_L, -\omega_L)$) may also generate a DC offset and produce distortion that masks the desired signal. While static offsets could be reduced by a one-time trimming, dynamic DC offsets often required some form of close-loop active systems (i.e., a digital signal processing) that monitor and compensate for the offset.

Other than the DC offset problem, the SNR of a homodyne receiver can also be impaired by flicker noise. When an active mixer is used in a homodyne receiver, flicker noise degrades the output SNR at the lower frequencies (see Fig. 2.16a). Note that the lower bound on the SNR is evaluated by integrating noise spectral density function over the bandwidth of the signal/IF filter. Often, this problem is alleviated by downconverting an IF of −10−100 kHz instead of DC. This low-IF architecture alleviates the DC offset and even-order distortion requirements as the baseband signal spectrum no longer encompasses the DC. By shifting to an IF slightly above DC, the IF signal will be subjected to less flicker noise resulting in a higher output SNR (see Fig. 2.16b).

The capability of building an RF front-end without image-reject and IF filters has made the homodyne and low-IF architectures an attractive choice for modern receivers where more receivers are realized on an integrated circuit in order to reduce assembly time and cost. Some recent examples of homodyne and low-IF receivers can be found in [39][40]. For wide dynamic range, these receivers employed techniques to compensate for the DC offset and mismatch in the quadrature signals, reduce 2nd-order nonlinearities, and lower the flicker noise corner frequency.

2-4-3. Image-Reject Downconverter

Aside from the zero- and low-IF architectures, the image-reject filtering in the heterodyne receiver can also be removed by using an image-reject downconverter. In an image-reject downconverter, the RF signal and image signals are processed differently, allowing the cancellation of the image signal. Unlike the zero- and low-IF architectures, the output of an image-reject receiver is not affected by the DC offset and flicker noise problems.
Usually determined by the desired data-rates and is specified in terms of the sensitivity, selectivity, and spurious free dynamic range of the receiver.

The overall noise and distortion of a cascaded chain are determined by the gain, noise, and IIP\textsuperscript{3}\textsuperscript{n} of the individual building blocks. For a front-end that consists of i-cascaded stages (see Fig. 2.18), the total noise figure \((\text{NF}_\text{Total})\) between the input and output is determined using the Friis' formula \([43]\),

\[
\text{NF}_\text{Total} = 10\log\left(\frac{1}{\text{NF}_{\text{Pre}}} + \frac{1}{\text{NF}_{\text{Img}}} + \ldots\right)
\]

where \(\text{NF}_{\text{Pre}}\) and \(\text{NF}_{\text{Img}}\) are the available power gain of the i\textsuperscript{th} stage and the input-referred n\textsuperscript{th}-order intercept point of the m\textsuperscript{th} stage in watts. Equation 2-30 predicts the worse case IIP\textsuperscript{3}\textsuperscript{n} because it assumes that IMD produced by each stage is uncorrelated and independent from all other \([15]\). The gain, noise figure, and linearity performance of the individual building blocks are often chosen via iteratively evaluating Equation 2-29 and Equation 2-30 until the desired performance for the overall receiver is achieved. The overall noise figure and linearity can be traded-off and optimized the SFDR by proper distribution of gain, noise figure, and linearity among the individual stages.

Results in Table 2-1 show that the trade-off of noise figure for linearity in a mixer helps to improve the overall IIP\textsuperscript{3}\textsuperscript{n} and dynamic range of a receiver by 6.5dB and 2.6dB, respectively. With higher IIP\textsuperscript{3}\textsuperscript{n}, the receiver in Case B is less susceptible to interferers that are picked up by the antenna thereby allowing the radio to sustain an acceptable BER when operating in a heavily networked environment. An application where the high linearity of the receiver in Case B will be highly beneficial is the wireless data communication via WiFi in an office building where wireless link performance is often limited by interference and distortion instead of sensitivity and noise. Conversely, being able to tolerate interferers with higher power level also means that the receiver in Case B can have a wider operating bandwidth \([5]\) and supports various wireless standards. As mentioned in Section 1.2.2, this multi-standard support will help to reduce complexity, part counts, and current consumption. The values in Table 2-1 also showed that the trade-off of noise figure for linearity in the mixer resulted in a
degradation of 2.6dB in the overall noise figure of the receiver in Case B. However, the effect of the larger noise figure of the mixer can be suppressed by the preceding preamplifier. For example, increase the gain of the preamplifier by 5dB will help to lowe the overall noise figure of the receiver in Case B by 2.6dB, allowing the receiver in Case B to have similar sensitivity performance and a better IIP3 performance.

Table 2-1: Performance of the receiver shown in Fig. 2.19 with two different mixers.

<table>
<thead>
<tr>
<th></th>
<th>Case A</th>
<th></th>
<th></th>
<th></th>
<th>Case B</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>Cumulative Gain, dB</td>
<td>-2.0</td>
<td>13.0</td>
<td>11.0</td>
<td>11.0</td>
<td>9.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative NF, dB</td>
<td>2.0</td>
<td>3.5</td>
<td>3.6</td>
<td>4.1</td>
<td>4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative IIP3, dBm</td>
<td>100.0</td>
<td>2.0</td>
<td>2.0</td>
<td>-6.6</td>
<td>-6.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cumulative SFDR, dB</td>
<td>N/A</td>
<td>115.0</td>
<td>115.0</td>
<td>108.8</td>
<td>108.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For example, an IEEE802.11b receiver at the edge of the range of a base station, operates at its lowest data rate (1Mbps) will be unaffected by interferers to maintain the desired BER of 10^-5. From specification given in [2], the noise figure of the receiver must be lower than 13.6dBm while its IIP3 must be lower than -6.5dBm. Note that the values were determined for a desired sensitivity of -80dBm and the rejection of two -32dBm adjacent channel interferers, a 3dB SINAD margin was also included in the calculation. Similarly, for the desirable BER of 10^-5, a IEEE802.11a/g receiver must have a noise figure of 10.2dBm or lower along with an IIP3 of -2.8dBm or better. Again the receiver is assumed to be at the edge of the range of the base station and is operating at the lowest data rate (6Mbps). Comparing the noise figure and IIP3 requirement for the receivers in this example to the values shown in Table 2-1 indicated that the Case B receiver in Table 2-1 will meet the specification whereas the Case A receiver in Table 2-1 will not. The 6.4dBm improvement in IIP3 help to reduce the 3rd-order IMD generated by in-band interferers by 13dB, allowing the Case B of Table 2-1 to tolerate stronger interferers.

Much research is still needed to develop integrated receivers that generate lower IM distortion while operating with a low voltage and current consumption. Since the overall linearity of the receiver is usually limited by the first downconverting mixer, the subsequent discussion will focus on reducing IM distortion within a mixer, thereby improving the linearity of the overall receiver.

2.5. Mixer

A mixer is a building block that has two inputs and one output. Its main operation is to transfer a desired signal (one of the two inputs) from one frequency band to another band utilizing the second input (local oscillator). Although this operation is non-linear in nature, the mixer should have a linear relationship between the desired signal and the output (a one-to-one correspondence). In its simplest form, this frequency shift can be realized by multiplying the desired signal with another sinusoidal signals in time domain or

\[
\{A_1 \cos(\omega_t + \phi_1)\} \{A_2 \cos(\omega_2 t + \phi_2)\} = 0.5A_1A_2 \cos(\phi_1 + \phi_2)\left[\cos(\omega_1 t + \omega_2 t) + \cos(\omega_1 t - \omega_2 t)\right]
\] (2-31)

where \(A_1\), \(A_2\), \(\phi_1\), and \(\phi_2\) are the amplitude, angular frequency, and phase of the sinusoidal inputs. For downconversion, the difference term \((\omega_1 - \omega_2)\) is kept and the plus term \((\omega_1 + \omega_2)\) is filtered. Similarly, the desired frequency is kept while the difference frequency is filtered out after up-conversion.

In practice, the mixing function is accomplished by multiplying the desired signal by a ±1 square wave. Fig. 2.20 shows a conceptual view of this multiplication process. The switch alternatively connects the IF to either of the unity gain amplifier outputs at a rate equal to the LO frequency. This is equivalent to multiply the RF signal by the square wave

\[
S(t) = \frac{1}{2} \sum_{n=-\infty}^{\infty} \sin^2 \left(\frac{\pi n f_{\text{LO}}}{f_{\text{RF}}} t\right) = \frac{1}{2} \cos(5\sin(2\pi f_{\text{LO}} t)) + \frac{1}{2} \sin(5\sin(2\pi f_{\text{LO}} t)) \ldots
\] (2-32)

and generating the frequency shift. This mixing function is often implemented using BJT and CMOS transistor, where an LO voltage signal is applied to the base and gate terminal to turn on and off continuously to produce the mixing effect.

Two common mixer topologies, active and passive, will be described in this section. The discussion will provide review of the basic operation of each topology along with its attractiveness and drawbacks. For simplicity, a downconverting mixer (the difference term in Equation 2-31) will be used as example in the subsequent discussion.

2.5.1. Active Mixer

Fig. 2.21 shows the schematic of a double-balanced active mixer. Although the mixer was originally proposed as a frequency demodulator [44] and analog multiplier [45], it (and its variant) has since become the most commonly used mixer in radio transceiver. This mixer is commonly known as the Gilbert multiplier and in its conventional form consists of three stages: voltage to current (transconductor), switching stage (switching quad), and current to voltage stage (transimpedance stage or load). These three stages are highlighted in Fig. 2.21. The mixer is classified as an active mixer because the switching stage dissipates bias current. A CMOS variant of Gilbert mixer can be realized by replacing the BJT by FET.

For normal operation, the input RF signal is applied to base terminals of transistors Q1-2 at the input transconductor stage generating a differential output current \(i_{t,\text{out}}^+\) and \(i_{t,\text{out}}^-\). This output
current is proportional and related to the input voltage by the transconductance of the input stage. It is injected into the common-emitter nodes of the switching quad. At the switching quad stage, a balanced LO signal drives the four transistors \( Q_{3,4} \) toggling the four transistors to turn on and off completely. Whenever the voltage across the base terminals of each transistor pairs \( Q_{3.4} \) and \( Q_{5.6} \) is larger than \( 4V_T \) (~100mV or equivalent to -13dBm dissipated across 100Ohms differential termination), one transistor within the pair will be on while the other is off (e.g., during positive LO incursion, transistors \( Q_3 \) and \( Q_5 \) are on while \( Q_4 \) and \( Q_6 \) are off). This toggling mechanism forces the RF and bias current to alternate flow through between the two pairs of transistors, and is akin to multiply the RF current with a square wave that swings between +1 (Equation 2-24). This multiplication process shifts the frequency of the output of the transconductor from RF to IF, and this IF current is then converted back into voltage/power in the transimpedance stage. In Fig. 2.21, this is realized by using two resistors to develop voltages at the IF output terminals. Besides resistive load, a tuned load or a transimpedance amplifier (more details in Chapter 3) can also be used to convert the output to drive subsequent stages.

One of the main reasons that Gilbert multiplier is the popular topology for single-chip transceiver is because it requires a relatively small LO swing. As mentioned earlier, the four transistors within the switching quad turn on or off whenever the voltage difference between the base terminals of each transistor pairs exceed \( 4V_T \). This required voltage swing can be easily generated by an on-chip oscillator and is much lower than what is needed in passive mixer (more details in the ensuing sub-section). Besides the advantage in needing a smaller LO swing, the Gilbert multiplier topology is also chosen because it allows the mixer to have different conversion gain while providing superior port-to-port isolation. Assuming the switching mechanism is modelled by \( g_{m1} \), the voltage conversion gain for the mixer shown in Fig. 2.21 is \( g_{m1}V_{if} \), where \( g_{m1} \) is the transconductance of transistors \( Q_{3.4} \). Consequently, transconductance (bias current, \( I_{bias} \)) of the input stage and/or load resistance can be adjusted to achieve the desired voltage conversion gain. Since the Gilbert-multiplier has a double-balanced structure (differential LO and RF), the mixer provides isolation for the three ports from each other when the differential inputs and output are observed. This helps to minimize LO feedthrough at the RF and IF terminals, where they could result in self-mixing and output desensitization, respectively.

Although the Gilbert multiplier mixer exhibits many advantages, it introduces distortion that compromises and reduces the BER of the overall receiver. The primary source of the distortion is contributed by the inherent nonlinearity of the transistors (see Section 2.3). Circuit techniques that reduce this distortion is the focus of this dissertation and will be detailed in the upcoming chapters.

### 2-5-2. Passive Mixer

A different category of mixer is the passive mixer, where the switching stage does not consume any bias current. Fig. 2.22 shows a simulation testbench for a double-balanced passive FET mixer. Similar to the active mixer, the FETs are also operated as switches and turned on/off alternatively by a balanced LO signal. In the on-state, the transistors are on and have low impedance (few Ohms) between their drain and source terminal, shorting and connecting the RF and IF terminals. In the off-state, the transistors are off and there exists a high impedance (larger than tens of kilo-Ohms) between the drain and source terminals, isolating the RF and IF terminals. This modulates the RF signal, shifting the frequency of the input signal to IF. It should be noted that the connection for the four transistors of the passive mixer is the same as the switching stage of the Gilbert multiplier shown in Fig. 2.21 (replacing the BJT of the switching quad with FET).

Besides the obvious advantage of dissipating zero power (absent of bias current), a passive mixer also offers two other advantages over their active mixer counterparts: better performance in flicker noise corner [46] and higher linearity. As discussed in Section 2-1, flicker noise is generated by the random trapping and detrapping of charge at the silicon-silicon dioxide interface and is proportional to its bias current. In passive mixer, the transistors do not have bias current resulting in flicker noise corner that are significantly lower than its active mixer counterpart. For example, a recent work demonstrating FET passive mixer with a flicker noise corner of 45kHz [46], that is comparable to a BJT design and superior to FET Gilbert multiplier (typical flicker noise corner is in the hundreds of mega-Hertz). Although the absence of bias current flowing through the transistors helps reduce the flicker noise corner of the passive, flicker noise cannot be totally removed (reducing flicker noise corner to...
increasing LO swing, the conversion increases and reaches a value of -5.27 dB at $V_{\text{LO,pk-pk}} = \frac{DD}{2}$.

For example, for peak-to-peak swing less than 0.2VDD, the conversion is lower than -16.42 dB. With the amplitude of LO is small and the transistors are operating like cascode transistor instead of a switch.

The terminals is normalized to the supply voltage (i.e., $I_{\text{2V}}$). The performance is only attainable when the passive mixer is driven by high LO power. Fig. 2.23 shows the frequencies at 100MHz and 90MHz, respectively. The value of the peak-to-peak LO swing at the gate terminals is

As shown in the figure, the conversion gain of the passive mixer's conversion gain is lowest when the amplitude of LO is small and the transistors are operating like cascode transistor instead of a switch. For example, for peak-to-peak swing less than 0.2VDD, the conversion is lower than -16.42 dB. With increasing LO swing, the conversion increases and reaches a value of -5.27 dB at $V_{\text{LO,pk-pk}} = V_{\text{DD}}$. This is the conversion gain value is 1.35 dB lower than the theoretical limit of -3.92 dB (or equivalent to 85.6% of maximal voltage gain, $2/\pi$). Further increase in LO amplitude yields minimal increase in conversion gain (as shown in the figure, where increasing the swing by a further 10% yielded a minute improvement of 0.1 dB in conversion gain).

On the other hand, the $I_{\text{IP3}}$ does not have a monotonic trend and its behaviour is divided into two regions by a critical LO swing. Below a critical LO swing, increasing amplitude of the LO increases the magnitude of IM3 and decreases the $I_{\text{IP3}}$. For the passive mixer shown in Fig. 2.23, the critical LO swing is about 0.272VDD and the corresponding $I_{\text{IP3}}$ is -2.35 dBm. Beyond this critical LO swing, increasing the amplitude of LO decreases the passive mixer's IM3 and increases the $I_{\text{IP3}}$. A larger LO swing increases the time within each LO period where each of the four transistors are either completed on or off. In either state, the four transistors behave as short-circuit or open-circuit generating minimal or no distortion. Consequently, the passive mixer exhibits low distortion at high LO swing. For instance, the $I_{\text{IP3}}$ at $V_{\text{LO,pk-pk}} = V_{\text{DD}}$ and $V_{\text{LO,pk-pk}} = 1.1V_{\text{DD}}$ are +10.30 dBm and +10.95 dBm, respectively.

Unfortunately, the available LO swing is often limited by other design constraints (e.g., maximum allowed power dissipation of LO buffer, allowble LO feedthrough and radiation without additional shielding or filtering), resulting in suboptimal linearity. Nevertheless, the time where the four transistors reside either in the complete on or off states per single LO period can also be increased by raising the slew rate of the driving voltage wave. This can be accomplished by using a square wave LO to drive the four transistors of a passive mixer instead of a sinusoidal LO. Since an ideal square wave

\( n = \infty \) in Equation 2-32 does not exist, a square wave modeled by up to the 5th odd-harmonics of the fundamental signal (\( n = 9 \) in Equation 2-32) was used for simulation purposes. The conversion gain and $I_{\text{IP3}}$ when the same passive mixer is driven by this square wave were simulated and also included in Fig. 2.23 for comparison. As seen in the figure, the simulated conversion gain and $I_{\text{IP3}}$ for a square wave LO resemble those obtained for a sinusoidal LO. The simulation results also show that the conversion gain and linearity are improved when the slew rate of the LO increases. For example, for the same peak-to-peak swing of $V_{\text{DD}}$, the square wave LO improves the conversion gain and linearity by 0.48 dB and 6.24 dB, respectively. Similarly, the time per LO period where all four transistors within the passive mixer are either complete on or off can also be reduced by increasing the aspect ratio of the four transistors. This reduces the required effective voltage (or LO swing) needed to turn the individual transistors on and off.

Although increasing the slew rate of LO and the aspect ratio of the four transistors can provide significant improvement to the $I_{\text{IP3}}$ of the passive mixer, both approaches become limited when the operating frequency is in the GHz (or higher) range. At low frequency, gate terminals of the four transistors appear as open-circuit and do not load the LO buffer. However, the gate terminals begin to look like a capacitor connected to ground at higher frequency as parasitic capacitors become to dominate. The effects from parasitic capacitors are further highlighted when LO with high slew rate is required, since it is often difficult to realize all the higher odd harmonics. Consequently, there is often a compromise over the maximum allowable aspect ratio of the four transistors and power dissipation of the LO buffer when a passive mixer intended to be used within a single-chip transceiver design.

### 2.6 Summary

This chapter presented background information on how dynamic range and BER of a wireless receiver are affected by its sensitivity and selectivity performance. Material presented in this chapter emphasized on selectivity since the BER for short and medium range wireless data transfer in a heavily network environment is often limited by selectivity instead of sensitivity. Limitation in selectivity...
Background information

Chapter 3

Distortion in Active Mixers

In the previous chapter, the function of the mixer within a radio receiver was discussed and different types of mixers were presented. It was noted that the upper limit of the overall dynamic range of a receiver is determined by the distortion of the first downconverting mixer, leading to research activities aimed at developing a more linear mixer. This chapter will focus on the distortion generated by an active mixer, more specifically the Gilbert mixing quad, since this topology and its derivatives are the most commonly used in integrated radios. The schematic diagram for the Gilbert mixer was shown in Fig. 2.21. It consists of three basic stages: input transconductor, switching quad, and output load. The total distortion present at the output of the mixer is the sum of distortion from all three stages. However, it will be seen that the distortion of the mixer is usually dominated by one of the three stages (depending on the operating condition of the mixer).

In this chapter, the nonlinear behavior of each stage within the Gilbert mixer will be described, beginning with the switching quad stage, followed by the input transconductance, and ending with the output stage. The discussion, aided by low frequency simulated results, will provide an insight into the nonlinear behavior of each stage and the influence of circuit parameters (e.g., bias current) on the distortion it produces. Common linearization techniques from the survey of the literature will also be presented and discussed. Again, low frequency analysis will be used as illustration on how these linearization techniques work to reduce the distortion. Finally, an example of a CMOS Gilbert mixer linearized by inductive degeneration is presented at the end of this chapter.

3.1. Switching Stage

In an ideal Gilbert mixer, a differential large amplitude square-wave LO drives the switching quad, turning the four transistors on and off as quickly as possible. This steers the differential output current of the transconductance stage between two pairs of transistors and produces the desired frequency conversion. In practice, the transistors do not turn on/off instantaneously and imperfect switching reduces the overall conversion gain and increases the noise figure. The common practice is to minimize the time length during which all four transistors are conducting current by using as large an LO voltage swing as possible. Although a larger LO is desired for improving conversion gain and minimizing noise figure, an excessively large LO can actually generate distortion that can dominate the overall linearity of both BJT and CMOS variants of the Gilbert mixer [50][51].

Fig. 3.1 shows a schematic of a Gilbert mixer that can be used to analyze the distortion generated by the switching quad. Although a BJT is presented here, similar behavior can also be expected from a CMOS switching quad [51]. Iq and Cq in the schematic represent the bias current and decoupling capacitor, respectively. An ideal transconductor with finite output impedance is used to model the emitter-coupled pair (ECP). For simplicity, the LO is a differential sinusoid and the source impedances at both RF and LO ports are 50Ω. Since the input and output signals of the switching quad have different frequencies, the distortion generated by the switching quad cannot be quantified or modeled by Taylor or Volterra series. However, the distortion can be quantified by IP3 obtained from a two-tone test calculated by a computer simulator (e.g., via Cadence SpectreRF™). The ratio between the fundamental and IM3 components at the differential IF output current is used as the AM/AM for the IP3 calculation. For this set-up, distortion observed in the differential IF output current will be solely due to the switching quad, since the input transconductance stage is ideal and distortionless. The transconductance of the voltage-controlled current sources are chosen to be 20mS, which is equivalent to an ECP biased at 1mA. Terminating resistors are also added to the differential RF ports. Simulation
Distortion in active mixer

Figure 3.1: Schematic of a Gilbert modulator with an ideal transconductance input stage for analyzing the distortion of the switching quad.

The primary sources of distortion within the switching quad are the base-resistance and base-emitter capacitance [50]. Since these two parameters are not zero in a real BJT, they generate distortion, resulting in finite IIP3. The influence of LO amplitude at the base terminals on the overall simulated IIP3 for various bias currents is shown in Fig. 3.2. As the amplitude of the LO increases from a very small value (i.e., < 1mV-pk), the transistors (Q3-Q6) are turned on and off with ever increasing speed. This reduces the time per LO cycle where all four transistors are conducting, increasing the ∆MO at the IF current output and the overall IIP3 of the mixer. However, the IIP3 then peaks and decreases rapidly with increasing LO swing before levelling off at an asymptotic level. For a fixed transistor size, decreasing the bias current reduces the LO amplitude where the peak IIP3 occurs. The graphs also show that higher linearity is attained by increasing the bias current, similar to the ECP that will be presented in Section 3.2.

This peaking in IIP3 has been observed in both BJT and CMOS switching quad variants [50][51], and was attributed to the finite impedances between the common-emitter nodes (node X and Y in Fig. 3.1) and ground [51]. Finite impedances at these two nodes result in the generation of time-varying voltages (having a frequency of 2fLO), which are proportional to the LO amplitude applied to the base terminals [51]. These voltage swings at the common nodes cause currents to flow through the switching quad that generate a time-varying distortion at the IF differential current output.

At DC, the parasitic capacitances appear as an open-circuit and the transconductance stages have high output impedance (e.g., 10kΩ or larger for a BJT) and the common-emitter nodes of the switching quad behave as virtual grounds. However, the impedance seen at the common-emitter nodes decreases with increasing frequency and can affect the distortion generated by the switching quad. Its impact can be studied by varying only the output impedance of the input transconductance stage (rout). This reduces the simulation time since the RF and LO frequencies are still relatively low in frequency (RF at 100MHz instead of GHz). The simulated IIP3 of the mixer for three different values of rout are shown in Fig. 3.3. These graphs show that when there is minimal loading at the common-emitter nodes (e.g., the 10kΩ case in Fig. 3.3), the ECPs of the switching quad operates like a current commutating stage. Increasing the LO amplitude increases the switching speed, reduces the distortion contribution from the switching stage, and improves the IIP3 monotonically. However, as the LO amplitude is increased further, the IIP3 then levels off. This saturation behaviour is due to the non-zero base-emitter capacitances of Q3-6. These capacitances, together with the input RF signal, generate non-zero voltages across the base-emitter terminals of Q3-6 and generate distortion components at their collector terminals.

However, when the impedance shunting the common-emitter nodes decreases, increasing the LO swing above ~30mVpk actually decreases the IIP3 of the mixer. With higher LO amplitude, excessive currents are pumped into the common-emitter nodes (node X and Y in Fig. 3.1) through the parasitic
Distortion in active mixer

base-emitter capacitance, leading to a decrease in linearity [50]. For example, the IIP3 decreases by 0.4dB and 1.5dB when the LO amplitude is doubled from 150mVp^ to 300mVp for r^out = 1KΩ and r^out = 0.1KΩ respectively. The simulated results also showed that the optimal LO swing for IIP3 peaking decreases as the load on the common-emitter nodes increases.

\[ \text{frf1 = 100MHz frf2 = 101MHz flo = 90MHz} \]
\[ \text{lB = 2mA IM/Ib = 0.01 Emitter length/width = 4μm/0.1μm} \]
\[ r^\text{out:} \quad 0.1KΩ \quad 1.0KΩ \quad 10KΩ \]

Figure 3.3: Effects of LO swing on linearity of switching quad for fixed bias current and varying r^out.

The influence of base resistance and base-emitter depletion capacitance on the distortion generated by the switching quad is studied by comparing the simulated IIP3 of the mixer for various transistor sizes (see Fig. 3.4). Similar to the dependency on bias current, the IIP3 graphs also follow a general trend where they first increase with increasing LO, peak sharply, and then approach some asymptotic values. However, unlike the dependency on bias current, the LO amplitude where IIP3 peaks decreases with increasing emitter length. Simulation results show that a lower base resistance (e.g., by increasing the emitter length) will generally reduce distortion for a fixed LO swing. Linearity of the switching quad improves with a smaller base resistance as the voltage drop across the resistors decreases, maximizing the switching speed. For example, keeping LO amplitude constant at 100mVp^ while increasing the emitter length by a factor of eight reduces the base resistance, increases the switching speed, and increases the IIP3 by 6.7dB. However, increasing the transistor size also increases the depletion capacitance across the base-emitter of Q3_6, where it will lead to lower linearity. This decrease in linearity comes about as larger base-emitter parasitic capacitors will increase the currents that are pumped into the common-emitter nodes by the LO signals at the base terminals. Furthermore, larger parasitic capacitance will also shunt the common-emitter nodes to ground, similar to r^out in Fig. 3.1 having a small value, leading to an increase in distortion at the output of the switching quad.

\[ \text{frf1 = 100MHz frf2 = 101MHz flo = 90MHz} \]
\[ \text{lB = 2mA r^out = 1KΩ IM/Ib = 0.01 Emitter length/width = 4μm/0.18μm} \]
\[ \text{Emitter length:} \quad 1μm \quad 2μm \quad 4μm \quad 8μm \]

Figure 3.4: Effects of LO swing on linearity of switching quad for fixed bias current for various emitter length.

The Gilbert mixer is often driven with large LO amplitude such that only one transistor within each BJT pair is active. The active transistors behave like cascode amplifiers where their base terminals are small-signal ground. Subsequently, the parasitic base-emitter capacitors become in parallel with r^out reduces the total output impedance of the transconductor, and lead to higher distortion (as shown in Fig. 3.3). Finally, the increase in parasitic capacitance at the base and collector terminals of the transistors in the quad will also load the LO and lower the IF bandwidth. Usually, distortion due to the base resistance is much smaller than that generated by the base-emitter depletion capacitance [52].

Although the discussions and simulation results presented in Fig. 3.2 to Fig. 3.4 focus on distortion, LO swing, bias current, and transistor size also affect the noise figure of the mixer. For example, a smaller LO amplitude increases the time per LO cycle when all four transistors conduct current and contribute to the overall noise, leading to an increase in the total noise at the IF terminals. Similarly, the noise figure can also be degraded by using smaller transistor or biasing the switching quad with larger bias current [52]. In the former scenario, shortening the emitter length of a BJT increases the base resistance resulting in a corresponding increase in thermal noise. In the latter scenario, increase the bias current will increase the shot noise and flicker noise contribution of transistors Q3_6. From Fig. 2.3, the base shot noise and flicker noise are represented by \(2qI_b/\sqrt{f_c} \) and \(2qK_b/b/\sqrt{f_c} \), respectively, when referred to the collector terminal. These two noise sources increase with bias current \(I_b\) and directly corrupt the desired signal. Unfortunately, the conditions for the switching quad to exhibit minimal noise and distortion are often contradictory (e.g., larger bias current decreases
Distortion in active mixer

distortion but increases noise figure). Consequently, the final design choice for LO swing, bias current, and transistor sizes is typically a compromise between noise figure and distortion [52].

The discussion presented in this subsection showed that although the transistors within the switching quad are designed to turn on and off quickly like switches, they are not distortionless (i.e., finite IIP3). Furthermore, the distortion introduced by the switching quad can be larger than the contribution of the input transconductor. It is typically the limiting factor in the design of a Gilbert mixer intended for GHz frequency operation [51]. A feedforward linearization technique that addresses the distortion generated by a switching quad will be presented in Chapter 5.

3.2. Transconductor (input) Stage

As described in Section 2-5-1, the input transconductance stage of the classical Gilbert modulator is usually implemented using a differential pair (e.g., an emitter-coupled transistor pair, or ECP). The ECP converts an input voltage into a current output that can then be frequency translated by the switching quad. Due to the nonlinear behaviour of the transistor, distortions appear in the output current of the ECP. The input voltage and output current of the ECP in the low frequency regime are related by the \( \tanh \) function (as given by Equation 2-24) and their relationship is approximately linear for very small input voltages (i.e., \( |V_{in}| \ll V_T \)). The influence of this nonlinear behaviour on the overall distortion of the mixer can be studied by replacing the switching quad in the Gilbert mixer by ideal switches and observing the distortion generated at the IF current output \( (i_{RF}, -i_{RF}) \), as shown in Fig. 3.5.

These switches turn on/off instantaneously. The idealized switches produce an output which is equivalent to multiplying the differential output current of the ECP \( (i_{Q1} - i_{Q2}) \) by an ideal square wave (as in Equation 2-32, which showed that downconversion introduces no distortion). Consequently, distortion observed at the IF is due solely to the input stage. This is carried out by quantifying its IIP3 and is determined by two-tone transient simulations using Cadence SpectreRF™.

Since the output is a differential current, the ratio between the fundamental and IM3 components observed in the current at the IF (i.e., \( I_{IF} \)) is defined as \( \text{AIM3} \) and is then used to determine the IIP3. Since IIP3 can vary widely when there is mismatch between the source and input impedances, each RF port of the mixer shown in Fig. 3.5 was resistively terminated with 50 \( \Omega \). The simulated IIP3 and AIM3 at low frequency (downconverting 100MHz RF signal to 10MHz IF) for a range of bias current are shown in Fig. 3.6. SiGe BJTs with a maximum of 120GHz are used in the ECP [26]. For each bias current, the IIP3 is extrapolated from the condition where magnitude of the small-signal linear fundamental response \( (i_{RF}) \) is equal to 0.025 \( I_b \) or \( \omega/\omega_b = 0.025 \). To ensure that the distortion components generated in the output current are primarily due to weakly nonlinear behaviour.

Fig. 3.6 shows the simulation results where linearity of the ECP increases with increasing bias current. Raising the bias current of mixer \( (i_b) \) from 0.5mA to 8mA increases the IIP3 from -10dBm to -1.3dBm. The 10dB improvement in IIP3 is due to changes to the input power delivered to the ECP and AIM3, these two components that are used to determine IIP3 (Equation 2-9). The input power delivered to the ECP and AIM3 increased by 6.8dB and 6.4dB, respectively, when the bias current was raised from 0.5mA to 8mA. The input power delivered had to be increased since the transconductance did not increase linearly with bias current, as suggested by Equation 2-24. In simulation, quadrupling the bias current from 1mA to 4mA increased the transconductance of Q1 and Q2 by only 3.6 times (from 17.74mA/V to 63.86mA/V). Although Equation 2-15 suggests that AIM3 is solely dependent on the ratio of \( \omega/\omega_b \), simulation data in Fig. 3.6 indicates otherwise. To maintain a constant ratio of \( \omega/\omega_b \) for increasing \( I_b \) necessitates a corresponding increase in the amplitudes of the sinusoidal signals applied to the base of Q1 and Q2. As input amplitude increases (\( V_{in} \) in Equation 2-24), higher-order distortion components (e.g., 5th and 7th) become more significant, affecting the AIM3.

As shown by the simulated results in Fig. 3.6, +10dBm or higher in IIP3 can only be attained by increasing the bias current and the overall power consumption, which is undesirable. The ensuing discussion in this section will concentrate on different linearization techniques that has been demonstrated and applied to the input stage. As the nonlinear controlled current sources in a transistor are the primary contributor to the overall nonlinear behaviour, the impact of the different techniques
will be investigated at a low frequency (i.e., ignoring reactance elements) thus providing an intuitive understanding. Examples illustrating how these techniques can be adapted for high frequency operation will also be provided. The discussion will encompass both circuit and system levels approaches to realize an input stage with higher linearity than the simple ECP. The linearization techniques provided in the following sections are applicable in reducing the distortion of source-coupled MOS pair, because the dominant distortion in CMOS transistor is also due to its nonlinear voltage-controlled current source.

Although not the focus of this chapter, it should be noted that the input transconductance stage should also have high gain and low noise. These requirements are necessary to suppress noise from the switching quad and subsequent stages in the receiver chain (e.g., IF amplifier). Often, linearization applied to the input stage implies a sort of trade-off between linearity and noise performance.

### 3.2-1. Smoothing the Transconductance

At low frequency, an input transconductor is distortion-free when the second-order and higher coefficients in the Taylor series describing its DC transfer characteristic are equal to zero. In other words, a distortion-free transconductor has a derivative for the DC transfer characteristic that is a non-zero constant for all values of input voltage. Unfortunately, the first derivative of ECP transfer characteristic is related to the input voltage by the \( \text{sech}^2 \) function, where it is maximal for zero input voltage and decreases with increasing input voltage. The effect of this nonlinearity is particularly pronounced in a Gilbert mixer when the magnitude of \( F_{in} \) is larger than \( 4V_f \). For example, the output current is saturated when \( V_{out}/V_f > 0 \).

The value of \( F_{in} \) is limited to a small magnitude of \( F_{in} \) (much less than \( 4V_f \)). The minimum value of \( F_{in} \) reduces the effective voltage that appears across the base-emitter terminals, increasing the linear range by approximately \( 1V_{BE} > 0 \). For example, the output current saturates when \( V_{out}/V_f > 0 \).

Besides increasing the linear range before the output current saturates, feedback also reduces the sensitivity of the first derivative to change in \( V_{out} \). The desensitization process is shown in Fig. 3.8a, where the first derivatives \( g_{1}(F_{in}) \) of the DC transfer characteristic for the ECP with three values of negative feedback are illustrated. The graphs show that the product \( B_{FE} \) decreases, the input voltage range where \( g_{1}(F_{in}) = g_{1}(F_{in} = 0) \) also increases, damping the variation of transconductance to the magnitude of the input voltage. This results in a transconductance that has a more horizontal profile (i.e., more flat about \( F_{in} = 0 \)) and as result, the magnitude of its successive derivatives will be smaller. For example, the effect of emitter-degeneration on the 3\(^{rd}\)-order derivative of the DC transfer characteristic (i.e., \( g_3 \)) is highlighted in Fig. 3.8b for the same three values of \( B_{FE} \).

The degeneration impedance can also be realized using reactive elements such as the capacitor and inductor. Resistive degeneration usually leads to an increase in noise figure as the resistors introduce thermal noise that is proportional to the resistance value. Furthermore, resistors will also consume voltage headroom limiting its application in modern IC design where the supply voltage is about 1V. Emitter degeneration can also be realized using inductor, which introduces minimal DC voltage drop and thermal noise. However, on-chip inductors are physically large on-chip, raising the overall silicon area and cost. The emitter-degeneration can also be realized using capacitors, however, this is generally avoided as capacitive degeneration can increase the distortion [24]. Capacitive degeneration can also lead
to instability as the impedance between the base terminal and ground are modified by the beta-transform \([53]\) to become \((1 + \beta(\omega))Z_g\) where \(\beta(\omega)\) is the current gain and \(Z_g\) is the capacitor's reactance. The negative resistance seen from the base terminal can result in common-mode oscillation. Other negative feedback configurations, such as shunt-shunt configuration and global feedback, can also be used to increase the linearity of the ECP.

Although negative feedback is an effective method to reduce nonlinear behaviour, it does have an acute deficiency. For a fixed bias current, reduction in \(|v^2/2f_v|\) is accompanied by a reduction in transconductance \((g_{m}(V_{re}=0)\) and higher noise. Consequently, linearization via negative feedback often requires a trade-off between desired gain and acceptable distortion. The benefits of negative feedback as a linearization technique is greatest when the ECP has large open-loop gain. Unfortunately, the open-loop gain of ECP at high frequency (i.e., GHz or higher) is limited, due to the influence of parasitic capacitance, lowering the capability of negative feedback for linearization. Although the decline in gain can be compensated by increasing the bias current, this leads to higher overall power consumption.

Other than negative feedback, other circuit topologies also reshape the transconductance for higher linearity. The multi-tanh cell, where multiple ECPs operate in parallel is often found in the recent literature \([54] [55]\). The basic principle of the multi-tanh is to design individual ECPs to be linear over different input ranges, such that when summed the overall response is linear over a larger input range than a single ECP. Fig. 3.9 shows the schematic of a two-stage multi-tanh transconductance stage. Instead of biasing the two ECPs to be symmetrical about \(V_{re}=0\), the two stages are biased with the same current and offset by using asymmetric transistors pair (different emitter-area) within each ECP. This forces the two stages to be centered at

\[
V_{re,eff} = \pm V_{th} \log(4) \tag{3-1}
\]

where \(A\) is the ratio between the two emitter areas.

The simulated transconductance of a two stage multi-tanh input stage, with the ratio of emitter areas (\(A\) in Equation 3-1) of 3.75 that is chosen for maximal flat \(g_m\) \([55]\), is shown in Fig. 3.10a. The transconductance of the individual parallel ECP \((Q_{1a},Q_{2a} & Q_{1b},Q_{2b})\) and an ECP biased with the same total bias current are also included in the figure. The individual stage within the multi-tanh input stage is centered at 12.35mV \((-0.5V_{f})\) offset on either side of \(V_{re}=0\). The maximal for the overall transconductance occurs at \(V_{re}=0\) and is 15.16mA/V. This is approximately one-fifth of the transconductance for the single ECP biased at the same total bias current. Although the large signal behaviour of the two-stage multi-tanh and the ECP are almost identical (both transconductance graphs are almost identical and decrease to zero for \(V_{re}/V_f>3\)), the multi-tanh generates significantly less weakly nonlinear distortion. This is illustrated by comparing the third-order derivative of the DC transfer characteristics of both the multi-tanh and the ECP (see Fig. 3.10b). The peak values for the multi-tanh is approximately 0.385 times smaller in magnitude than the ECP at \(V_{re}=0\), indicating that the IM3 and HD3 observed at the current output of the multi-tanh will be approximately one-half that generated by the ECP (assuming that the effects of reactive elements are negligible). The linear range and weakly nonlinear behaviour of the two-stage multi-tanh can be improved further by increasing the number of parallel stages.

Although the distortion requirement of the transconductor can be satisfied and improved upon by increasing the number of parallel stages, this approach is limited for circuits that has to operate in GHz range. This limitation is the direct consequence that each additional parallel stage introduce parasitic capacitance at the input and output terminals, thereby lowering the maximum bandwidth. For RF applications, three stages are typically used in order to limit the capacitive loading at the output terminals \([56]\).

The two linearization techniques described thus far, emitter-degeneration and multi-tanh, strive to increase the linear range of the DC transfer characteristic of the ECP by adding extra circuit elements. These linearization techniques are effective in reducing the IM3 and HD3 as they reduce the magnitude of the 3rd-order coefficients of Taylor series expansion for the effective output current. Therefore, this suggests that the input transconductor stage can be made linear by replacing the ECP by another transconductor that has a Taylor series expansion with reduced magnitude in its higher-order coefficients. One such transconductor is the Class-AB stage \([57]\) shown in Fig 3.6a, where its DC transfer characteristic can be described by the hyperbolic sine (sinh) function.

\[
\sinh(x) = \frac{e^x - e^{-x}}{2}
\]

Figure 3.9: A two-stage multi-tanh transconductor using asymmetrical emitter

Figure 3.10: DC transfer characteristic: (a) first derivative and (b) third derivative of a two-stage multi-tanh transconductor.
Distortion in active mixer

Figure 3.11: A single-ended to differential sinh transconductor: (a) basic topology, (b) with degeneration, and (c) with IM3 cancellation.

Transistors Q1 and Q2 are configured as common-base and common-emitter stages, respectively. These two stages generate the in-phase \( (Q_1) \) and anti-phase \( (Q_2) \) output currents at their respective collector terminals. A diode-connected transistor \( Q_3 \) is used as a current mirror to ensure transistors \( Q_1 \) and \( Q_2 \) have the same bias current and transconductance. For small signal input voltage, the differential current output of this transconductor can be expressed as

\[
I_{out} = I_{out}^- - I_{out}^+ = I_c \left[ e^{\frac{V_{out}}{V_T}} - e^{-\frac{V_{out}}{V_T}} \right] = 2I_c \sinh \left( \frac{V_{out}}{V_T} \right)
\]

where \( I_c \) is the bias current of \( Q_1 \) and \( Q_2 \).

The differential output current can also be represented by substituting the \( \sinh \) function by its Taylor series expansion, which for a small-signal output is

\[
I_{out} = 2I_c \left[ \frac{V_{out}}{V_T} + \frac{1}{6} \left( \frac{V_{out}}{V_T} \right)^3 + \frac{1}{120} \left( \frac{V_{out}}{V_T} \right)^5 + \ldots \right]
\]

A comparison of the Taylor series expansion for this Class-AB transconductor to the ECP (Equation 2-24) shows that the ratio of third-order to the fundamental coefficients of this transconductor is twice that of the ECP for the same current consumption. The linearity performance of this Class-AB transconductor can be improved further by introducing local feedback. One possible enhancement is the MICROMixer (see Fig. 3.11b), where emitter-degeneration \( (R_E) \) is used to further reduce the magnitude of the higher-order coefficients [57]. Degeneration resistor \( (R_E) \) is used here for bias current matching. A cascode transistor can also be added the collector of \( Q_2 \) in Fig. 3.11b for a better matching of the bias currents between \( Q_1 \) and \( Q_2 \) [57]. Another possible modification to the basic Class-AB configuration is shown in Fig. 3.11c, where selecting \( R_E = V_c/(2I_c) \) will result in a simultaneous third-order coefficient cancellation and 50-Ohm input match [58].

One major disadvantage of using the combination of common-base and common-emitter stages to realize a linear transconductor is the compromised even-order linearity. This is especially true as the operating frequency increases. Unlike the ECP, the two single-ended current outputs, \( I_{out}^- \) and \( I_{out}^+ \), are affected by different parasitic capacitances. Only the anti-phase current is affected by the collector-base capacitance \( (C_{c3}) \) of \( Q_2 \), generating a mismatch in gain and phase between the current outputs, resulting in a common-mode signal. This mismatch can be compensated for by using two Class-AB transconductors driven differentially at the cost of double the bias current.

A transconductor stage that has all higher-order coefficients equal to zero is realized with only passive elements thereby circumventing the nonlinearities introduced by the transistors. Fig. 3.12 shows an example of passive input stage that is realized using an on-chip transformer [59]. The transformer \( (T_1) \) is used to convert a single-end input from the primary side into a pair of differential output currents and a pair of differential output voltages on the secondary side. Tuning capacitors \( (C_{1,3}) \) are added at both primary and secondary of \( T_1 \) for maximum gain. Besides being distortionless, the turns ratio of the transformer can also be used as a design variable for impedance matching at both primary and secondary sides. Although the passive transconductor introduces minimal distortion, it suffers from two drawbacks. Firstly, a transformer usually occupies a large silicon area (typically 200x200um² or larger) thereby increases the overall chip cost. Secondly, the transformer transconductor is narrowband in nature. Increasing the passband for wideband application would reduce the gain resulting in an increase in noise.

3-2-2. Out-of-Band Termination, Pre-distortion, and Feedforward Cancellation

Distortion that appears at the output of the input transconductor stage can also be reduced using other means such as out-of-band termination, pre-distortion, and feedforward linearization. Unlike the linearization techniques presented previously, these approaches focus on exploiting distortion cancellation instead of reducing the magnitude of the higher-order coefficients of the Taylor series.

Out-of-band termination is a technique that exploits mixing of the first-order and second-order responses via the second-order nonlinearity to cancel IM3 distortion. This technique is extensively used in single-ended and differentially driven common-emitter amplifiers, where impedances at low frequency (i.e., very close to DC) and even-order harmonics connected to the input and output
Distortion in active mixer

terminals can affect the IM3. This is evident from the expression for IM3 of a common-emitter amplifier (given in Equation 2-21) as

$$|IM3| < \left| \frac{H_i(\omega_1)}{H_i(\omega_2)} \right| \left[ 1 + H_i(\Delta\omega) \left[ 1 + 2C_{EE}Z_d(\Delta\omega) \right] \right]$$

where $H_i(\omega)$ is given by Equation 2-17 and $Z_d(\omega)$ is the sum of the base and emitter impedances. $\Delta\omega$ and $2\omega$ are the difference frequency (e.g., $|\omega_1\pm\omega_2|$) and harmonic frequency (e.g., $j2\omega_1 + j2\omega_2 + j2\omega - 2\omega$). The "-1" term in the Equation 3-4 represents the IM3 contributed by the third-order nonlinearity of the transconductor ($g_3$ in a Taylor series) and is the dominant contributor. The second and third terms in Equation 3-4 represent second-order mixing between the first-order and second-order response via the second-order nonlinearity ($g_2$ in a Taylor series) of the transconductor, and typically $H_i(\Delta\omega)H_i(2\omega)$ [24]. This out-of-band termination linearization technique came about by observing that proper selection of $Z_d(\Delta\omega)$ and $Z_d(2\omega)$ can yield a real number from the second and third terms in Equation 3-4 and cancel with the "-1" term, resulting in lower IM3.

An example of the out-of-band termination technique implemented to reduce the distortion of a common-emitter amplifier is shown in Fig. 3.13 [24]. The bias circuit is designed such that it presents a low impedance around DC, increasing the second-order interaction ($H_i(\Delta\omega)$) on IM3. This impedance value is chosen in conjunction with the emitter-degeneration such that the first two terms in Equation 3-4 cancel each other, resulting in lower IM3 and higher linearity. Further developed examples of this out-of-band termination concept can also be found in [60] and [61] where the contribution from $H_i(2\omega)$ is accounted for.

The out-of-band termination scheme has two deficiencies: bandwidth limitation and asymmetric IMD tones. For the out-of-band termination to be effective, impedances $Z_d(\Delta\omega)$ and $Z_d(2\omega)$ must be kept constant over the range where interferers could potentially be present. For example, the output impedance for the bias circuit shown in Fig. 3.13 has the desired value only for frequencies close to DC.

![Bias Circuit](image)

Figure 3.13: An inductive degenerated common-emitter with out-of-band termination for high linearity performance

---

Nonlinear behaviour of a circuit can also be compensated for by cascading nonlinear building blocks (e.g., amplifiers) together. If the nonlinear characteristics of the two building blocks are precisely complementary, the output will be distortion-free. In pre-distortion linearization, which is an example of this method, a circuit that has the inverse nonlinear relationship is placed before the input of the desired nonlinear amplifier. This concept is illustrated in Fig. 3.14a where two nonlinear amplifiers with complementary nonlinear behaviours (expansion and compression) are cascaded together to achieve a distortion-free output. Conversely, a circuit with the inverse nonlinear relationship can be cascaded at the output of the desired nonlinear amplifier. This is known as post-distortion linearization. The pre-distortion stage can be realized as either analog circuitry or using digital signal processing. Although pre/post-distortion is mainly used in RF power amplifiers, this approach can be realized at the

![Figure 3.14](image)

**Figure 3.14:** Pre-distortion linearization: (a) block diagram illustration and (b) BJT example.
circuit level. Fig. 3.14b shows an example of analog pre-distortion where the ECP is preceded by two diode-connected transistors \((Q_3, Q_4)\). The diode-connected transistors have a natural logarithm (i.e., compressive) input current to output voltage relationship, where the differential voltages \(V_{\text{in}}\) and \(V_{\text{out}}\) are related to the input current by the function \(V_{\text{in}} = f_{\text{in}}(I_{\text{in}})\). This is the inverse of the exponential current-voltage relationship of the ECP. Cascading the diode-connected stage and the ECP as shown in the figure can result in a linear output [64].

Although pre-distortion appear to be an attractive solution for high linearity, it does have several drawbacks. Firstly, a pre-distortion stage with the exact opposite nonlinear behaviour usually does not exist (in analog case). Secondly, pre-distortion will not compensate for changes in distortion due to aging as the pre-distortion stage is often calibrated for a specific nonlinear behaviour. Thirdly, fidelity of the desired signal can be compromised by noise. In the analog pre-distortion example (shown in Fig. 3.14b), the desired signal that flows through the diode-connected transistor can be at or below the noise floor. Finally, gain of the system will also be modified with pre-distortion. In the example shown in Fig. 3.14a, the input-output transfer function exhibits less small-signal gain.

Feedforward compensation is another wideband technique that is commonly applied in the design of power amplifier to achieve a distortion-free output. The underlying principle of this technique is to isolate and cancel the distortion generated by a nonlinear amplifier. The basic building blocks that illustrate the feedforward compensation operation is shown in Fig. 3.15. As shown in the figure, the input signal is first split into a distortion and a reference path. The main amplifier amplifies the input signal with a gain \(A\) but it also generates distortion in the distortion path. This signal is then sampled and scaled by an attenuator (i.e., by \(1/A\) in Fig. 3.15), generating a signal that contains the original input signal along with attenuated distortion components. At the distortion isolation node, this scaled signal is summed destructively with a time-delayed version of the input signal, resulting in an output signal (error signal) that contains only the attenuated distortion components of the main amplifier. This error signal is then amplified to raise the amplitudes of the distortion components to equal those at the output of the main amplifier. The amplified error signal is then summed destructively with the time-delayed signal from the distortion isolation node. Ideally, this combination remove all distortion components generated by the main amplifier yielding a distortion-free output. Furthermore, since feedforward linearization does not use global feedback loop, the linearization technique is a wideband technique that is inherently stable.

**Figure 3.15: Block diagram of a typical feedforward linearization** [65].

Despite its tremendous linearization ability, feedforward compensation can be rather complex and is not truly distortion-free in practice due to a number of factors. One of the most critical factors is gain and phase matching between the signals that are fed into the distortion isolation and distortion cancellation nodes. A gain or phase mismatch larger than 0.15dB and 1°, respectively, will limit the distortion suppression to 35dB [65]. Another factor that could affect the effectiveness of the feedforward compensation is the linearity performance of the error amplifier. The discussion presented so far assumed that the error amplifier is distortion-free and only amplifies the error signal. Unfortunately, the error amplifier also has its own nonlinear characteristic and will introduce distortion that will not be present in the final output. Typically, the error amplifier should generate several orders of magnitude less distortion than the main amplifier in order for feedforward linearization to be effective [65]. Lastly, although feedforward linearization is wideband and does not use global feedback, it is still limited by the individual building blocks' bandwidth and other feedback loops that created by parasitic capacitances.

**Figure 3.16: Schematic for a Cascode Compensation (CasComp)** [66].

As a practical example of the feedforward method, cascode compensation (CasComp) removes the distortion generated by a differential cascode amplifier by using an error amplifier [66]. The schematic of the CasComp is shown in Fig. 3.16. In cascode amplifier \((Q_3-Q_4)\), most of the distortion is generated across the base-emitter junctions of the input pair \(Q_1-Q_2\) and then replicated across the base-emitter junctions of cascode transistors \((Q_3-Q_4)\). The distortion is then sensed and amplified by the transistor pair \(Q_5-Q_6\). Distortion compensation is realized by adding an anti-phase compensating current, via cross-coupling the output of \(Q_3\) to \(Q_4\) and the output of \(Q_5\) to \(Q_6\) and from the error amplifier to the collector currents of \(Q_3-Q_4\). Cancellation of the original distortion occurs when the transconductances of both cascode and error amplifiers are equal [66]. Since only local feedback is used, the CasComp had demonstrated the capability to operate up to approximately one-fifth of the transistor max \(f_T\) [66]. Sensitivity to processing variations and component mismatch in a CasComp stage is relatively low, because it depends upon the ratio of the amplifier gains and not on absolute component values.
3.3. Output Stage

A classical set-up for output stage of the Gilbert mixer is to attach resistive loads to the outputs of the switching quad ($Z_L$ in Fig. 2.21). The resistive loads convert the current outputs into differential voltages, and the resistance value can be chosen for the desired voltage/power gain. This current-to-voltage conversion is typically used since most baseband circuits (e.g., IF amplifier) that follow the voltageages, and the resistance value can be chosen for the desired vohage/power gain. This cuiTent-to-
to the swhching quad ($Z_L$ in Fig. 2.21). The resistive loads convert the current outputs into differential 2). Generally, this is avoided by choosing a load resistance ($R_{L,\text{opt}}$) for maximal allowable voltage and mixer are voltage driven.

where $V_{\text{in}}$ is set by $V_{\text{ce,QB}}$, $V_{\text{ce,Q1}}$, and the saturation of $Q_3$. Consequently, the optimal resistance is $R_{L,\text{opt}} = \left(\frac{V_{\text{CC}} - V_{\text{min}}}{L}\right)^2$ and choosing a resistance value larger or smaller than this optimal will result in waveform clipping in the current or voltage domain, respectively. Since the load resistance is proportional to the overall conversion gain, the final resistance value is often a trade-off between the gain and linearity.

A different approach to convert the output current of the switching quad into voltage is to use a transimpedance amplifier load (see Fig. 3.18). For illustration, only the half-circuit is shown. The virtual grounds prevent any voltage swings, while the current output ($I_{\text{out}}$) is forced to flow through the feedback resistor ($R_F$) to generate an output voltage ($V_{\text{out}}$) at the output. Provided the loop gain is sufficiently large, the transimpedance amplifier does not introduce additional distortion to the output voltage. Unlike the resistive load output stage, the feedback resistor ($R_F$) can be used to set the desired conversion gain and avoiding voltage waveform clipping (within the current driving capability of the operational amplifier). For a downconverting mixer, a filtering capacitor ($C_p$ in Fig. 3.18) is added at the input of the transimpedance amplifier to shunt all higher harmonics that falls outside the bandwidth of the operational amplifier (OpAmp). The removal of the higher-order harmonics is necessary as the plus term ($\omega_{RF} + \omega_{LO}$) can cause waveform clipping at the output, and generate intermodulation distortion that will lowers the linearity of the mixer (Section 2-3-2). The effects of higher-order harmonics at the IF output on the overall linearity will be highlighted in Section 3.4 using a simulation example. The data show that the filtering capacitor ($C_p$ in Fig. 3.18) attenuates higher frequency components thereby improving the IIP3 by more than 20dB.

Furthermore, the virtual grounds introduced by the transimpedance reduce the distortion caused by drain-source voltage swing (i.e., by DIBL introduced in Section 2.3) in a CMOS implementation.

3.4. A Reference Mixer Example

As a reference circuit, a 1.8V double-balanced mixer utilizing local feedback for lower distortion was implemented in a 0.18μm CMOS process. This reference mixer illustrates the limitation of this linearization technique. It will also be shown that distortion observed at the IF is caused by both the input transconductance stage and the switching quad, as outlined in the previous sections of this chapter and in [50][51]. The mixer schematic is shown in Fig. 3.19. The input transconductor consists of two transistors ($M_1$ and $M_2$) configured as differentially driven common-source amplifiers (instead of ECP in a classical Gilbert mixer). Two source degeneration inductors ($L_1$ and $L_2$) are used to reduce the nonlinear behaviour of transistors $M_1$ and $M_2$. The switching quad consists of four transistors ($M_3$-$M_6$) operating as switches. With the mixer designed as a frequency downconverter, two filtering capacitors ($C_p$) are connected at the IF output to remove the upconverted signal and all high frequency harmonics.

Starting with some initial calculated values, the aspect ratios for all the transistors are determined after several iterations of simulation. All transistors have the minimum gate length (0.18μm) and the final values for transistor widths are shown in Fig. 3.19. The single symmetric degeneration inductor has a value of 2.7nH when driven differentially [67]. A microphotograph of the mixer is shown in Fig. 3.20. The filtering capacitors ($C_p$) are implemented by using multiple parallel metal layers available in the process. For experimental measurements, two external Bias-Ts (RF and $C_p$) in Fig. 3.19) are used at the IF terminals. This allows the drain terminals of transistors $M_3$-$M_6$ to be biased with the highest allowed supply voltage (1.8V).
Distortion in active mixer

Two 47Ω-transimpedance amplifiers are connected at the IF terminals to facilitate conversion gain and IIP3 measurements. The transimpedance amplifiers drive a differential 100Ω load (from a 2-to-1 balun) while presenting low impedance (~0Ω at IF) to the drain terminals. As described in Section 3.3, this setup ensures that the drain voltage is biased at the highest allowed DC voltage with minimal AC

time needed for simulation while allowing the use of a simple lowpass filter networks at the IF terminals to attenuate the summed components. The lowpass filters were realized using 4.08pF capacitors shunting the IF terminals ($C_p$ at the input of the transimpedance amplifier as shown in Fig. 3.18). The two on-chip capacitors ($C_{onChip}$ in Fig. 3.19) were excluded from the simulation. These capacitors have an effective impedance of 8.1Ω at the $\omega_{pp} + \omega_{LO}$ frequency, effectively attenuate the summed components. In simulation setup, the transimpedance amplifiers were realized by two operational amplifiers with a unity gain frequency of 3GHz. For both scenarios, the power level of the fundamental and IM3 components observed at the IF output were simulated for various input power level. For comparison purposes, the results for the two scenarios are plotted together in Fig. 3.21.

The simulation results clearly show that shunting capacitors at the IF terminals improve the linearity (IIP3) of the mixer. The lowpass filter reduces the voltage swing at the drain terminals of transistors $M_{3,6}$, thereby minimizing the distortion generated by DIBL from transistor $M_{3,6}$. As a result, the IM3 components at the IF output were lowered by over 35dB, thereby increasing the IIP3 from
Figures 3.22 and 3.23 demonstrate the measured and simulated conversion gain and mismatch losses, as well as the IIP3 and SSB noise figure. The graphs illustrate the performance of the mixer across different frequency bands, showing the improvement achieved by the input shunting capacitors. The measured conversion gain is higher than the simulated results, especially at lower frequencies, which can be attributed to lower parasitic capacitance in the experimental samples. The discrepancies in IIP3 and SSB noise figure are attributed to process variation and the simulator's estimation of distortion. The dissertation will explore further methods to improve IIP3 via feedforward linearization in subsequent chapters.
end of the frequency band, the noise figure increases to about 18dB as the inductive degeneration reduces the conversion gain. For comparison, the simulated noise figure at 15MHz is also shown in Fig. 3.23. Up to 4.5GHz, the measured and simulated noise figure differ by less than 2dB (except at 3.25GHz where the measured noise figure is 3dB higher). Beyond 4.5GHz, the difference between simulated and measured results increases. The simulated and measured noise figures are 25.8dB and 17.4dB, respectively, at 6GHz.

Overall, performance of the reference double-balanced CMOS mixer shows that inductive degeneration suppresses distortion only when the feedback factor is high. Unfortunately, this reduces the conversion gain and increases the noise figure of the mixer.

3.5. Summary

This chapter described the distortion generated in an active mixer (Gilbert mixer) by the input transconductor, switching quad, and output load stages. A review of techniques that either suppress or cancel distortion were also presented, along with their limitations. A summary of how the various linearization techniques operate along with their major deficiencies are listed in Table 3-1. The measured performance of a reference CMOS mixer was also presented. The next two chapters of this dissertation will focus on applying feedforward to the input transconductance and switching quad stages for distortion suppression.

<table>
<thead>
<tr>
<th>Stages</th>
<th>Linearization techniques</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transconductor</td>
<td>Negative feedback</td>
<td>Increases linearity at the expense of gain. Limited capability at GHz operation due to lower available gain.</td>
</tr>
<tr>
<td></td>
<td>Multi-tanh</td>
<td>Multiple stages that are linear over different input voltage amplitude that work in parallel. Each additional stage introduce addition parasitic capacitances that will reduce the operating bandwidth.</td>
</tr>
<tr>
<td></td>
<td>sinh input stage</td>
<td>A topology where the 3rd-order coefficients of its Taylor series have smaller magnitude than the ECP. Generation of even-order distortion components.</td>
</tr>
<tr>
<td>Out-of-band termination</td>
<td>Transformer input stage</td>
<td>Inherently distortion free. Bandwidth limitation and typically occupies a large silicon area.</td>
</tr>
<tr>
<td></td>
<td>Predistortion</td>
<td>Make use of the mixing of the 1st-order and 2nd-order responses via 2nd-order nonlinearity to cancel 3rd-order distortion. Suppression of distortion generated by interferers that are sufficiently wide apart in frequency can be limited.</td>
</tr>
<tr>
<td></td>
<td>Feedforward linearization</td>
<td>Cascade two amplifiers with complementary nonlinear behaviour. Two nonlinear amplifiers with complementary nonlinear behaviour usually does not exist and the overall gain is typically lower. Broadband technique that can isolate and cancel distortion components to offer linear output. Distortion suppression is sensitive to matching of magnitude and phase.</td>
</tr>
<tr>
<td>Output stage</td>
<td>Class A biasing</td>
<td>Select a resistive load for simultaneous maximal voltage and current swings to avoid voltage/current clipping. The optimal resistance might not yield the desired gain.</td>
</tr>
<tr>
<td></td>
<td>Transimpedance amplifier</td>
<td>Utilizes an active circuit with selectable transimpedance to convert IF current output from mixer to a voltage/power output. Bandwidth of the IF output is limited by the transimpedance amplifier.</td>
</tr>
</tbody>
</table>
Chapter 4

Gilbert Multiplier with Feedforward-Compensated Input Stage

Distortion generated by the widely used Gilbert multiplier cell was examined in the previous chapter. The total distortion at the output terminals contributed by the three basic stages within the mixer (i.e., input transconductor, switching quad, and output load) was analyzed. In this chapter, feedforward-compensation will be used to demonstrate lower overall distortion for the mixer by linearizing the input stage. Feedforward-compensation was chosen as it offers the capability of providing a linear output over a multi-GHz frequency range.

The first half of this chapter describes a bipolar transconductor consisting of 2 amplifiers working in parallel to achieve higher linearity. This transconductor utilizes various local feedback factors (e.g., choosing different values of transistor transconductance and resistive degeneration) to realize feedforward distortion cancellation (also called feedforward-compensation in this thesis). A design example outlining the selection of circuit parameters will be provided. Simulation results show that feedforward-compensation offers improvement in distortion suppression compared to the linearization techniques detailed in the previous chapter. In the second half of this chapter, a CMOS equivalent transconductor (i.e., same circuit topology as the feedforward-compensated bipolar) will be presented. The CMOS transconductor is designed to operate between one and six gigahertz and incorporated as the input stage of a Gilbert mixer. Simulation and measured results of this CMOS mixer will be presented and discussed.

4.1. BJT Transconductor

The CasComp topology presented in Section 3-2-2 is a transconductor that provides excellent distortion suppression. Even-order distortion is rejected by employing a differential structure, while the third-order distortion is reduced via feedforward linearization. This is accomplished by using an error amplifier to sense and replicate the distortion components of the main amplifier. The replicated distortion is then added destructively at the output of the main amplifier, resulting in higher linearity. However, the phase shift in the signal paths of the main and error amplifiers differ, and this imperfection increases with operating frequency. Consequently, the operating bandwidth is limited by the speed of the stages. It will be seen that the third-order distortion is no longer cancelled by the feedforward addition of currents (summing of collector currents from Q3 and Q4 in Fig. 3.16) as operating frequencies approach one-fifth of the transistor frequency.

Fig. 4.1 shows a transconductor that offers wider operating bandwidth, generates less distortion than the CasComp, and offers a low input impedance required by RF and many high-speed applications. The transconductor, which is called feedforward-compensated differential pair in this thesis, consists of two differentially-excited common-base transistors (Q1-Q2) and emitter-coupled pair (Q2-Q3). Distortion compensation of the weakly nonlinear behaviour of the transistors is achieved by using emitter degeneration to modify the phase and amplitude of the third-order distortion products produced by each amplifier stage. This is possible because the phase relationship between fundamental and distortion components (i.e., either in-phase or anti-phase) produced by a degenerated common-base or common-emitter transistor depends upon the product of emitter degeneration resistance and transistor transconductance. Consequently, the differential common-base and common-emitter amplifiers of Fig. 4.1 are designed such that the fundamental components of their respective output currents add in-phase, while the third-order distortion components (both harmonic and intermodulation products) are anti-phase and cancel each other.
Gilbert modulator with feedforward-compensated input stage

Figure 4.1: Feedforward-Compensated Differential-Pair.

Compared to classical feedforward linearization techniques (e.g., CasComp), the transconductor shown in Fig. 4.1 differs in how the third-order distortion components and the compensating components are generated. In the feedforward-compensated differential pair, both the distortion and compensating components are developed concurrently (i.e., by both amplifiers). In a CasComp stage, the main amplifier generates the undesired third-order distortion components that are then isolated and used to generate the compensating components.

4-1-1. Analysis of the Feedforward-compensated Differential Pair

Various combinations of bias currents and resistors for distortion cancellation can be determined analytically by performing a harmonic balance analysis [68] and using the superposition principle. A harmonic balance analysis is first used to predict the distortion generated at the outputs of each of the common-base and common-emitter amplifiers alone. This involves exciting the two amplifiers by a sinusoidal input source that contains only one fundamental frequency component or

$$V_{in}(t) = V_{in} \cos(\omega t)$$  \hspace{1cm} (4-1)

Due to the inherent nonlinearities of the transistors, each amplifier generates a differential collector current

$$i_C(t) = c_1 \cos(\omega t) + c_2 \cos(2\omega t) + c_3 \cos(3\omega t) + ...$$  \hspace{1cm} (4-2)

encompassing all harmonics that are generated in the collector current. Coefficients for the power series are first determined by summing the voltages at the input using Kirchhoff’s voltage law, followed by the substitution of the nonlinear base-emitter junction voltage by a power series expansion. The individual collector currents are then combined in accordance with the superposition principle to determine the condition for distortion cancellation.

Using one half-circuit of the differential-excited common-base transistors in Fig. 4.1 as an example, Kirchhoff’s voltage law applied to the input loop of Q3 gives

$$V_{in} = -I_g R_{b3} - R_{b3}$$  \hspace{1cm} (4-3)

or

$$V_{in} = R_{b3} \ln(1 + c_3 I_g)$$  \hspace{1cm} (4-4)

Replacing the natural log (ln) function in Equation 4-4 by its series expansion, and substituting Equation 4-4 into Equation 4-3, results in an expression that relates the input voltage to the emitter/collector current ($i_C(t)$). After substituting and replacing the input voltage and emitter current in Equation 4-3 with the expressions from Equation 4-1 and Equation 4-2, coefficients $c_i$ are determined from a harmonic balance analysis. The coefficients for the first three harmonics of common-base and common-emitter differential currents are listed in Table 4-1. For sufficiently small input (e.g., $v_1 = V_p$), higher-order harmonics are ignored since $c_3$ is usually the dominant distortion contributor to the third-order distortion.

Table 4-1: Coefficients for the Differential Currents.

<table>
<thead>
<tr>
<th>Harmonic Coefficient</th>
<th>$i_{Out,CE}$</th>
<th>$i_{Out,CB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>$\frac{g_m}{1 + R_g g_m}$</td>
<td>$\frac{g_m}{1 + R_g g_m}$</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$c_3$</td>
<td>$\frac{g_m}{48g_m^2(1 + R_g g_m)^3}$</td>
<td>$\frac{g_m}{96g_m^2(1 + R_g g_m)^3}$</td>
</tr>
</tbody>
</table>

When differential common-base and emitter-coupled stages are configured as in Fig. 4.1, the total output current can be written as

$$i_{Out}(t) = i_{Out,CE} + i_{Out,CB}$$  \hspace{1cm} (4-5)

Given the signed coefficients of Table 4-1, third-order distortion cancellation via phase and amplitudes matching is now possible. For,

$$R_g g_m \lesssim 0.5$$  \hspace{1cm} (4-6)

the third-order distortion coefficient of the common-base pair is 180° anti-phase to that of the emitter-coupled pair. This condition can be realized through proper selection of $R_g$ and bias current $I_g$ (which sets $g_m$). Amplitude matching between the third-order coefficients is then realized by selecting $R_A$ and bias current $I_A$ satisfying

$$\frac{2I_g}{1 + R_A g_m^2} \approx \frac{I_g (1 + 2R_g g_m^3)}{1 + R_g g_m^2}$$  \hspace{1cm} (4-7)

$R_A$ and $R_B$ in Equation 4-6 and Equation 4-7 are the total emitter degeneration resistances. This represents the sum of the intrinsic resistance of the transistor and the resistance connected to the emitter terminal of the transistor.
A sample calculation for the resistor and bias current selections is outlined in the following example. Firstly, the ratio of bias currents between the common-base and common-emitter stages is chosen to be 2, with a total bias current of 5mA assumed. This results in \( I_\text{E} \) and \( I_\text{B} \) having values of 1.875mA and 0.625mA, respectively. The product of \( R_3 \) and \( g_{m3} \) is then chosen at 0.25. This value is selected to accommodate situations in which there may be large variations in resistance due to processing variations. With the value for \( I_\text{E} \) pre-determined, a value of 10Ω for \( R_3 \) is chosen. Accounting for the 1.7Ω intrinsic emitter resistances of the bipolar devices, amplitude matching between the third-order coefficients is ensured with \( R_\text{E} \) of 20.81Ω from Equation 4-7.

Although the analysis given in this section considers only a single-tone excitation and ignores all parasitic capacitances, the analysis is still valid and provides valuable insight in reducing the combined third-order intermodulation distortion (IM3) of the output current and linearity (IIP3). For weakly nonlinear behaviour (i.e., close to small-signal operation), the HD3 and IM3 caused by single- and two-tone excitations are proportional to each other [68]. Namely,

\[
IM_j = k \cdot HD_j,
\]

or the \( k \)th intermodulation product is \( k \)-times larger than the corresponding harmonic distortion product. Consequently, cancellation of HD3 by the circuit will also lead to a reduction of IM3 at the output. The parasitic capacitances were ignored in the analysis for simplicity since they are distortion-free (as described in Section 2-3-3-3). Their effects on the overall distortion behaviour in the GHz range will be studied through simulation.

4-1-2. Simulation Results for the Feedforward-compensated Differential Pair

Simulations of the proposed circuit were performed using the Spectre simulator for a 0.5μm SiGe process with a transistor \( f_T \) of 47GHz [69]. Values for the transistor bias currents and resistors in the circuit simulation are identical to those determined in the previous section. For simplicity, all NPN transistors in Fig. 4.1 were chosen to be 20μm × 0.5μm. The suitability of this transconductor for RF applications was then verified by quantifying its performance at the 2.4GHz frequency band (e.g., applicable to IEEE802.11b/g WLAN).

The simulated frequency response (normalized to the transconductance at 1MHz) of the feedforward-compensated differential pair is shown in Fig. 4.2. The CasComp frequency response was also simulated and included in this figure for comparison. Both the bias current and feedback resistor values for the CasComp were chosen appropriately (e.g., choosing 40Ω/3.3mA for \( R_\text{E} \) and 10Ω/1.67mA for \( R_\text{E} \) in Fig. 3.16) as suggested in [66]. Simulation results show that the 3dB bandwidth of the feedforward-compensated differential pair is 38GHz, which is an improvement of 24GHz over the equivalent CasComp.

Fig. 4.3a shows the simulated differential input impedance (single pole, parallel equivalent resistor and capacitor). The equivalent shunting resistor is 53.4Ω at 0.1GHz. The resistance decreases with frequency and is 38.9Ω at 1GHz. The equivalent shunting capacitor also has the similar frequency behaviour and varies between 395fF and 480fF. This result gives an input return loss better than -10dB for frequencies up to 5.25GHz, offering broadband (50Ω) input matching. The input impedance is 45.4-j16.1Ω at 2.4GHz, corresponding to a return loss of -15.3dB. Minimum noise figure (NFmin) and 50Ω noise figure (NF50) were also simulated for a 100Ω differential output load, and are plotted in Fig. 4.3b. The noise figure of the feedforward-compensated differential-pair is much higher than for an emitter-coupled pair (which can have NF of ~1dB). This dominant sources of noise in the feedforward-compensated differential-pair are the common-base amplifier and the degeneration resistors. Between 0.1GHz and 10GHz the difference between NFmin and NF50 is very small (~0.8dB at 10GHz). This indicates that a simultaneous power and noise match can be achieved since the optimum source resistance for minimum noise (giving NFmin) is close to 50Ω. The noise figure at 2.4GHz is 3.49dB.

Effectiveness of the distortion cancellation was verified by quantifying the THD and IM3 behaviour of the feedforward-compensated differential pair when it is driven by 2.4GHz input tone(s). Harmonics up to the ninth order were used to determine the THD (using Equation 2-7), and the amplitudes of individual harmonics were determined from discrete Fourier transform (DFT) computed from a time-domain simulation. The simulated THD performance for input amplitude swings of 1mVpk to 100mVpk are plotted in Fig. 4.4. For comparison, the THD performance for the CasComp along with three other transconductors (i.e., resistive degenerated differential-pair, two-stage multi-tanh, and Class-AB transconductor) were also simulated and included in the figure. All transconductors draw the same bias current (5mA). Design parameters for the multi-tanh (\( A = 4 \) in Fig. 3.9), the degenerated differential-pair (9.45Ω/5mA for \( R_{\text{EE}/\text{IE}} \) in Fig. 3.7), and the Class-AB transconductor (\( R_\text{E} = 5\Omega \) in Fig. 3.11) were chosen for optimal linearity as outlined in Section 3.2.

As seen in the figure, the feedforward-compensated differential pair not only achieves, but maintains THD performance of 0.055% (or less) up to an input amplitude of 100mVpk (equivalent to \(-4\text{V}_\text{pk}\)). Furthermore, the simulation results also show that the feedforward-compensated differential pair generates significantly less distortion than the other transconductor (e.g., at 25mVpk, the THD is 0.003% compared to the 0.232% of the degenerated diff-pair). This superior performance is highlighted further when the input sinusoid has a larger amplitude, where the distortion cancellation mechanism excels and yields a current output that clearly has significantly less distortion. For example, it produces less than one-tenth the distortion of a comparable Class-AB stage (0.055% versus 1.392%) at 100mVpk.
The intermodulation distortion behaviour of the feedforward-compensated differential pair was also simulated in order to determine its linearity (IIP3). Two tones at 2.40 and 2.41 GHz were input to the circuit, and DFT results of the power dissipated across a differential 100Ω load used to determine fundamental and IM3 products. The simulated results are shown in Fig. 4.5. The simulated IIP3 and input referred P1dB at 2.4GHz are +18.0dBm and -5.5dB, respectively. For small inputs (up to -30dBm), the IM3 increases by 3dB for every 1dB increase in input power. However, the IM3 increases sharply at about P1dB = -25dBm before levelling off. The sharp increase in IM3 results from higher-order intermodulation distortion (e.g., IM5), and distortion generated by clipping at the amplifier output. Unlike the distortion caused by the third-order intermodulation, distortion caused by clipping adds in-phase, resulting in an increase in IM3.

As the distortion cancellation is directly dependent on accurate matching of phase and magnitude of both distortion components, the effects of processing variations were also simulated. A change of ±10% from the nominal resistor values were used for these test cases. The two worse cases for the feedforward-compensated differential pair along with the nominal results for the other transconductor are summarized in Table 4-2.
Gilbert modulator with feedforward-compensated input stage

\[ f_1 = 2.40 \text{GHz} \quad f_2 = 2.41 \text{GHz} \quad \text{IIP}_3 = 18.0 \text{dBm} \]

Figure 4.5: Simulated output vs. input power for fundamental and IM3.

**4-1-3. Discussion**

These simulation results indicate that the feedforward-compensated differential pair would provide a more linear transconductor for the input stage to an RF mixer. The linearized pair exhibits superior linearity and wider operating bandwidth compared to other transconductors. It is also robust to process variations as summarized in [70]. As mentioned earlier in this section, the noise figure of the feedforward-compensated differential pair is relatively high. However, poorer sensitivity can be acceptable when the proposed input transconductor is used as the input stage for a mixer or IF amplifier where the emphasis is on linearity rather than noise. Finally, the noise figure of the feedforward-compensated differential pair could also be improved by using the technique suggested in [71]. The noise compensation technique utilizes a common-gate and a common-source amplifiers, and is similar to the feedforward-compensated differential pair presented in this subsection. The two amplifiers of the noise compensation technique operate in parallel and are biased such that the noise generated by the individual amplifier compensates the noise generated by the other, reducing the overall noise at the output terminal.

**4.2. CMOS Transconductor**

The transconductor presented in Section 4.1 can be modified into an all-CMOS design by replacing the emitter-coupled pair and common-base stages in Fig. 4.1 with a source-coupled pair and a pair of common-gate transistors, respectively. Similar to the BJT case, local feedback loops are used to manipulate the output currents by appropriate selection of degeneration resistors, bias currents, and MOS transistor aspect ratios. With proper selection, the fundamental tones of the output currents are summed constructively while the third-order distortion tones cancel each other, thereby improving the composite linearity of the transconductor.

### 4-2-1. Design of the CMOS Version

Fig. 4.6 shows the schematic of a double-balanced mixer with a feedforward-compensated differential pair input stage. The input stage consists of parallel amplifiers, that is, a pair of common-source \((M_{1,2})\) amplifying stages and a pair of common-gate \((M_{3,4})\) amplifying stages. The common-source stages are used instead of a source-coupled pair to permit DC biasing for low voltage operation. Switching quad \((M_{5,6})\) cascades the input stage to downconvert the differential output current of the transconductor from RF to IF. These IF output currents are then converted to a voltage by driving a resistive load (not shown in the figure). Similar to the BJT transconductor design, resistive feedback loops (via \(R_{CS}\) and \(R_{CG}\)) modify the transistor output currents so that the fundamental components remain in-phase, while the IM3 components are equal in magnitude but anti-phase. When the two sets of output currents are summed (by connecting the drain terminals of \(M_1\) to \(M_3\), and \(M_2\) to \(M_4\) as shown in Fig. 4.6), the fundamental components add and the distortion components cancel, yielding a "distortion-free" RF current at the input to the mixing quad for downconverting.

It should be noted that although the final schematic shown in Fig. 4.6 is similar to the superposition derivative method [72], the final circuit topology was derived from the CasComp. A circuit that

![Figure 4.6: Schematic of a CMOS mixer with feedforward-compensated input stage.](image)
employs superposition derivative linearization usually consists of two or more parallel common-source amplifiers (with different widths and gate biases), where each amplifier generates distortion with different magnitude and phase. Summing the outputs from these amplifiers will result in cancellation for the distortion components and consequently, higher linearity. In the proposed topology, series local feedbacks are used to reduce sensitivity of distortion matching to process variation (see Table 4-2) and a differentially-excited common-gate amplifier is used to facilitate input matching.

The mixer was designed to downconvert the RF signal between 1 and 6GHz in a 0.18μm CMOS technology. During normal operation, the mixer is powered from a 1.8V voltage supply. Each of the common-source and common-gate transistors is biased with 0.75mA current. The width of transistors M1,4 and values for resistors R_CG and R_CS were determined through simulation for optimal IM3 cancellation (at nodes A and B in Fig. 4.6). Their respective values in microns and in Ohms are shown beside each component. Transistors M1,4 have a minimum gate length (0.18um), while M9,10 have a gate length of 0.26um (for higher output impedance). At the switching quad, a sinusoidal LO with amplitude of 375mVpk is applied to each gate. A width of 45um (aspect ratio of 250) was chosen for transistors M1,4. The aspect ratio is a compromise between the desire for higher switching speed and the parasitic capacitance of the gate terminals. At the IF output, filtering capacitors (C_P) are added to attenuate high frequency components which can cause clipping and desensitization [28]. Distortion caused by DIBL is kept at a minimum by biasing the drains of M5,8 at the highest possible voltage (1.8V). This is accomplished by using chokes (RFC) and coupling capacitors (C_C). Two 8Ω transimpedance amplifiers are connected to the differential IF output terminals (IF^+ and IF^-). The transimpedance amplifiers also minimize voltage swing at the drain terminals and distortion caused by DIBL from transistors M5,8.

Parasitic capacitances in both the common-gate and common-source amplifying paths cause magnitude and phase variations to both fundamental and IM3 components. Effects of these parasitics on the magnitude and phase matching between the two parallel amplifying stages for various RF input frequencies were simulated and are plotted in Fig. 4.7. The figure shows that the ratio between the common-gate and common-source amplifier currents at node A and B in Fig. 4.6, which is defined as \( \frac{I_{out}}{I_{comp}} \), varies from -16.63° to 11.16°. These imperfections will ultimately limit the linearity of the input stage and (potentially) the mixer. For example, distortion suppression is limited to 35dB or lower when the gain and phase mismatch are larger than 0.15dB and 1°, respectively. The parameters shown in Fig. 4.6 were optimized for an RF of 2.4GHz and the IM3 tones from the two amplifying stages are equal in magnitude, while there is a phase error between output IM3 currents of 6.96° (\( \angle I_{IM3} \)) and decrease with increasing input (i.e., RF) frequency. The conversion gains at 1 and 6GHz were simulated and are plotted in Fig. 4.7. The figure shows that the ratio between the common-gate and common-source output currents at the fundamental and IM3 frequencies.

As expected, parasitics cause the conversion gain to peak at the lowest frequency (1GHz in Fig. 4.8) and decrease with increasing input (i.e., RF) frequency. The conversion gains at 1 and 6GHz are 3.99dB and 2.13dB, respectively. On the other hand, simulation results show that the IIP3 is -9.99dBm at 1GHz and increases with increasing RF, reaching a peak of +12.14dBm at 1.25GHz. Raising the operating frequency further, IM3 components with larger magnitude are observed at the output of mixer, leading to a drop in IIP3 and cumulating in a minimum IIP3 of +1.37dBm at 3.6GHz. The simulator predicts that any further increase in input frequency will result in a decline in IM3 magnitude and corresponding increase in IIP3. This trend continues until 5.6GHz, where a local maximum IIP3 of +7.16dBm was obtained in simulation. Moreover, as the RF increases from 5.6GHz to 6.0GHz, the IIP3 decreases and is +5.88dBm at 6.0GHz. The maximum and minimum IIP3 occur at
Gilbert modulator with feedforward-compensated input stage

Figure 4.8: Simulated conversion gain and IIP3 of the mixer in Fig. 4.6 along with the IIP3 of the input stage

between different frequencies for the input transconductor and the complete mixer. Between 1-6GHz, IIP3 of the input stage varies between +6.22dBm (minimum at 4.75GHz) and +14.20dBm (maximum at 1.80GHz). Comparing the IIP3 of the input stage and the mixer in Fig. 4.8 shows that distortion is introduced during the downconverting process (except for the RF range between 1.0-1.3GHz and 5.5-5.75GHz). This drop in linearity is especially noticeable between 1.60GHz and 4.70GHz where the overall IIP3 decreases by at least 3dB.

Fig. 4.9a shows the simulated noise figure (SSB, 50Ω) versus IF for three different input (RF) frequencies (i.e., 1.0, 3.5, and 6.0GHz). Simulation predicts that flicker noise will be present for IFs up to 10MHz and the flicker noise corners for RF input frequency of 1.0GHz and 6.0GHz are 2.8MHz and 14.9MHz, respectively. At higher IF, flicker noise contributions become less significant as thermal noise dominates (as described in Section 2.2) for IFs of 100MHz or higher (from simulations, as shown in Fig. 4.9a). Noise figure of the mixer for RF input frequency of 1.0 and 6.0GHz is 80.0dB and 8.76dB, respectively. As the operating frequency increases from 1 to 6GHz, the conversion gain decreases (from the effects of parasitic capacitances) thereby compromising the noise figure. The decline in noise figure is greatest at IFs where the flicker noise dominates. This is readily seen from the plots of the noise figure of mixer versus RF for fixed IF shown in Fig. 4.9b.

All three plots share the same frequency profile, where the noise figure is lowest when the RF input is 1GHz (lowest) and highest at when the RF is 6GHz. When the mixer is configured to downconvert to an IF of 15MHz, the noise figure is 11.79dB (for an RF of 1GHz) and increases to 16.56dB when the RF is 6GHz. As shown in the figure, changing the IF changes the contribution of flicker noise and the overall noise figure. For example, increasing the IF to 30MHz helps to reduce the noise figure at 1GHz and 6GHz to 11.08dB and 15.50dB, respectively. On the other hand, reducing the IF to 10MHz will increase the noise figures at 1GHz and 6GHz to 12.41dB and 17.58dB, respectively. Although not implemented in the design, flicker noise present at the IF terminals can be reduced by incorporating baseband processing techniques such as correlated double sampling and chopper stabilization [73], at the expense of complexity, in the baseband signal path. The former is a sample-and-hold approach where an amplifier samples the unwanted flicker noise and then subtracts the flicker noise from the instantaneous value. Effectively, the flicker noise is high-pass filtered and significantly reduced at low frequencies. The latter approach is a continuous time baseband processing technique where it first upconverts the desired IF signal to a frequency above the flicker noise, amplifies the desired IF signal, and then downconvert it back to the baseband.

4-2-2. Measured Results for the Mixer with Linearized Transconductor

Fig. 4.10a shows the die photo of the double-balanced mixer prototype which measures 0.97x1.10mm2 (including bondpads). The fabricated chip encompasses all elements shown in Fig. 4.6 except for the chokes and coupling capacitors. On-chip filtering capacitors (Cf in Fig. 4.6) were implemented using multi-layered vertical parallel plate (VPP) capacitors and are highlighted in
Gilbert modulator with feedforward-compensated input stage

Figure 4.10: (a) Photomicrograph of CMOS mixer with feedforward-compensated input stage and (b) Testboard used for evaluation.

Fig. 4.10a. These VPP capacitors were also used to bypass high frequency components at biasing lines \( V_{BI} \) and \( V_{BG} \). Common-centroid layout was also employed for the input transconductance stage to ensure good matching between the differential RF paths. For measurement purposes, the die was packaged in a 32-pin quad ceramic flatpack, which was then mounted on a custom-made printed circuit board (PCB). The PCB also incorporates passive surface mount device (SMD) for bypassing and decoupling of DC nodes (see Fig. 4.10b).

The mixer was characterized for operation in the 1-6GHz frequency band, which includes the IEEE802.11a/b/g band allocated for the high-speed wireless local-area networking. Differential RF and LO signals were derived from single-ended sources using external passive microwave power dividers and hybrid baluns. At the IF output terminals, external Bias-Ts and transimpedance amplifiers were used to bias the mixer and convert IF output current into a voltage for measurement.

The measured conversion gain, measured IIP3, and measured OIP3 between 1 and 6GHz are shown in Fig. 4.11. In the measurement setup, the mixer was configured to downconvert an RF signal to an IF of 50kHz, and two fundamental tones spaced 100kHz apart were used for the two-tone tests. Similar to the simulation results, the measured conversion gain emulates the frequency behaviour where it is maximum at low frequency, and it decreases with increasing frequency. For comparison purpose, conversion gain, IIP3, and OIP3 predicted by SpectreRF simulator are also plotted in the figure (in dash lines). The conversion gain varies by 4.76dB with the highest and lowest conversion gain measured at 1GHz (1.62dB) and 6GHz (-3.14dB), respectively. Comparing the measured and simulated conversion gains shows that the simulator provides an optimistic result. The measured data is 2.37dB lower than the prediction of the simulator at 1GHz, but this difference increases to 5.31dB at 6GHz. This may be caused by factors such as lower than expected transconductance of transistor or unaccounted for parasitic capacitance due to process variations, or slower switching speed from the switching quad.

As shown in Fig. 4.11, the IIP3 of the mixer is relatively constant between 1 to 6GHz and varies between +3.58dBm and +6.32dBm. However, unlike the frequency behaviour of the conversion gain, the measured IIP3 is concave in representation, which differs from the frequency response of simulated IIP3 (convex response). Difference in the frequency behaviours is probably caused by the imperfect anti-phase relationship between the distortion components of the common-gate and the common-source amplifiers. This error in phase shift between the distortion components is probably caused by the fact that different parasitic capacitances are present within the two amplifier stages, introducing different phase delays. Below 2GHz, measurement results indicate that the mixer exhibits significantly more distortion than what the simulator predicts. For example, the measured IIP3 at 1.5GHz is +3.65dBm while the simulated IIP3 is +9.92dBm. This suggests that the IM3 components generated by the mixer could be due to lower than expected distortion suppression and/or the switching quad generating higher than expected distortion components. As the RF increases beyond 2.0GHz, the measured IIP3 becomes...
better than the prediction of the simulator and this better than the expected behaviour continues up to an input RF of 5.0GHz. For example, the measured IIP3 at 3.5GHz is +6.23dBm and this is 4.84dB better than the value predicted by the simulator. However, as the RF increases beyond 5.0GHz, the measured IIP3 decreases and the distortion components measured at the mixer outputs were larger than that predicted by the simulator (e.g., at 5.5GHz, the measured and simulated IIP3 are +4.86dBm and +6.92dBm, respectively. The measured OIP3 of the mixer (or the sum of conversion gain and IIP3) are also plotted in Fig. 4.11. Between 1-6GHz, the OIP3 varies from +1.53 to +5.96dBm. 

Noise figure (SSB 50Ω) when the mixer downconverts various RF were also measured and the noise figure at 15MHz IF are shown in Fig. 4.12. The noise figure is lowest (11.31dB) at 1GHz and increases steadily with frequency, rising to 18.26dB at 6GHz. As mentioned earlier, parasitic capacitances, inaccuracy in transconductance prediction at higher frequency and slower switching speed of the mixing stage can all contribute to a degradation in conversion gain and increase in noise figure. At higher operation frequency, the reduction in the transconductance of the input stage and lower operating speed of the switching quad will lead to an decrease in the output SNR and increase in the overall noise figure. For comparison, the measured noise figure at an IF of 15MHz (first shown in Fig. 4.9b) is also plotted in Fig. 4.12. As shown in the figure, the measured noise figure compares favourably to the prediction of the simulator. The difference between the measured and simulated noise figures between 1-6GHz range differs up to 2.94dB (occurring at 3.00GHz where the measured and simulated noise figures are 16.13dB and 13.19dB, respectively).

The second-order distortion performance (IIP2) of the mixer was also measured and the data obtained from one sample is included in Fig. 4.12. Simulated IIP2 of the mixer is not included as the simulations did not include statistical data. Across the 1-6GHz band, the IIP2 was measured to be between +37.79dBm and +32.04dBm. Similar to the noise figure performance, the IIP2 also shows the general trend where highest performance was achieved in the lower frequency range. The even-order distortion performance can be expected to improve when the mixer is integrated with other RF building blocks, thereby, eliminating mismatch introduced by external cables and hybrid couplers used in the measurements. Furthermore, the IIP2 can also be improved using the techniques as outlined in [39][60].

4.2-3. Discussion

Simulation and measurement results presented in the previous two subsections demonstrate that the feedforward-compensated differential-pair topology (presented in Section 4.1) can also be realized in CMOS technology and used as an input stage to a Gilbert-type mixer. Although the CMOS variant offers distortion compensation over a range of input frequencies, distortion suppression is limited (as seen by the IIP3 plots in Fig. 4.8 and Fig. 4.11). This can be caused by parasitic capacitances or the generation of intermodulation distortion due to even-order mixing. The former introduces mismatch in magnitude and phase for optimal distortion cancellation. Parasitic capacitances also lower the conversion gain and increase the noise figure. Even-order distortion generated at the output of the transconductance stage can generate third-order intermodulation distortion in the switching quad via an even-order intermodulation mechanism.

Although simulation results predict good magnitude and phase matching between the IM3 components of the two parallel stages (see Fig. 4.7), measured IIP3 indicates that parasitic capacitances introduce magnitude and phase mismatch resulting in sub-optimal linearity (e.g., simulated IIP3 for the mixer operating at 1.5GHz was +9.92dBm while measured IIP3 was +3.65dBm). Between 1-6GHz, the measured IIP3 is relatively stable and within ±1.5dB of +5dBm. This differs from the simulated IIP3, where variation between ±1.37dBm and +12.14dBm were predicted. The difference between simulation and measurement can be attributed to inaccuracy in distortion modeling and the effects of parasitic capacitances. Both introduce mismatch in magnitude and phase, leading to sub-optimal linearity. Simulated and measured conversion gains possess a similar frequency response (peak gain at low frequency and decreasing with increasing RF). However, the parasitic capacitances that are present on the fabricated die appear to be higher than predicted in simulation. Consequently, conversion gain obtained in measurement was 2 to 5dB lower than in simulation. The decline in measured conversion gain result in lower SNR at the IF output port, resulting in a noise figure that is up to 3dB higher than predicted from simulation.

4.3. Summary

This chapter presented two transconduction stages where third-order distortion (both harmonic and intermodulation) at their output currents are reduced via feedforward-compensation.

The first circuit consists of an emitter-coupled pair and differentially-excited common-base stage that are driven concurrently by a balanced input voltage. Local resistive series degeneration is applied to both parallel stages, controlling the magnitude and phase of their respective third-order distortion. Therefore, third-order distortion components generated by the emitter-coupled pair and the differentially-excited common-base stage can be of the same magnitude, but anti-phase, by choosing the appropriate values for the bias currents and degeneration resistors. When the output currents are summed, the distortion components cancel, resulting in a combined output current that is distortion-free. Simulation results show that this transconductor offers improvement in harmonic distortion over other transconductors (e.g., differential-pair with series feedback). The IM3 generated at the output is also expected to be much lower (≥15dB) than other linearized transconductor. A 2.4GHz amplifier realized using this transconductor as an input stage would exhibit an IIP3 of +18dBm. Besides
Gilbert modulator with feedforward-compensated input stage

suppressing third-order distortion at the output, the transconductance stage also offers wide operating bandwidth and low input impedance as required for RF and many high-speed applications.

The second circuit is a CMOS version of the first transconductor and consists of differentially-excited common-source and common-gate stages. Similarly, local feedback (via proper selection of bias currents, transistors sizing, and degeneration resistors) was applied to the individual stages such that the IM3 components from these two stages are equal in magnitude but anti-phase. When the output currents are summed, the IM3 components combine destructively while the fundamental components combine additively, resulting in an output current with higher fundamental components and lower IM3 distortion. This transconductor was used as an input stage to a Gilbert mixer, converting the input voltage into a distortion-free output current. Measured results showed that although the distortion cancellation predicted by simulations was not achieved, the mixer still exhibits an IIP3 of between +4dB and +6dBm. Compared to the commonly used linearization techniques such as local series feedback, the feedforward-compensated differential pair offers comparable or better linearity. For example, the Gilbert mixer with inductive degeneration described in Section 3.4 had a measured IIP3 of less than +5dBm. Furthermore, measurements also indicated that the feedforward-compensated differential pair offers comparable or better spur-free dynamic range (SFDR) than the Gilbert mixer with inductive degeneration. Note that the SFDR comparison is carried out by comparing the difference between the IIP3 and noise figure or (IIP3-NF) of the two mixers. SFDR of the feedforward-compensated differential pair is 6.4 and 0.1dB at 2.5 and 5.5GHz, respectively, better than the Gilbert mixer.

This chapter also presented (simulation) results that demonstrate distortion being introduced by the switching quad stage of the mixer. Circuit topologies that compensate the distortion introduced by the switching quad stage will be described in the next chapter.

Chapter 5

Feedforward-Linearized Switching-Quad

This chapter describes a feedforward-linearized topology where two parallel mixers are used to implement distortion cancellation. This topology was used to realize two double-balanced mixers in both BJT and CMOS technologies. The BJT mixer is described in this chapter, while the CMOS variant is described in Chapter 6. In both designs, the feedforward-linearized topology was shown to suppress the distortion generated by a switching quad. The prototype bipolar mixer described in this chapter is designed to improve the linearity of a receiver downconverter for wireless applications in the 2-6GHz range, where interference arising from intermodulation distortion can potentially affect the bit-error performance of a wireless receiver. Parameters affecting the distortion observed at the output of a single BJT switching quad will first be presented and discussed. This is followed by simulation and measured results that will illustrate the capabilities of feedforward linearization technique.

5.1. Linearized Mixer Concept

In the previous chapters, feedforward linearization (i.e., CasComp) was advocated as a linear and broadband technique to reduce IM distortion and circuit sensitivity to processing variations and component mismatch. However, the feedforward scheme used to linearize an amplifier cannot be adapted directly to the switching quad (between the RF input and IF output) because of the frequency translation in the signal path between RF input and IF output. In a classical feedforward linearization scheme, distortion introduced by the nonlinear amplifier is derived by comparing signals at the input and output of the nonlinear amplifier. However, the RF input signal cannot be used to derive an estimate of the distortion at the IF output without doing another frequency translation.

This problem is avoided when the compensating signal is developed independently of the output by using a second, parallel signal path (see Fig. 5.1). As shown in the figure, the feedforward-linearized mixer utilizes two switching quads operating in parallel, where each mixer produces in-phase
Ensuing discussions in this chapter are devoted to the design and realization of various mixers to be used in a feedforward manner as shown in Fig. 5.1.

5.2. BJT Mixer Design

Typically, an LO input that is many times higher than the thermal voltage (e.g., 200-500mV-pk differential) is used to drive the switching quad of a balanced mixer (e.g., see Fig. 5.2). This reduces the period of time when all 4 transistors (i.e., Q1-Q4) are conducting, and as a result the quad behaves like a common-base amplifier between the RF (input) and IF (output) over most of the LO cycle. Therefore, one might expect intuitively that local feedback from emitter resistor R would null or cancel the IM distortion in this mixer, similar to the distortion nulling observed in a common-base stage with local feedback. This is indeed possible, however, the IM distortion generated by the quad is a function of LO amplitude, transistor bias current and the emitter area of the transistor [50].

As an example, the simulated IM3 phase (relative to the fundamental components) and IIP3 of the IF output current are plotted in Fig. 5.3 for the circuit of Fig. 5.2 as the feedback resistor R varies. The transistors in the switching quad are biased at 4mA and driven by a 235mVpk sinusoidal LO for the simulations. The results show that as the feedback increases (i.e., an increase in R), the IM3 components fall to zero (or null), thereby maximizing the mixer IIP3. For example, the peak IIP3 is +9.8dBm at an RF input frequency of 250MHz. The behavior at higher frequencies is similar, but parasitic capacitances lower the maximum IIP3 and change the conditions required to null the IM distortion (e.g., peak IIP3 occurs at 163Ω instead of 172Ω when the frequency increases from 0.25 to 5.75GHz).

The effects of the process variation and mismatch in the mixer IIP3 were analyzed using Monte Carlo analysis in Cadence Spectre™. The nominal operating conditions of $I_{\text{bias}}=3.5\,\text{mA}$, $R=75\,\Omega$, and $V_{\text{LO}}=235\,\text{mV}_{\text{pk}}$ (single-ended) for the mixer of Fig. 5.3 produced the desired distortion cancellation and an IIP3 of +7.93dBm. The IIP3 (for constant $I_{\text{bias}}$) results obtained from the Monte Carlo simulations are plotted as a histogram in Fig. 5.4. IIP3 for the linearized mixer follows a Gaussian distribution with a mean of +8.30dBm and standard deviation of 0.46dB.

The preceding discussions and simulation results showed that when the proper combination of bias current, LO swing, and degeneration (R) is chosen, distortion nulling is achieved at the output of the
mixer. However, effectiveness of the distortion nulling is easily compromised by process variation and mismatch (as suggested by the simulation results plotted in Fig. 5.4), leading to different mixer samples having different linearity performance, even if the samples are fabricated on the same wafer. More importantly though, simulation results indicated that the phases of the IM3 components and fundamentals at the IF outputs can be made in-phase or anti-phase to each other.

5.2-1. Design and Simulation

The capability to manipulate the phase of the IM3 components lend itself to realize a linear mixer by utilizing two modified switching quads in feedforward operation, similar to that shown in Fig. 5.1. A feedforward-linearized mixer that uses two switching quads operating in parallel is shown in Fig. 5.5. The same balanced RF and LO signals are used to drive both switching quads. Appropriate selection of the bias currents and degeneration resistances (e.g., $R_A$ and $R_B$ in Fig. 5.5) then creates the condition where Mixer A and Mixer B produce IM3 components with equal magnitude but anti-phase. Distortion cancellation is then realized by summing the RF currents of the individual switching quads (by shorting the collector terminals of transistors as shown in Fig. 5.5). Similarly, process variation and mismatch will also affect the distortion performance of both switching quads. However, if the two switching quads are fabricated on the same wafer, both switching quads will experience the same deviations and their IM distortion will track each other (this will be shown in forthcoming discussions).

Degeneration resistors add thermal noise to the RF input giving the mixer a relatively high noise figure. Inductors could be substituted to provide the necessary degeneration impedance because they do not (ideally) add thermal noise. However, inductive degeneration restricts the range of possible bias currents and device sizes where the phase shift between the two RF paths can be matched for IM cancellation, thereby restricting the RF input frequency range. In addition, resistors consume significantly less area than on-chip inductors. In an integrated receiver, the output impedance of the driving stage could also be used to degenerate the mixers (see Section 5.2-3). Since the emphasis in this design was proof of concept for a broadband mixer with improved linearity, degeneration resistors are used in the following design.

For validation of the feedforward-compensation concept, all transistors in the mixer of Fig. 5.5 were chosen to be $5 \times 0.2 \mu m^2$ and the total bias current limited to 10mA. Degeneration resistors were restricted to between 50Ω and 150Ω to limit the thermal noise contribution and voltage drop across them. Both mixers are (nominally) driven by a sinusoidal LO of 166V$_{pk}$ at each base terminal, which ensures that the quads switch completely using an LO that is easily sourced from an on-chip oscillator in a wireless application.

Feedforward compensation is realized by matching the IM3 magnitudes for mixers A and B. In simulation it was found that 75Ω and 120Ω for $R_A$ and $R_B$, respectively, at a total bias current of 3mA for mixer A and 5mA for mixer B set this condition. This selection of degeneration resistors and bias currents gives approximately the same (50 Ohm) input impedance for the two mixers. A differential IF load of 470Ω and a 2.2V supply were used for these simulations. The fundamental and IM3 output powers vs. the input power for the feedforward-linearized mixer as well as each mixer individually are plotted in Fig. 5.6 ($f_{RF} = 5.75$GHz and $f_{IF} = 150$MHz). There is an improvement of more than 7dB in $I_{IP3}$ for the feedforward-linearized mixer (overall $I_{IP3} = 14.6$dBm) compared to either mixer operated stand-alone. These simulations also show that the individual mixers and the feedforward-linearized mixer have approximately the same gain compression point. This is expected, because local feedback only affects distortion generated by the nonlinear transconductance of transistor. It does not affect distortion generated by waveform clipping, which is typically the source of gain compression.
Linearity and conversion gain for the feedforward-linearized mixer from a simulation over the RF input frequency range from 1-10GHz are plotted in Fig. 5.7. The conversion gain is approximately 12.3dB at 1GHz, and decreases gradually with increasing frequency to 11.7dB at 10GHz. This slight drop in conversion gain is expected as parasitic capacitance dominates the high frequency response. When operated at a lower supply voltage, voltage/current clipping at the IF output lowers the -1dB compression point. However, intermodulation distortion (dominated by nonlinearity of the active devices) remains essentially the same, so an IIP3 similar to what is achieved at higher supply voltages is expected.

Similarly, the OIP3 is highest at lower frequencies; it is +31.6dBm at 1GHz and drops to +24.3dBm at 10GHz. Simulations predict that gain and phase mismatch in IM3 components generated by the two mixers are 0.2dB and 1° at 1GHz, but increase to 1.7dB and 13° at 10GHz. Distortion cancellation is affected as parasitic capacitances play a more significant role in the response of the mixer at higher frequencies.

A single Gilbert mixer operating at 8mA bias current (3V supply) was also simulated for comparison. The same LO swing (166mVpk) and IF load (470Ω) are used. The conversion gain of the mixer is adjusted via a degeneration inductor to equal the feedforward mixer at 5.75GHz (RF). Simulation predicts that the single mixer with a 10nH degeneration inductor, has a conversion gain and IIP3 of 12.2dB and 11.56dBm ($f_{RF} = 5.75$GHz), respectively. Linearity of is still 3dB lower than the feedforward-linearized mixer at $f_{RF} = 5.75$GHz, and deteriorates further as the RF decreases below 5.75GHz. For example, at 2GHz, simulation predicts a conversion gain and IIP3 of 20.4dB and -1.7dBm, respectively, for the single mixer at 8mA with the same 10nH degeneration. On the other hand, the simulated IIP3 for the feedforward-linearized mixer is better than 12.6dBm for frequencies ranging from 1-10GHz.

Results of Monte Carlo simulation of the complete feedforward-linearized mixer are plotted in Fig. 5.8 ($I_{MixerA} = 3mA$, and $I_{MixerB} = 5mA$). Unlike the mixer with feedback distortion nulling (see Fig. 5.2), the distribution is not Gaussian but a type III generalized extreme value. The mean and standard deviation of the distribution are 14.4dBm and 0.70dBm, respectively. Over 80% of the simulated results for IIP3 fall between 14 and 16dBm. Statistical simulation suggests that although the highest IIP3 cannot be obtained for all samples, high linearity (i.e., within 1-2dB of the max.) is still achieved for the majority of the simulation cases, promising higher yield.

The second-order intermodulation distortion (IM2) of the feedforward-linearized mixer (important in low/zero-IF applications) is primarily determined by component matching and the physical layout of the circuit. Monte Carlo simulation results for the IIP2 of the mixer (refer to Fig. 5.9) are Gaussian with a mean and standard deviation of 67.15dBm and 7.48dB, respectively. A similar analysis on the circuit of Fig. 5.2 designed for IM3 nulling gave similar results with a mean IIP2 of 66.13dBm and a standard deviation of 5.94dB, respectively. Fewer components give a slightly smaller spread for a single mixer. Even-order distortion and IIP2 may be reduced further by correcting for even-order mismatches using a scheme such as that proposed in [60].

5.2-2. Measurement

The prototype of the feedforward-linearized mixer was fabricated in a SiGe BJT process [26] and characterized over the frequency range from 1 to 6GHz. This covers the spectrum currently occupied
Feedforward-linearized switching quad

by commercial wireless applications (including Groups A and B of IEEE 802.15.3a UWB). The experimental test set-up is shown in Fig. 5.10. Differential RF and LO signals are derived from single-ended generators using microwave power dividers and 0°/180° hybrid baluns. At the RF input, two differential signals are used such that the individual mixers can be biased individually. The mixer was configured as a downconverter for testing, so 5pF filtering capacitors (included on-chip) are used to filter the unwanted (upper) sideband at the IF output. The mixer IF is ac coupled to two 470Ω single-ended loads (RIF) and balun-coupled to single-ended test instruments (e.g., a spectrum analyzer). The prototype mixer including the I/O pads occupies 1.10x1.24mm² while the active circuitry of the mixer occupies just 80x180μm², as seen from the chip micrograph at the center of Fig. 5.10. The supply voltage was varied between 1.2V and 2.2V in testing, although most of the characterization work was performed using the (nominal) 2.2V supply. The measured IM3 performance is expected to improve when the mixer is used in an integrated transceiver design (i.e., RF integrated with LO and IF/baseband stages), because gain and phase mismatch in the RF, LO and IF paths inherent in the experimental setup can be minimized with an integrated solution.

The bias and LO input power that resulted in the highest linearity (ΔIM3) was first determined experimentally from a conventional two-tone test. In each case, 2.4GHz RF signals spaced 100Hz apart and with an input amplitude of -13dBm were downconverted to a 50kHz IF when the mixer was driven by a differential LO power of -1.6dBm (note: these are the powers at the mixer input ports corrected for cable losses, etc.). Close spacing of the RF input tones was used so that the tones can be resolved more accurately at the IF. The largest measured dB-difference between the fundamental and third-order intermodulation distortion products (i.e., the ΔIM3) is plotted in Fig. 5.11 for various combinations of bias current in each quad. The ordered pair of numbers next to each data point represents the bias currents flowing in mixers A and B (from Fig. 5.5), respectively.

For portable applications, both low power dissipation and high ΔIM3 is desired. The bias combination (3mA,3mA) from a 2.2V supply was therefore selected as a compromise between low intermodulation distortion and total current consumption for further characterization. At this bias current, the measured fundamental and IM3 output powers at the IF versus the RF input power are plotted in Fig. 5.12 (fIF = 50kHz and fRF = 2.4GHz). An LO input power of -2.6dBm was used for the measurements. The measured conversion gain and IIP3 are 12.6dBm and 14.3dBm, respectively. The conversion gain can be easily adjusted by modifying the IF load impedance (i.e., current limiting) and introduce distortion while high IF load impedances increases the dominant pole at the IF terminal and decreases the IF bandwidth. As seen from the transfer curves of Fig. 5.12, the IF output of the feedforward mixer is linear up to an RF input power of approximately -4dBm, while the IM3 distortion power expands for input powers larger than about -5dBm.

The measured conversion gain, IIP3 and OIP3 of the feedforward-linearized mixer in the frequency ranges from 1.25-2.4GHz and 4-6GHz are compared in Fig. 5.13 (note that frequencies between 2.5 and 4GHz could not be covered with the test set-up at that time). The measured values compare favorably with those predicted from simulation. For example, the simulated conversion gain and IIP3 at 5.7GHz are 12.7dB and 14.4dBm, respectively, compared to the measured values of 12.6dB and 13.3dBm. The conversion gain is relatively constant at approximately 12.5dB between 1 and 6GHz. The IIP3 is better than +13dBm for operating frequencies between 2.1 and 5.8GHz. The IIP3 varies by
3.5dB across the measured frequency range and is lowest at 1.25GHz (IIP3=11.7dBm). Similar performance can be expected over the entire 1-10GHz RF operating range.

Matching between the various single-ended paths in the test set-up is difficult to achieve in practice. For example, matching the amplitudes of all 4 mixer RF inputs and ensuring 180° phase difference between signals at each RF input pair over frequency, would require extensive calibration and characterization work and extensive trimming. However, the effects of amplitude mismatch, phase error and phase mismatch are easily studied in simulation. In this way, the differences observed between the measurements and the nominal simulation results may be accounted for. The measured amplitude and phase errors for the external passive components used in the experimental set-up (e.g., power splitters, 0°/180° hybrids, cables, etc.) are frequency dependent, and can cause up to ±1dB difference in amplitude, and up to ±10° deviation from the nominal phase condition. Simulations predict that such errors cause the conversion gain to vary by up to ±1dB, so the conversion gain is relatively insensitive to input amplitude and phase errors. Linearity (enhanced by IM3 feed-forward cancellation) is more sensitive to amplitude and/or phase variations at the RF inputs. The simulations show that IIP3 degrades by up to 5dB at fRF = 5GHz and up to 8dB at fRF = 2GHz. More consistent performance (i.e., similar to that predicted in Fig. 5.7) could therefore be expected when the mixer is integrated with a low-noise preamplifier and LO synthesizer on the same chip. This would reduce the amplitude variation and phase mismatch with frequency at the RF and LO inputs introduced by the external components used to test the mixer (as predicted by simulation of such variation and mismatch).

With only the mixing quads and bias sources connected between the supply rails (i.e., 2 transistors in series), the feedforward mixer is well-suited to low-voltage operation. The conversion gain and IIP3 are plotted in Fig. 5.14 when operating at 2.4GHz RF from a 1.2V (minimum) supply. Again, the bias current combination of (3mA,3mA) was used for these measurements. The variation in gain and distortion is minimal due to the decrease in supply voltage. It should also be noted that lower voltage operation (down to approximately 1V) is possible if the bias current sources were removed and the supply current regulated using the LO common-mode (i.e., bias) voltage. The IIP3 varies by just ±0.6dB as the supply voltage drops from 2.2V to 1.2V.

The even-order distortion (i.e., IIP2) was also measured for one sample at 2.4, 5.3, and 5.8GHz, where it is 54.5, 36.6, and 36.3dBm, respectively. These values are lower than predicted by simulations (i.e., 67dBm at 5.8GHz). It is likely that the difference is caused by mismatching in amplitude and phase of the signals in the RF and LO paths at the inputs to the mixer test circuit (mostly determined by mismatch in the external hybrid baluns and cables). Again, better performance could be expected from a fully integrated solution, where these sources of error are lower and are more tightly controlled.

Degradation of the distortion cancellation and IM3 performance due to minor variation in bias current and LO power was also investigated. Bias current of each mixer was varied by ±0.5mA (-15%), which resulted in a maximal drop of 3.6dB in IIP3, Similarly, the maximum change observed in IIP3 was 6.7dB for a ±2dB variation in the LO input power.

5.2-3. Discussion

The previous two subsections demonstrated that the magnitude and phase of the IM3 components at IF output of a switching quad can be varied (by the selection of LO swing, bias current, and degeneration resistor) and used in a feedforward operation to create a linear mixer. The measured and simulated results of the feedforward-linearized mixer are summarize in Table 5-1. The data indicate
that optimal linearity is achieved when higher LO power and lower bias current are used compared to simulations (i.e., 3dB lower LO power and 2mA less current). The measured conversion gain is also slightly larger than predicted by simulation (by 0.8dB). The highest linearity observed in the measurements and simulations are comparable, but occur at different LO input powers and bias currents as noted earlier, which is likely due to processing variations and imperfections in the experimental test setup (e.g., phase inaccuracy of the 0°/180° hybrids).

Two other bipolar mixers reported in the recent literature, one developed for receiver base-stations [74] and a second mixer that also uses feedforward-linearization [75], but only in the RF input stage (i.e., the input differential pair of a Gilbert mixer) are also included in Table 5-1. Linearity of the mixer developed in this work (i.e., output third-order intercept point) is comparable to the receiver basestation mixer from [74]. Comparing output intercept points (OIP3) removes the effect of conversion gain, which varies widely among the mixers listed in the table. The LT5527 mixer described in [74] uses on-chip transformers at the RF and LO inputs for input coupling and matching purposes. It is designed for use in 3G cellular telecommunication basestations, where the electronics are not portable or powered from a battery, so current consumption and supply voltage are not important design constraints. As shown in Table 5-1, the feedforward-linearized mixer offers comparable linearity (from the OIP3) but with a current consumption of 6mA for the core of the mixer compared to 78mA (total, including LO buffers) from a 5V supply for the LT5527. The feedforward-linearized design brings higher linearity to portable applications where current consumption and operating time on a fresh battery charge are important. Further power savings are possible for the feedforward design if only one of the two mixers is activated in response to less demanding operating conditions using an adaptive power management scheme.

The noise figure of the feedforward design is compromised by the use of resistive degeneration at the RF input, which adds thermal noise. However, this also allows the mixer to work over a wide range of frequencies (from baseband up to 6GHz), whereas the narrowband RF and LO inputs of the LT5527 chip restrict its operating range to between 400MHz and 3.7GHz. The total RF input impedance of the mixer is close to 50Ω (57.4-j16.6Ω at 5.8GHz) across most of the 1-6GHz band, thereby simplifying impedance matching to other components (e.g., an image-reject filter between LNA and mixer). The deleterious effects of high noise figure for the feedforward-linearized mixer are mitigated somewhat by the higher conversion gain of the active mixer. However, the feedforward mixer would require greater noise suppression from a low-noise preamplifier in order to realize a given overall receiver noise figure compared to that of the other two mixers listed in Table 5-1.

As mentioned earlier, resistive degeneration at the mixer RF inputs were used to implement feedforward cancellation of IM distortion over a wide range of frequency. However, thermal noise produced by the degeneration resistors (R_A and R_B in Fig. 5.5) is among the largest contributors to the total mixer noise. In order to improve the noise figure, these resistors should be either reduced or eliminated. One way to achieve this is to use the output impedance of the RF driving stage to degenerate the mixers. As seen in Fig. 5.15, the RF sources for the mixers may be represented by Thevenin equivalent voltage sources (V_A/V_B) with a series output resistance (R_{Out,A}/R_{Out,B}). The desired values for R_{Out,A} and R_{Out,B} can be realized in numerous ways, with one possibility at RF being the use of transformers with different turn ratios (for narrowband applications) to reflect the desired impedance levels at the mixer inputs. A driving stage having a constant output impedance over frequency (e.g., amplifier with shunt output feedback) can also be used to synthesize the desired R_{Out,A} and R_{Out,B} for broadband applications. In both these cases the noise figure is improved, as R_{Out,A} and R_{Out,B} are now included as a part of the source impedance.
Table 5-1: Performance summary and comparison

<table>
<thead>
<tr>
<th>Parameter</th>
<th>FF-linearized mixer</th>
<th>LT 5527 [74]</th>
<th>FF-gm [75]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simulated</td>
<td>Measured</td>
<td></td>
</tr>
<tr>
<td>RF input, GHz</td>
<td>2.4</td>
<td>1.9</td>
<td>0.87</td>
</tr>
<tr>
<td>IF, in Hz</td>
<td>250M</td>
<td>50k</td>
<td>240M</td>
</tr>
<tr>
<td>Conversion gain, dB (IF load resistance, Ω)</td>
<td>12.3 (470Ω)</td>
<td>12.6 (470Ω)</td>
<td>2.3 (50Ω)</td>
</tr>
<tr>
<td>RF separation (2-tone test)</td>
<td>10MHz</td>
<td>100Hz</td>
<td>1MHz</td>
</tr>
<tr>
<td>OIP3, in dBm</td>
<td>27.4</td>
<td>26.9</td>
<td>25.8</td>
</tr>
<tr>
<td>IIP2, in dBm</td>
<td>15.1</td>
<td>14.3</td>
<td>23.5</td>
</tr>
<tr>
<td>IIP3, in dBm</td>
<td>67.1 (for 5.75GHz)</td>
<td>54.5</td>
<td>N/A</td>
</tr>
<tr>
<td>SSB noise figure (50Ω), in dB</td>
<td>17.1</td>
<td>18.6</td>
<td>12.5</td>
</tr>
<tr>
<td>Zin at 2.4GHz, in Ω</td>
<td>64.5-j4.2</td>
<td>60.3-j7.1</td>
<td>64.6-j13.9</td>
</tr>
<tr>
<td>Differential LO, in dBm</td>
<td>1.1</td>
<td>-2.6</td>
<td>-3.0</td>
</tr>
<tr>
<td>Supply voltage, in Volts</td>
<td>2.2V nominal</td>
<td>1.2V &lt; Vcc &lt; 3.3V</td>
<td>5.0</td>
</tr>
<tr>
<td>Bias current, in mA</td>
<td>8.0</td>
<td>6.0</td>
<td>78.0</td>
</tr>
<tr>
<td>Power dissipation, in mW</td>
<td>17.6</td>
<td>13.2</td>
<td>390</td>
</tr>
<tr>
<td>Active chip area, in mm² (including testpads)</td>
<td>0.014 (1.4)</td>
<td>N/A</td>
<td>0.73 (2.07)</td>
</tr>
<tr>
<td>Device technology</td>
<td>0.2μm SiGe-HBT</td>
<td>Bipolar</td>
<td>SiGe-HBT</td>
</tr>
</tbody>
</table>

5.3. Conclusions and Summary

This chapter presented a feedforward-linearization topology that suppresses both third-order harmonic and intermodulation distortion generated by a mixer switching quad stage. In the proposed topology, two mixers are designed to operate in parallel. The LO swing, bias current, transistor size, and degeneration resistance are selected such that IM3 components of their IF outputs are anti-phase and equal in magnitude. When the IF outputs of the two mixers are summed, their IM3 components cancelled each other. The fundamental components of the IF outputs are summed in-phase and therefore added constructively.

Simulation indicated that the feedforward approach could enhance a BJT mixer IIP3 by at least 7dB at 5.75GHz. Simulated results also predicted that high linearity could be attained over a broadband with an IIP3 better than 12.7dBm over the 1-10GHz range (see Fig. 5.7). The feedforward-linearization was verified by measurement, albeit with the two mixers having different bias currents than those predicted by simulator. Over an RF frequency of 1 to 6GHz, the feedforward-linearized BJT mixer exhibited an IIP3 that ranged from 11.72 to 15.22dBm (see Fig. 5.13). The results presented in this chapter were published in [76][77]

An implementation of the feedforward linearization technique presented in this chapter on a CMOS mixer will be presented in the next chapter. A CMOS variant will allow the

Figure 5.14: Measured conversion gain and IIP3 at various supply voltages (I_{MixerA} = 3mA, and I_{MixerB} = 3mA)

Figure 5.15: RF input degeneration using impedances from the preceding stage
feedforward-linearized switching quad to be integrated with the other RF building blocks and baseband circuits, leading to further cost saving.

Chapter 6
Feedforward-Linearized Switching-Quad in CMOS

In the previous chapter, a feedforward-linearized bipolar mixer topology that utilized two switching quads was presented. In that design, two bipolar switching quads were used in parallel and designed such that third-order intermodulation distortion (IM3) components generated at their IF outputs are equal in magnitude but anti-phase. When the outputs of the two switching quads are summed, their IM3 components cancel, resulting in a distortion-free output. Similar to the feedforward-linearized differential-pair design presented in Chapter 4, this feedforward-linearized mixer topology can also be adapted to CMOS technology in order to facilitate integration with other radio-frequency (RF) building blocks and digital circuitry as part of a system on a chip.

A highly linear mixer, such as that presented in the previous chapter, is imperative in the realization of a receiver that is capable in tolerating unwanted signal/interferers. This is so as a mixer is typically placed after a preamplifier in a receiver and the linearity (input third-order intercept point or IIP3) of the mixer typically determines the overall IIP3. With higher IIP3, a receiver can tolerate interferers of higher power thereby allowing the receiver to maintain the desired bit-error-rate (BER) and data throughput when the receiver is operating in a heavily networked environment. Unfortunately, increasing the linearity of a mixer is typically achieved at the expense of other circuit parameters such as an increase in noise figure. However, as discussed in Section 2.4.4, trading higher noise figure for higher IIP3 in a mixer is acceptable or recommended. It can ultimately lead to a higher dynamic range.

The first half of this chapter focuses on design of the CMOS feedforward-linearized mixer, similar to the BJT design presented in the previous chapter. Magnitude and phase of the IM3 components at the IF output of the CMOS switching quad are also dependent on the LO amplitude, bias current, and transistor aspect ratio. Besides replacing all BJTs in Fig. 5.5 with FETs, the CMOS design also includes several modifications that address some of the shortcomings of the BJT feedforward-linearized mixer. These modifications include the use of an on-chip trifilar transformer and a LO buffer where they take a balanced RF/LO signals and generate a pair of balanced RF/LO signals. The second half of this chapter describes the measured performance of the CMOS feedforward-linearized mixer over a frequency band from 2 to 6GHz. Similar to the BJT design, the measured results will also showed that high linearity (third-order intercept point, or IIP3) can be attained by the CMOS variant. Finally, a brief discussion of the noise performance and summary will be provided at the final section of this chapter.

6.1. CMOS Implementation

Unlike the BJT variant, the CMOS prototype has two balanced input (RF and LO) and a balanced outputs (IF). This eliminates the need for a hybrid splitter (shown in Fig. 5.5) for characterization, thereby reducing the number of external components and their associated magnitude and phase errors. An LO buffer is also included in the CMOS prototype. The buffer allows different LO swings to be used when driving the two switching quads. This permits optimization of the distortion at the IF terminals by changing the LO swings during the testing. Finally, as suggested in the previous chapter, degeneration of the mixer RF input is provided by the output impedance of the previous stage (in this case, 50Ohm output impedance of test equipment). This reduces excess thermal noise added by the degeneration resistors and lowers the overall noise figure.
Feedforward-linearized switching quad

6-1-1. Design and Simulation

The feedforward-linearized CMOS mixer prototype is designed in a 0.13μm CMOS technology [49]. This mixer is intended to be the first downconverting mixer within an RF receiver and was designed for a supply voltage of 1.2V in the 2-6GHz, which encompasses the frequency ranges of the IEEE802.11a/b/g standards. A schematic of the mixer is shown in Fig. 6.1. The CMOS mixer consists of two parallel switching quads (consisting of $M_{1,g}$ and $M_{2,g}$). The switching quads are configured for feedforward operation to reduce intermodulation distortion in the output currents. Similar to the BJT design, the two switching quads are designed such that the IM3 components at their output terminals are equal in magnitude but opposite in phase. When the outputs of the parallel switching quads are summed by connecting the drain terminals as shown in Fig. 6.1, the IM3 components cancel resulting in an internal linear output. As mentioned before, the condition for distortion cancellation is realized when the appropriate LO swings, transistor transconductance (aspect ratio and bias current), and degeneration resistance are used in the design of the mixing quads.

As shown in Fig. 6.1, RF and LO signals for the two switching quads are derived from balanced RF and LO signals, respectively. At the RF port, magnetic coupling between the primary and secondary windings of an on-chip trifilar transformer couples RF signals to both switching quads. At the LO port, appropriate LO swings, transistor transconductance (aspect ratio and bias current), and degeneration resistance are used to present the optimal LO swings to each switching quad.

The Spectre™ simulator was used to determine the optimal values for transistor aspect ratios, bias currents, and LO swings in order to realize distortion cancellation at the IF outputs. The transistors of the mixer are biased with 3.5mA and have an even-order resistive termination of 200Ω ac-coupled via the centre-tap of the transformer secondaries. The optimal LO swings are 0.6V pk-pk and 0.2V pk-pk for Mixer A and Mixer B, respectively.

Fig. 6.2 shows the top view of the planar trifilar transformer that is used to couple RF signals to both switching quads. The terminals of the primary winding (solid) are at the top of the figure, while the secondary terminals (shaded and dotted black) are grouped together on the right side. Applying balanced RF signals to the terminals of the primary winding induces matching RF signals at both secondary windings. These signals are then fed to the switching quads by connecting the secondary windings to the source terminals of $M_{1,g}$ (see Fig. 6.1). The transformer has a physical turn ratio of 1:1 and is realized as a planar interwinding of three microstrip lines. The primary and secondary windings are implemented in 3μm thick metal layer over ~9μm of inter-metal dielectric. All three conductors have a width of 3.24μm and are separated from each other by 2μm (i.e., minimum spacing between the conductors to increase magnetic coupling). Lower metal layers are used to interconnect the primary and secondary windings. The trifilar transformer was optimized for a passband at 3GHz and occupies a 250x250μm² silicon area but could be smaller if multiple metal layers were used in its construction, rather than the planar design.

The GEMCAP2 program [78] was used to generate a lumped-element model for simulation. This model was also used to extract the low frequency electrical parameters of the trifilar transformer. The self inductance and DC resistance on the primary winding are 3.73nH and 6.35Ω, respectively. Both secondary windings have the same low frequency electrical parameters, with a self-inductance and DC resistance of 3.42nH and 7.81Ω, respectively. The secondary windings are coupled to each other with a k-factor and mutual inductance of 0.84 and 2.96mH, respectively. The secondary windings are coupled to both primary and secondary terminals with a k-factor and mutual inductance of 0.8 and 2.72nH, respectively. Tuning capacitors ($C_t$ in Fig. 6.1) are placed in shunt at both positive and negative terminals of the primary windings to compensate for magnetic leakage at the expense of bandwidth. To accommodate an operating frequency of 2.6GHz, MOSFET switches (not shown in Fig. 6.1) in series between the tuning capacitors and ground are used to fine-tune the passband. From simulation, a value of 1.5pF was chosen for the tuning capacitors. Each tuning capacitor was realized using four parallel capacitors (0.1pF, 0.2pF, 0.4pF, and 0.8pF), this allow for tuning to account for variation and uncertainty during the fabrication process.

Using the trifilar transformer as a power splitter to generate matching balanced RF signals for both switching quads eliminates the need for an external power splitter such as that shown in Fig. 5.10. This improves matching between the pair of differential RF signals and within each pair of balanced RF signals. Although the trifilar transformer balun also introduces phase and magnitude errors post-fabrication, they are expected to be minimal. Simulations predict that phase error was less than 1 degree while magnitude error was less than 0.02dB between 2 and 6GHz. These errors are mainly determined by the resolution of the fabrication process and should be significant lower than those caused by external components (i.e., accuracy on the width of the trifilar’s conductor is ~0.25μm).

Aside from providing RF connectivity for both switching quads, the trifilar transformer is also used to synthesize the series degeneration resistance at the source terminals of the mixing quads required for distortion cancellation (see Fig. 5.15). For this stand-alone mixer prototype, the 50Ω output impedance of test equipment will be transformed from the primary to the secondary windings to provide the necessary degeneration while adding minimal noise. Finally, the bias current to the switching quads are provided at the center-taps of the two secondary windings. Centre-tapping does not affect the small signal operation and allows the switching quads to be biased independently. Also shown in the figure are two...
even-order resistive terminations ($R_1/C_C$ and $R_2/C_C$) which are designed to decrease the overall IM distortion via second-harmonic mixing (as suggested for BJT amplifier in [61]) for optimal IM cancellation. These termination resistors appear in the common-mode circuit and will (ideally) not affect the differential signal gain or increase the noise figure.

Fig. 6.3 shows the schematic of the two-stage LO buffer that is used to generate LO swings for the two switching quads. The first stage consists of two static inverters between the supply rails driven by balanced sinusoidal LO signals. The gate terminals of transistors $M_{9/12}$ are biased at $V_{DD}/2$ via a resistive divider (not shown in the figure), while the differential LO signals are ac-coupled to the static CMOS inverters. The static inverters increase the slew rate of the input signals and generate a pair of differential rail-to-rail square-wave LO signals ($LO_{A^-}$ and $LO_{A^+}$). Each of these “square” LO signals is connected to, and drives two limiting amplifiers to generate the desired LO swing for each switching quad.

Each of the four amplifiers consists of a transmission gate with a pull-up PMOS transistor driven by complementary LO signals. The output voltage swing of each amplifier is limited to $V_{DD}$ and $V_{BB}$ (or $V_{BB}$). This clamping operation can be understood by considering the behaviour of a single amplifier, for example transistors $M_{13/14}$. When $LO_{A^-}$ is low (0V), transistor $M_{13}$ is off while transistors $M_{14/15}$ are on, creating a low-impedance path between $V_{DD}$ and $V_{BB}$. During the other half cycle, $LO_{A^-}$ is high (1.2V) causing transistors $M_{13/14}$ to turn off, while transistor $M_{15}$ is on and short-circuits the output to $V_{DD}$ (or 1.2V). Similarly, the outputs of the other three limiting amplifiers constrain the swing between $V_{DD}$ and $V_{BB}$, thereby generating the desired LO swings for the switching quads.

Similar to the trifilar transformer, the two-stage LO buffer is also designed to operate between 2 and 6GHz. The minimum gate length (0.12μm) is used for all transistors to minimize their parasitic capacitances. The gate width for all transistors in the two-stage LO buffer are annotated in Fig. 6.3. In the first stage of the LO buffer, the optimal gate widths for the NMOS and PMOS transistors were determined via simulation to be 9μm and 27μm, respectively. The ratio between the gate widths of PMOS and NMOS transistor widths is chosen such that there is a symmetrical transfer characteristic. The outputs of the static inverters toggle between 0V and $V_{DD}$ whenever the voltage at the gate terminals crosses $V_{DD}/2$. Unlike the first stage buffer, the voltage outputs of the second stage swing from $V_{DD}$ to $V_{BB}$ instead of down to 0V. From simulation, the optimal gate width for the transmission gate transistors (i.e., $M_{13}$ and $M_{14}$) and the pull-up PMOS transistors is 20μm. During normal operation, the $V_{BB}$ and $V_{BB}$ are set to 0.6V and 1.0V, respectively, to generate the optimal LO swings to Mixer A (0.6Vpk) and Mixer B (0.2Vpk). Spectre™ predicted that the output swings of the LO buffer can be limited and adjusted to generate the desired condition for distortion cancellation. An input differential LO with an amplitude of 1.2Vpk-pk is used to drive the two-stage LO buffer, with an input power of +8.5dBm across the 100Ω input termination. The LO buffer drives Mixer A and Mixer B. Although the input sinusoidal signals driving the LO buffer have a relatively large swing (1.2Vpk) for characterization purposes, simulation results indicate that smaller amplitude can be used without compromising the integrity of the LO signals applied to the two mixers. For example, a sinusoid with an amplitude as low as 0.5Vpk-pk and 1Vpk-pk for 2GHz and 6GHz operation, respectively, can be used without compromising the LO signals used to drive Mixer A and Mixer B. For comparison purposes, the simulated performance at 2 and 6GHz is plotted over one LO period in Fig. 6.4.

At an operating frequency of 2GHz, the differential LO outputs for Mixer A and Mixer B swing between ±0.6V and ±0.2V, respectively, with square-wave-like waveforms. Positive and negative transitions for LO of Mixer A ($LO_{A^-}$) require approximately three-tenths of a LO period (equivalent to 3ns). Similarly, the LO outputs applied to Mixer B ($LO_{B^-}$) also switch between the positive and negative states quickly, and the transition time is approximately one-tenth of a LO period (equivalent to 0.5ns). The rise and fall-times of $LO_{A^-}$ are shorter because its amplitude swing is much lower. Both $LO_{A^-}$ and $LO_{B^-}$ have a square-wave profiles and this helps to turn on and off the transistors of both
switching quads rapidly, thereby reducing the overall noise at the IF outputs. The overshoots at the transitions between positive and negative states are caused by feedforward via the gate-drain parasitic capacitances. As the operating frequency increases to 6GHz, transition time for LO output increases and both waveforms become more sinusoidal, and the amplitude swing for LOA also decreases. These changes happen as higher odd-order harmonics of LO are attenuated by parasitic capacitances present at the gate terminals of transistors M1,4.

Besides attenuating higher-order harmonics, parasitic capacitances also introduce phase delay to both LOA and LOB. Since the gate widths of transistors M1,4 are larger than transistors M5,6, there is greater parasitic capacitance at the gate terminals of M1,4 and a larger phase shift in LOA. The effects on phase matching between the IF outputs of the two individual mixers were simulated and compared to the zero crossing of the differential LO voltage outputs. Simulation results predict that LOA leads LOB by 8.6° at 2GHz but increases to 13.5° at 6GHz. This difference between the LO signals for the two mixers compromises the matching of both fundamental and IM3 components at the IF outputs, eventually placing a limit on the effectiveness of the feedforward-compensation.

In this prototype design, the VB^ and VBB in Fig. 6.3 are defined by off-chip voltages. However, on-chip voltage sources would be used when the feedforward-linearized CMOS mixer is used in a single-chip transceiver. When integrated with an on-chip oscillator, the resistive termination at the input terminals of the two-stage LO buffer can be omitted. Removing the resistive termination reduces the LO voltage swing required at the input to the LO buffer, and would lower the overall power consumption.
As mentioned earlier, the IF outputs of the two individual mixers will experience different delays due to parasitic capacitance and phase mismatch differences between the LOs. Fortunately, simulation results indicated the phase difference between the fundamental components of the IF outputs is minimal. The largest simulated phase difference between the fundamental components occurred at 5.0GHz, when the outputs of Mixer A lag the outputs of Mixer B by 9.8°. This phase error resulted in a 0.02dB reduction in the fundamental output power.

As shown in Fig. 6.5, Spectre™ predicts that the feedforward-linearized mixer will exhibit an IIP3 greater than +15.4dBm within the frequency band of 1-6GHz. Similar to the conversion gain, the IIP3 fluctuates within the frequency band resulting in local maxima and minima. This behaviour is caused by parasitics, as they introduce phase and magnitude errors between the IM3 components of the two individual mixers, resulting in suboptimal distortion cancellation. For example, magnitude and phase errors between the IM3 components at 3.5GHz were determined via simulation to be 0.18 and 10.5° (IM3 components from Mixer B is 18% larger in magnitude and lead by 10.5°) resulting in an IIP3 of +18.9dBm. On the other hand, the simulator predicted that at 6GHz, the magnitude and phase errors will be -0.18 and 15.3° (IM3 components of Mixer A is larger than that of Mixer B) resulting in an IIP3 of +15.4dBm. Note that when both magnitude and phase errors equal zero, the IF output will be distortion-free and IIP3 will approach infinity.

Errors in phase and magnitude are introduced due to the differences in transistor aspect ratio in the quad. Mismatch between transistors M1,4 and transistors M5,8 introduce different delay times between the source and drain terminals and attenuation (magnitude of IM3 components). Phase and magnitude errors are also introduced by the LO buffer when parasitic capacitor affect the slew rate and magnitude of LOA and LOB resulting in sub-optimal LO waveforms. As described earlier and shown in Fig. 6.4, parasitics also introduce phase errors between the LOs that compromising distortion cancellation.

Figure 6.6: Variation in IIP3 due to process and mismatch from 500 Monte Carlo simulations of the feedforward-compensated CMOS mixer.

Although these inherent phase and magnitude errors place an upper limit on the effectiveness of feedforward-compensation, simulation results suggest that an IIP3 of better than +15dBm can still be achieved between 2-6GHz.

Similar to the BJT mixer described in Section 5.2 statistical analysis was used to determine the effects of process variation and mismatch. Fig. 6.6 shows the results of Monte Carlo simulation on the complete (i.e., including the two-stage LO buffer and input trifilar transformer). The mixer is configured to downconvert RF signals from 5.75GHz to an IF of 100MHz, and the biasing circuits were adjusted after each iteration to ensure that the individual mixers always have a constant bias current (3.5mA each). As shown by the histogram, the distribution of the simulated IIP3 data for the CMOS mixer is similar to that obtained for BJT mixer (Fig. 5.8) and is not Gaussian distributed. Like the BJT mixer, most of the data is skewed towards higher IIP3, with just over 86.8% residing between 14.5 and 18.5dBm. Data for the Monte Carlo was fit to an extreme value distribution, and its probability density function is also included in Fig. 6.6. Mean and standard deviation of the IIP3 are 15.9dBm and 1.3dB, respectively.

Fig. 6.7 shows the results of Monte Carlo simulations of process variation and mismatch on the even-order distortion (IIP2) of the feedforward-linearized CMOS mixer. Similar to the IIP3 simulations described earlier, the mixer was configured to downconvert an RF signal of 5.75GHz. The statistical analysis also included process variation and mismatch of the LO buffer. Similar to the BJT mixer in Section 5.2, the even-order distortion at the IF output is primarily caused by random mismatch between balanced paths. As described earlier and shown in Fig. 6.4, the IIP2 data has a Gaussian distribution with a mean and standard deviation of 40.7dBm and 2.5dB, respectively. As mentioned in Section 5.2, this IIP2 performance can be further improved by using calibration techniques such as described in [60].

Figure 6.7: Variation in IIP2 due to process and mismatch from 500 Monte Carlo simulations of the feedforward-compensated CMOS mixer.
Comparing to the BJT variant, the CMOS mixer exhibits more even-order distortion (mean IIP3 of 40.7dBm versus 67.2dBm). This increase in even-order distortion is because of the higher component counts (i.e., transistors within the two-stage LO buffer) but also due to inferior matching of transistors in the CMOS design compared to the BJT implementation [79].

All simulation results presented in prior discussions assumed that the output impedance from the test equipment is exactly 50Ω, however, the actual impedance typically deviates from this ideal value. Since this impedance is presented to the RF input of the CMOS mixer and is used as a parameter in the feedforward-linearized operation, variation in the value could affect the linearity performance. Sensitivity of the linearity performance of the mixer to variation in impedance seen at its RF input was studied in simulation by placing an offset impedance in series between the RF source and the primary terminals of the input trifilar transformer. The offset impedance was modeled by a series combination of resistor with an inductor or capacitor, while the RF source was modeled by a voltage source with a 50Ω output impedance. Deviation of up to ±5Ω in both real part and reactance that correspond to an impedance range from (45-j45)Ω to (55+j55)Ω were considered in the sensitivity analysis. Inductance/capacitance of the offset impedance was chosen such that they only generate the desired offset reactance at the RF. For example, a 1.384nH inductor and 5.536pF capacitor were used to generate ±j5Ω at an RF of 5.75GHz.

Simulated variation in the ratio of IM3 components generated by Mixer A to the IM3 components of Mixer B (i\textsuperscript{R}) are plotted in Fig. 6.8. Both the magnitude and phase of the ratio are plotted as they ultimately determine the maximum distortion suppression that can be achieved. Total suppression is achieved when i\textsuperscript{R} = 1.180°. Instead of the absolute phase difference between the IM3 components (\angle i\textsuperscript{R}), the phase difference referenced to the desired 180° (\angle i\textsuperscript{R} - \angle i\textsuperscript{R}\textsuperscript{3}) is plotted in the figure.

Simulation results show that increasing the imaginary part of the series-equivalent source impedance (larger +jΩ) is desired for both phase and magnitude matching. Increasing the real component helps reduce the magnitude mismatch but introduces more phase error. For example, changing the source impedance from 50+j5Ω to 50+jjΩ changed i\textsuperscript{R} from 1.113\angle 163.9° to 1.065\angle 166.1°, whereas changing the source impedance from 45+j0Ω to 55+j0Ω changed i\textsuperscript{R} from 1.114\angle 169.9° to 1.093\angle 162.3°. Consequently, the linearity (IIP3) of the mixer fluctuated with the variation in the impedance at its RF port. The variation is shown by the family of IIP3 plots in Fig. 6.9.

Using the same example as before, changing the source impedance from 50+j5Ω to 50+j5Ω increases the IIP3 from 16.1dBm to 17.1dBm. This is expected since increasing the reactance reduces mismatch in both magnitude and phase, resulting in a 1dB increment in IIP3. On the other hand, changing the source impedance from 45+j0Ω to 55+j0Ω reduces the IIP3 from 17.2dBm to 16.1dBm. This 1dB drop in IIP3 transpires, as increasing the real component of the series impedance decreases the phase mismatch of IM3 components by 7.6° in spite of the fact that better magnitude matching is achieved (1.093 versus 1.114). The best IIP3 performance is achieved when the IM3 components from the two mixers have the same magnitude and are anti-phase to each other. In simulation, variation in the source impedance causes IIP3 of the mixer to fluctuate between 15.7 and 17.9dBm, but the variation is limited to within ±1dB of the nominal IIP3 (16.6dBm along with an i\textsuperscript{R} of 1.101\angle 162.3°). The best and worst IIP3 occurred when the source impedance seen by the mixer is 55+j5Ω and 45+j5Ω, respectively.

As a result, although IM3 cancellation via the feedforward-linearized operation is compromised by source impedance variations, the mixer is still expected to exhibit high linearity.
Finally, besides affecting the IIP3 of the mixer, variation in the impedance presented to the RF input of the mixer will also affect the conversion gain. However, simulation results indicated that the changes are minimal. Within the same range of offset impedance, Spectre™ predicted that the conversion gain would vary between 12.19 and 12.22dB.

6-1-2. Measurement

A prototype of the feedforward-linearized CMOS mixer was fabricated in a 0.13μm CMOS process (fT ~ 85GHz) [49] and characterized over the frequency range from 1 to 6GHz. Fig. 6.10a shows the microphotograph of the fabricated die which occupies an area of 1.49x1.34mm² (including the bondpads). Placement of the mixer, LO buffer, and input transformer are highlighted in the microphotograph. As shown in the figure, the supply voltage for the LO buffer is provided via a separate bondpad as this lowers coupling of switching noise generated by the LO buffer to other circuits. Two bondpads (VBA and VBB) were included in the design to compensate for variations in the bias currents of Mixer A and Mixer B caused by processing. Switches for the tuning capacitors are turned on and off when 0V/1.2V is applied to bondpads Vgci to VBC4. Multiple vertical metal-metal capacitors from the design kit were used as bypass capacitor at all DC pins to reduce high frequency noise. Metal-metal capacitors were also used to implement the tuning capacitors as well as the filtering capacitors (two 10pF) at the IF terminals.

Characterization of the prototype was carried out with the aid of a custom-made test board (see Fig. 6.10b) that allows the GHz signals (RF and LO) to be delivered directly to the bondpads via microwave probes while all bias pins (e.g., VDD, VBA, and VBB) are wire-bonded and incorporate additional bypass and decoupling filters. Using this setup, insertion loss and load inductance along the high speed paths are minimized while supply noise and spurious tones at bias nodes are attenuated or prevented. Furthermore, the input impedance of the RF port can be measured directly. The IF outputs were wired bonded to a pair of PCB traces connected to SMA connectors. External 470Ω transimpedance amplifiers along with a 180° transformer balun convert the differential IF output currents to a single-ended voltage driving a 50Ω load (e.g., a spectrum analyzer).

Similar to the BJT mixer described in Section 5.2, conventional two-tone tests were conducted experimentally to determine the combination of bias currents and LO swings that generate maximal IM3 suppression. Magnitudes of the IM3 components generated solely by each mixer for different bias current and LO swing were first determined by making one of the two mixers a cascode stage. This was realized by setting VBB to 1.2V, which forces transistors M5,6 in Fig. 6.1 to operate as cascode transistors. Consequently, Mixer B does not downconvert the input differential RF signal and the IF signals (both fundamental and IM3 components) observed at the outputs are then solely generated by Mixer A. Similarly, setting VBA to 1.2V allows the IM3 caused by Mixer B to be determined. With the
IM3 generated by each mixer for various bias currents and LO swings determined, the optimal selection that generates the largest IM3 cancellation observed at the IF outputs can be chosen. In each two-tone test, the mixer was configured to downconvert two 6.0GHz signals spaced 100Hz apart with an input amplitude of -8dBm to 50kHz IF. Fig. 6.11 shows the measured IM3 power level generated by each mixer as their LO swings are swept from 0.1 to 1.2VPk for three bias currents (3.0, 3.5, and 4.0mA). The power levels of the IM3 components provide an indication of the both the input and output third-order intercept points, since lower IM3 corresponds to better IIP3 and OIP3 generally. Also included in the figure, is the simulated IM3 output predicted by Cadence Spectre™. A bias current of 3.5mA is used for the simulation in both cases. Measured results show that both the bias current and LO swing of a mixer switching quad influence the magnitude of the IM3 observed at the output by up to 20dB.

As shown in the figure, the magnitude of IM3 components generated by Mixer A and Mixer B have different behaviour as the LO swing increases from 0.1 to 1.2VPk. The IM3 generated by Mixer A has a convex response while those generated by Mixer B is characterized by local maxima and minima. The measured results also show that increasing the bias current of Mixer A for a fixed LO swing reduces the IM3 magnitude. On the other hand, increasing the bias current of Mixer B shifts the occurrence of the local maxima and minima to a higher LO swing (e.g., shift the graphs in Fig. 6.11 horizontally, towards the right). Experimentally, these results showed that the magnitude of the IM3 components are functions of both the bias current (or transconductance) and the LO amplitude. This dependency is especially true for Mixer B.

The variation of IM3 with bias currents and LO swings is predicted by the simulator, and the causes of these phenomena were discussed in Section 3.1. However, measured results differ from values predicted by the simulator and ultimately affected the optimal voltages VB3 and VB4 for distortion cancellation. The difference between measurement and simulation is greater for Mixer A, where measured distortion components have output levels that are 10dB greater than those predicted in simulation. They may be caused by inaccuracy in the models used to predict the magnitude of the distortion components. The inverted concave property of Mixer B's IM3 also suggests that its phase changed by 180° as the LO swing is varied from 0.1 to 1.2VPk. Information on the phase could not be obtained as a spectrum analyzer was used for testing, however, a 180° phase shift is predicted by Cadence-Spectre™. This phase shift is a necessary condition to operate two parallel mixers in a feedforward-compensation manner. It generates anti-phase IM3 with equal magnitude that are then combined destructively.

The inverted concave shape could also have been the result of optimal bias current and LO swing. As magnitude of the IM3 components in measurement differed to those predicted in simulation, the values required for the bias currents and LO swings for maximum distortion suppression were determined again. Experimentally, the optimal bias currents for both mixers were determined to be the fundamental and IM3 tones predicted by the simulator. Namely, appropriate selection of transistor aspect ratios, bias currents, and LO swings will result in the optimal IM3 tones from the two mixers respectively. The LO swings were set by making the DC voltages for VB3 and VB4 to 0.15V and to 0.95V, respectively. Fig. 6.12 shows the measured spectrum at the IF outputs for two cases: Mixer B operating as a switching stage and then as a cascoding stage. In both cases, two RF signals spaced 7.5kHz were downconverted from 5.75GHz to an IF around 50kHz. An RF input power of -12.06dBm was used for the comparison. IM distortions generated at this input power were mostly due to weakly nonlinear behaviour, as compression was not observed in the measured fundamental response, even when the input power was -1.06dBm. The sideband skirts on either side of the lower fundamental tone were due to intermodulation. The sideband skirts on either side of the lower fundamental tone were due to phase noise produced by the synthesizer used in the test setup (Agilent HP 8665B). This phase noise show that although there was small improvement (-0.5dB) in the output fundamental tones by having Mixer B operating in parallel to Mixer A. The IM3 components at the IF outputs were suppressed by 11.9dB. This suggests that the IM3 components from the two mixers are about equal in magnitude and anti-phase to each other, translating to an improvement of about 6dB in IIP3. The IIP3 of the feedforward-linearized CMOS mixer was measured to be 14.7dBm. The fundamental tones did not experience a similar increase, as the fundamental tones generated by Mixer B were about 12dB lower than those generated by Mixer A.

By demonstrating the capability to suppress IM3 at the IF output and increasing the third-order distortion components, the two output spectra shown in Fig. 6.12 also validated the phase relationship between Mixer A and Mixer B. The phase shift is a necessary condition to operate two parallel mixers in a feedforward-compensation manner. It generates anti-phase IM3 with equal magnitude that are then combined destructively.

The measured conversion gain, IIP3, and OIP3, of the feedforward-linearized CMOS mixer in the frequency range from 1-6GHz are plotted in Fig. 6.13. Similar to the simulation, all 4-bits of the two-tone capacitors were switched on or off when the RF was below or at and above 3.5GHz. This setup was determined experimentally to produce the highest conversion gain at all frequencies within the 1-6GHz range. Across the frequency range, the measured conversion gain varied between 8.9 and 11.9dB. The lowest conversion gain was measured at 2.4GHz while the highest conversion gain of 5.5dB was measured at 6.0GHz. As the frequency of operation was increased from 1 to 6GHz, the IIP3 of the feedforward-linearized CMOS mixer was increased from 1.6 to 6.2dBm at 1.0GHz to 21.2dBm at 2.4GHz. Increasing the RF further led to a continual drop in IIP3 until the lowest IIP3 of 13.4dBm was measured at 4.5GHz. From
5-6GHz, the IIP3 of the mixer was relatively constant at approximately 15dBm. The 8dB measured variation in IIP3 with frequency was probably due to mismatch in magnitude and phase between the IM3 tones of the two mixers caused by parasitic capacitances as discussed in Section 6-1.1. Comparing the measured conversion gain and IIP3 to the values predicted from simulation (see Fig. 6.5) reveals that the frequency responses are different, especially for IIP3.

Values obtained in measurement are comparable to those predicted from simulation, however, frequency responses for the conversion gain and IIP3 differ from their simulated counterparts. For example, the measured conversion gain and IIP3 at 6.0GHz are 11.90dB and 15.1dBm, respectively, whereas the simulated values are 12.10dB and 15.40dBm. The measured OIP3 (conversion gain + IIP3) across the frequency band varied by 5.2dB but was measured to be better than 24.9dBm.

At DC, the LO buffers do not draw static/bias current, but this is no longer true as their transistors switch from one state to another. The amount of current drawn from the supply is proportional to the frequency of operation. Experimentally, the two-stage LO buffer drew a current of 1.95mA and 5.43mA at operating frequencies of 1GHz and 6GHz, respectively.

Noise figure of the feedforward-linearized CMOS mixer as it downconverts RF signal at 1 to 6GHz was measured, and the spot noise values at an IF between 10 to 300MHz were recorded for each RF input frequency. The noise profile at IF below 10MHz was not characterized due to limitations of the measurement setup. Fig. 6.14 shows the measured and simulated noise figure of the feedforward-linearized mixer as it downconverts an RF signal of 5.75GHz. As shown in the figure, flicker noise dominates below 50MHz IF, leading to relatively high noise figure in both simulation and measurement. As the IF increases, both the measured and simulated noise figures decrease to 21dB at an IF of 300MHz. The peaks in the measured data around 100MHz and 170MHz were traced to interference from broadcast FM radio stations and a local pager service, respectively. Overall, the noise figure values predicted by the simulator presented a pessimistic view when compared to the measured data. For example, the simulated noise figure at an IF of 10MHz was more than 5dB higher than in measurement. This better than expected noise figure was probably due to the combination of a larger LO swing for optimal linearity in measurement, and overestimation of the flicker noise in simulation.

The effect of having the second mixer operating in parallel on noise figure was also analyzed. This was carried out by measuring the noise figure of Mixer A alone and the results are included in Fig. 6.14. Comparison of the measured plots shows that distortion suppression was achieved at the expense of sensitivity. Over the measured IF, the degradation was approximately 2.8dB between 50 and 300MHz. Designing the transistors within Mixer B to have a smaller aspect ratio but the same bias current for distortion cancellation resulted in Mixer B having a higher flicker noise corner. Operating the two mixers in parallel raises the overall noise figure for the same IF. This increase is most significant at the lower IFs. For example, measured results showed that the noise figure increased by 4dB over the 10 to 20MHz range.

Noise figure of the CMOS mixer at other RF inputs have a similar IF profile to that shown in Fig. 6.14 (i.e., highest values at lower IF, spikes at 100MHz and 170MHz). The frequency behaviour of the spot noise at 50MHz as the RF changes from 1 to 6GHz is plotted in Fig. 6.15, and values predicted by the simulator are included in the figure for comparison. Within the 2 to 6GHz band, the measured noise figure fluctuates between 19.0 and 25.2dB, but is consistently better than the values predicted by Spectre™. The simulated noise figure within this frequency is approximately 26.0dB. Below an RF of 2GHz, very little of the RF input signal is magnetically coupled to the input of Mixer A and Mixer B.
This resulted in lower input SNR leading to higher noise figure in both measurement and simulations. For example, the simulated and measured noise figures at 1GHz are 34.3dB and 27.9dB, respectively.

Shown in Fig. 6.15 are the measured input-referred second-order intercept point (IIP2) and the differential return loss at the RF terminals. The IIP2 quantifies the even-order distortion generated at the IF output for various RF input frequencies obtained using a single die. Since the feedforward-linearized CMOS mixer utilizes a double-balanced structure, the main contributors of even-order distortion are on-chip mismatches due to limitation in lithography and magnitude and phase from the external couplers used in the test set-up. The measured IIP2 exhibits a general trend, where it is highest at low frequency and decreases with increasing frequency (with the exception of the dip at 2GHz). This is expected, since mismatch in phase and magnitude between each balanced pair of signals increases with frequency, thereby lowering the IIP2. Although not simulated, the IIP2 at other RFs should be similar to the measured values (i.e., highest at the low frequency and decreasing with increasing frequency). As a comparison, the measured and simulated IIP2 at 5.75GHz are 41.0dBm and 40.7dBm, respectively.

The measured RF input return loss of better than 10dB was achieved over 2.10-3.26GHz and 2.79-5.01GHz with all tuning capacitors turned on and off, respectively. The lower frequency band matches very well to that predicted from simulation, but the upper band is about 500MHz lower. This indicated that the simulator underestimates the parasitics associated with the tuning capacitors. The parallel equivalent of the RF input impedance measured at 2.4GHz consists of a 53.47Ω resistor and a 13.52nH inductor, while the parallel equivalent at 5.75GHz is a 50.1Ω resistor and a 495.8μF capacitor.

Measured results presented in this subsection shows that a high linearity (IP3) CMOS mixer can be attained between an RF of 2 and 6GHz using feedforward operation. The measured performance at 5.75GHz are summarized in Table 6-1 and the performance predicted by simulator are also included in the table for comparison. As shown by the numbers in Table 6-1, the measured results compare favourably to those predicted by the simulator. However, optimal distortion suppression were obtained in measurement and simulation with different LO amplitudes. This discrepancy was discussed earlier and was attributed mostly to limitation of simulation model in predicting the IMD components accurately. Although the range of values for the IIP3 obtained in measurement and simulation are comparable to each other, their frequency responses over 1-6GHz differ significantly from each other. This difference is mostly attributed to inaccuracy in modeling, processing variation, and imperfections in the test setup.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Feedforward-linearized CMOS mixer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Simulated</td>
</tr>
<tr>
<td>RF input, GHz</td>
<td>5.75</td>
</tr>
<tr>
<td>IF, in Hz</td>
<td>100M</td>
</tr>
<tr>
<td>Conversion gain, in dB</td>
<td>12.1</td>
</tr>
<tr>
<td>(IF load resistance, Ω)</td>
<td>(470Ω)</td>
</tr>
<tr>
<td>RF separation (2-tone test), in Hz</td>
<td>1M</td>
</tr>
<tr>
<td>OIP3, in dBm</td>
<td>27.5</td>
</tr>
<tr>
<td>IIP3, in dBm</td>
<td>15.4</td>
</tr>
<tr>
<td>IIP2, in dBm</td>
<td>40.7</td>
</tr>
<tr>
<td>SSB NF at 50MHz (50Ω), in dB</td>
<td>27.1</td>
</tr>
<tr>
<td>Parallel equivalent Zin at 5.75GHz</td>
<td>55.5Ω/300.5μF</td>
</tr>
<tr>
<td>LO swing for Mixer A/Mixer B, in Vpk</td>
<td>0.6/0.2</td>
</tr>
<tr>
<td>Supply voltage, in Volts</td>
<td>1.2</td>
</tr>
<tr>
<td>Bias current (including LO buffer), in mA</td>
<td>12.4</td>
</tr>
<tr>
<td>Power dissipation, in mW</td>
<td>14.9</td>
</tr>
<tr>
<td>Device technology</td>
<td>0.13μm CMOS</td>
</tr>
</tbody>
</table>

6.2. Discussion

Both simulated and measured results presented in this chapter demonstrated that the feedforward-linearized CMOS mixer topology exhibits high linearity. The measured results showed that when the
feedforward linearization topology is applied, IIP3 increased by 6dB while the noise figure degraded by 3dB. This increases the spurious-free dynamic range of the mixer by 2dB, as determined using Equation 2-1 and Equation 2-2 with n = 3. The feedforward-linearized CMOS mixer can be used as the first downconverting mixer of a receiver thereby allowing the receiver to tolerate interferers with high power level. This is especially beneficial when the receiver is expected to operate in a heavily networked environment where many radios operate in proximity to each other and have the potential to cause interference resulting in a decrease in BER performance. Furthermore, as the feedforward-linearized CMOS now able to endure more interferers, its operating bandwidth can be increased. This will enable the mixer to be used to receive signals of different wireless standards that are located at different frequencies, as mentioned in Section 2-4-4. This will facilitate the realization of a multi-standard radio described in Fig. 1-2b where the mixer can be shared to reduce complexity, parts count, and current consumption.

Both measurement and simulation data indicated that the flicker noise corner of the feedforward-linearized CMOS mixer to be in the MHz range. Consequently, this could compromise the BER of the receiver if the mixer is used as part of a low/zero-IF architecture receiver. Influence of the flicker noise at a particular IF frequency can be alleviated at the transistor level by either increasing the size of the individual transistor while keeping their aspect ratio constant or replacing the NMOS transistors with PMOS equivalents. The former approach lowers the flicker noise corner by increasing the product of the gate length and gate width (see Fig. 2.3) at the expense of speed, as more parasitic capacitances will be introduced at the gate terminals of transistors Mi.g. Changing to PMOS transistors with PMOS equivalents. The former approach lowers the flicker noise corner as the coefficient Kf in Fig. 2.3 for PMOS is generally much smaller than NMOS [10].

However, feedforward-linearized CMOS mixer exhibits a relatively high noise figure (e.g., noise figure of a Gilbert mixer is typically in the range of 10-14dB). This inferior noise performance arises due to the lack of transconductance stage at the input terminals of the feedforward-linearized CMOS mixer. The input transconductance stage in the Gilbert mixer provides gain and suppresses the noise generated by the switching stage. However, noise figure of feedforward-linearized CMOS mixer is acceptable if a low-noise preamplifier precedes the mixer. For example, a front-end realized using the feedforward-linearized CMOS mixer would need to have a preamplifier having a gain, noise figure and IIP3 of 18dB, 2dB and -2dBm, respectively. The receiver would then have a noise figure and IIP3 of 8.8dB and -2.2dBm, respectively, allowing it to be tolerant to interferers and complies with the IEEE802.11 standards. The noise figure of the feedforward-linearized CMOS mixer can also be lowered by increasing the impedance seen by its RF port. For example, as discussed in Section 2-3. When used as part of an integrated receiver, the RF port of the feedforward-linearized CMOS mixer sees the output impedance of the preceding stage or preamplifier. The output impedance of the preamplifier is typically much larger than the 50Ω used in the prototype design described in this Chapter (i.e., 100Ω to 200Ω). Increasing the source resistance at the RF port will increase the denominator of the second term in Equation 2-3, which in turn lowers the noise figure.

In the prototype design, an on-chip trifilar transformer was used to couple a balanced RF inputs to both switching quads. This implementation resulted in increment of silicon area (see Fig. 6.10) and reduction in RF bandwidth (see the input return loss plots in Fig. 6.15). For a single-chip solution intended for wideband operation, this trifilar transformer has to be replaced and a possible solution is shown in Fig. 6.16 where active circuits are used. For illustration purpose, only the half-circuits are shown. RF signals are coupled to the two switching pairs via two common-source amplifiers with local shunt-shunt feedback. Besides reducing the distortion of the common-source amplifiers, the local shunt-shunt feedbacks are also used to generate the desired output impedance. Therefore, the impedances (ZA.out and ZB.out) presented to the two switching pair can be chosen independently for optimal linearity performance. Although the proposed circuit is similar to a classic Gilbert mixer with source-degeneration, shunt-shunt feedback allows the desired impedance to be generated and presented to the two switching quads. This along with the appropriate selection in LO amplitudes, bias currents and transistor aspect ratio allow the distortion generated by the switching quad stage to be suppressed by the feedforward topology presented earlier in this chapter. In the gigahertz range, the impedance seen by the switching quad stage within a classic Gilbert-modulator is mostly the parasitic capacitance presented at the common-source nodes (e.g., nodes A and B in Fig. 6.16).

In single chip implementation, the input stage proposed in Fig. 6.16 can also be incorporated in the preamplifier design in which it become the second stage of a two-stage preamplifier, similar to that described in [20].

6.3. Summary

This chapter presented a CMOS variant of the feedforward-linearization topology that suppresses both the third-order harmonic and intermodulation distortion generated by a switching quad stage. Two switching quads are operated in parallel and are designed by appropriate selection of LO amplitude, bias current, transistor size, and impedance presented to its RF ports such that the IM3 components of their IF outputs are anti-phase and equal in magnitude. When the IF outputs of the two mixers are summed, their IM3 components cancel. The fundamental components of the IF outputs are in-phase and constructively.

A 2-to-4 on-chip trifilar transformer and a 2-to-4 LO buffer were also incorporated in the design to generate balanced RF and LO signals, respectively, for the two switching quads. Simulation predicted that the feedforward-linearized mixer can achieve high linearity, with an IIP3 better than 15.40dBm over the 2-6GHz RF frequency band (see Fig. 6.5). In measurement, high linearity was also obtained but with different LO amplitude. The CMOS mixer exhibited a range of 15.40 and 21.20dBm in IIP3 between the same frequency band (see Fig. 6.13), but the performance over frequency range differed from those predicted in simulation. As discussed earlier in this chapter, these differences were probably caused by a combination of factors such as magnitude and phase mismatch from the test setup, process variation and mismatch, and inadequacy of the models in predicting third-order distortion.
Furthermore, statistical analysis showed that like the BJT implementation, the feedforward-linearization implementation help to de-sensitize the IIP3 of the CMOS mixer to process variation and mismatch. Monte Carlo simulation showed that while linearity performance of each individual switching quad is affected by process variation and mismatch, their outputs track each other resulting in significant distortion cancellation. Consequently, the IIP3 predicted in simulation can be expected post-fabrication. As shown in Fig. 6.6, simulation results predicted that over 75% of the fabricated die would exhibit an IIP3 that fall within ±8dB of the nominal value (or better). This intrinsic high yield reduces the overall manufacturing cost as trimming may not be necessary post-fabrication.

This improvement in dynamic range and linearity is desired for the targeted wireless LAN applications. In the targeted application, BER of the wireless link will often be limited by the selectivity of the receiver since the “base-station” or wireless access point are often within a few tens or hundreds of metres from the receiver (~300m). This is especially true in a tightly-networked area where numerous wireless LAN receivers are operating in close proximity, each acting as an in-band interferers to the other receivers.

Chapter 7

Quality of service of a wireless link between two wireless devices is heavily dependent on the lowest achievable bit error rate. This in turn depends on the sensitivity and selectivity of the RF front-ends of these two wireless devices. While sensitivity had often been the more important parameter between the two, widespread adoption of wireless communications in the consumer market had created the situation and increased the likelihood where selectivity is now determining the overall BER. Proliferation of consumer wireless communication has increased the likelihood of multiple radios operating within a geographical space, each behaving like an interferer to other radios in their vicinity. These interferers will interact due to the nonlinear behaviour of each receiver, generating intermodulation and harmonic distortions that could potentially mask the desired signal.

This dissertation presents circuit topologies that mitigates the effects of interference on an RF front-end. The feedforward topologies were applied to various RF downconverting mixers, substantially lowering intermodulation and harmonic distortions at their outputs. All these mixers were designed to operate as the first downconverting mixer of a WLAN receiver where high level of proliferation of this wireless standard increases the likelihood of interference. Minimizing the distortion generated by a mixer as it usually dominates the overall selectivity performance of the receiver was the focus of this dissertation. Besides offering an output with significant lower distortion profile, the topologies presented in this dissertation also offered other desired factors. For example, low-voltage operation helps seamless integration with other digital circuits in modern integrated processes, broadband operation to accommodate both existing and emerging wireless standards, and robustness to process variation thereby reducing trimming cost and time.

Basic information such as terminology, source of distortion, and different types of distortion were first presented in Chapter 2. These discussions provided an understanding on how various forms of distortion are quantified, mechanism on how they corrupt the desired signal, and mechanism on how they are generated with illustration using basic BJT and CMOS amplifiers. A case study involving two receivers was then presented in this chapter, demonstrating the need for a mixer with high third-order intercept point to realize a robust receiver that is insensitive to interference. Following the case study, a brief discussion on the operation of mixer along with the two most commonly used mixer topologies in modern integrated RF front-end was presented. The first topology was the ubiquitous Gilbert-multiplier, an active mixer, while the second topology presented was the passive mixer. The discussion on passive mixer also included simulation results showing its distortion performance is directly related to the switching speed of the transistors. Increasing both the aspect ratio of the transistor and the slew rate of the LO helped to suppress distortion but places a premium on the LO buffer, which is difficult to implement when the operating frequency is in GHz range.

Distortion generated by the Gilbert mixer was detailed in the subsequent chapter. The first three sections presented in Chapter 3 dealt with the source of distortion generated by the three individual stages of the Gilbert-multiplier. In the first section, distortion performance of the switching quad stage was presented. Simulation was used and demonstrated that instead of behaving like ideal switches, the transistors generate distortion that appear at the IF outputs and are a function of the LO swing, bias current, and impedance seen at its input. The input transconductance stage was then presented where the limited linear range of its DC transfer function was identified as the largest contributor of nonlinearity. Various existing techniques that help to extend the linear range of the DC transfer function were then presented and discussed with the aid of simulator. In the third subsection, discussion was focus on the topology of the output stage. Advantages of using a transimpedance amplifier stage over
using a resistive load to convert the output current from the switching quad into voltage for subsequent building blocks were presented. Materials presented in these three subsections showed that all three stages of the Gilbert mixer contribute to the distortion components observed at the IF output, with one of the three stages generally being the dominant contributor. Distortion generated by the input transconductance stage is dominant when its output impedance is high (i.e., at low frequency). However, when the output impedance of the transconductance stage is low (i.e., in GHz operation), distortion generated by the switching quad stage is the dominant contributor. Distortion generated by the output stage can be the dominant factor if an improper choice of resistive load is chosen, causing output waveform clipping.

An active mixer with inductive degeneration input stage was also presented in the closing section of Chapter 3 to highlight the limitation of the existing techniques. Based on the information presented in Chapter 3, an on-chip feedforward methodology was proposed to realize a GHz active mixer with high linearity. The proposed feedforward approach differed from the classical feedforward linearization where it uses two amplifiers/mixers in parallel each generating anti-phase third-order distortion components with the same magnitude. When the outputs of the two amplifiers/mixers were summed, the distortion components cancel each other resulting in a distortion-free output.

Chapter 4 presented a feedforward-compensation topology that utilized two parallel amplifiers with different local feedbacks to realize the necessary condition for distortion cancellation. In the BJT variant, resistive feedbacks were applied to an emitter-coupled pair and differentially driven common-base stage such that the third-order distortion components generated at their output currents cancel each other. Simulation results showed that this topology offer substantially linearity (third-order harmonic distortion) improvement over other existing techniques at a low bias current (5mA). Furthermore, this feedforward-compensated transconductance stage also offer wider operating bandwidth (3dB bandwidth of 0.8/fc) and robustness to process variation, suggesting that it would be a suitable input stage for a highly linear wideband Gilbert mixer. This topology was then modified and implemented in CMOS, by replacing the emitter-coupled pair with a differentially driven common-source stage and the common-base stages with common-gate stages, and used as the input stage of a Gilbert mixer. Local feedbacks were again applied to the two amplifier stages such that they generate third-order distortion components that are equal in magnitude but anti-phase to each other. Results obtained for this CMOS mixer was presented, demonstrating the effectiveness of the proposed feedforward topology in realizing a highly linear transconductance at low power consumption (IIP3 better than 3.6dBm over 1-6GHz as described in Section 4.2).

Simulated and measured results presented in Chapter 4 also highlighted that the switching quad is not "distortion-free", but rather that it generates distortion as well. This contradicts the common assumption that the transistors behave as ideal switches and do not generate any distortion. Consequently, distortions generated by switching quad stage were considered leading to the design and implementation of a feedforward-linearized mixer topology. This topology was first demonstrated in its BJT variant and presented in Chapter 5. This feedforward-linearized mixer topology utilized two modified switching quads operating in parallel. Via appropriate selection in LO swing, bias current, and degeneration resistance, the two switching quads simultaneously downconvert the RF signals and generate equal magnitude but anti-phase intermodulation distortion at their IF outputs, allowing them to be cancelled. A CMOS variant of the feedforward-linearized mixer was also designed and implemented. The results for the CMOS variant were presented in Chapter 6. This topology had not been reported in literature before and was implemented for the first time.

For both BJT and CMOS variant, results were presented demonstrating how the undesired distortion generated by a switching quad can be utilized in a feedforward manner. Simulation and measured results demonstrated that the feedforward can suppress IM distortion at the IF output of the mixers by over 10dB, yielding improvement in at least 6-7dB improvement in IIP3 (see Fig. 5.6 and Fig. 6.12). Besides demonstrating broadband distortion cancellation, the proposed feedforward-linearized topology also demonstrated high resilience to process variation and mismatch allowing cost saving by reduction in trimming cost. Statistical analysis obtained via simulation showed that the two parallel mixers used for feedforward operation track each, offering distortion cancellation in spite of the variation. Simulator predicted that over 80% of the fabricated die would fall within ±dB of the targeted IIP3 (see Fig. 5.8 and Fig. 6.6).

7.1. Future work

Research works presented in this dissertation showed that feedforward operation could be utilized to suppress distortion generated by a mixer (switching quad) stage. Although the presented feedforward operation demonstrated superior performance, much work has yet to be done to refine the topology. Listed below are recommendations for future work.

Measured IM3 values for the two individual CMOS switching quads (Fig. 6.12) for various LO swings differ from those predicted by the simulator. For both switching quads, the simulator underestimated the magnitude of their IM3 components and the critical LO swing where the IM3 components experiences a 180° phase shift. It is assumed that these are attributed to limitation of the models and their abilities to predict accurately the distortion components when the transistors are operating in GHz range. This deficiency seems to be most acute when the transistors are operating like switches. Since the amount of IM3 suppression achieved via feedforward operation depends directly on the matching of magnitude and phase of the two IM distortion components, it is necessary that the model reflects accurately the behaviour observed in measurement. A closer collaboration with the modeling engineers will be required where it will ultimately lead to first-time correct designs, reducing the number of design iterations.

In Chapter 3, simulation results showed that IM distortion generated by a switching quad is heavily dependent on the input impedance presented to its common-emitter nodes (nodes X and Y in Fig. 3.1). It was shown via three values that the impedance shunting the common-emitter nodes has a profound influence on the overall distortion performance of the Gilbert mixer. More studies on this phenomena is required as it will facilitate a designer in determining which of the two stages (transconductance or switching quad) is the dominant distortion contributor and applying linearization to the correct stage. Furthermore, this research work mainly focus on how to utilize the low impedance presented to the switching quad as a tool for feedforward-linearization. Further research should also focus on presenting different impedance value for different frequency as suggested for the linear amplifier cases in [60] [61] (i.e., presenting an open-circuit at 2fLo).

Future research should also include designing a receiver that utilizes the feedforward-linearized mixer presented in Chapter 5 and 6 as its first downconverting mixer. This will help to verify that the proposed topology will ultimately offer a wider dynamic range with higher linearity performance. A receiver architecture that will greatly benefit from this feedforward-linearized topology would be the Weaver image-reject receiver. This architecture will allow the feedforward-linearized CMOS mixer and a passive mixer be used as the first and second mixing stage, respectively. This architecture will offer high linearity performance without any compromise in noise performance. This is possible by keeping the IF of the feedforward-linearized CMOS to be in the hundreds of MHz range, mitigating the impact of flicker noise. Operating the second mixer stage in the MHz range allows digital logics such as the static inverter to be used as LO buffer where high LO swings with high slew rate can be readily generated. This setup should ensure that high linearity performance can be obtained from the passive mixer (see Fig. 2.23). If single downconverting architecture is desired, the influence of flicker

Conclusion and recommendations

Future work

Listed below are recommendations for future work.

- Research works presented in this dissertation showed that feedforward operation could be utilized to suppress distortion generated by a mixer (switching quad) stage. Although the presented feedforward operation demonstrated superior performance, much work has yet to be done to refine the topology.
- Measured IM3 values for the two individual CMOS switching quads (Fig. 6.12) for various LO swings differ from those predicted by the simulator. For both switching quads, the simulator underestimated the magnitude of their IM3 components and the critical LO swing where the IM3 components experiences a 180° phase shift. It is assumed that these are attributed to limitation of the models and their abilities to predict accurately the distortion components when the transistors are operating in GHz range. This deficiency seems to be most acute when the transistors are operating like switches. Since the amount of IM3 suppression achieved via feedforward operation depends directly on the matching of magnitude and phase of the two IM distortion components, it is necessary that the model reflects accurately the behaviour observed in measurement. A closer collaboration with the modeling engineers will be required where it will ultimately lead to first-time correct designs, reducing the number of design iterations.
- In Chapter 3, simulation results showed that IM distortion generated by a switching quad is heavily dependent on the input impedance presented to its common-emitter nodes (nodes X and Y in Fig. 3.1). It was shown via three values that the impedance shunting the common-emitter nodes has a profound influence on the overall distortion performance of the Gilbert mixer. More studies on this phenomena is required as it will facilitate a designer in determining which of the two stages (transconductance or switching quad) is the dominant distortion contributor and applying linearization to the correct stage. Furthermore, this research work mainly focus on how to utilize the low impedance presented to the switching quad as a tool for feedforward-linearization. Further research should also focus on presenting different impedance value for different frequency as suggested for the linear amplifier cases in [60] [61] (i.e., presenting an open-circuit at 2fLo).
- Future research should also include designing a receiver that utilizes the feedforward-linearized mixer presented in Chapter 5 and 6 as its first downconverting mixer. This will help to verify that the proposed topology will ultimately offer a wider dynamic range with higher linearity performance. A receiver architecture that will greatly benefit from this feedforward-linearized topology would be the Weaver image-reject receiver. This architecture will allow the feedforward-linearized CMOS mixer and a passive mixer be used as the first and second mixing stage, respectively. This architecture will offer high linearity performance without any compromise in noise performance. This is possible by keeping the IF of the feedforward-linearized CMOS to be in the hundreds of MHz range, mitigating the impact of flicker noise. Operating the second mixer stage in the MHz range allows digital logics such as the static inverter to be used as LO buffer where high LO swings with high slew rate can be readily generated. This setup should ensure that high linearity performance can be obtained from the passive mixer (see Fig. 2.23). If single downconverting architecture is desired, the influence of flicker
Conclusion and recommendations

Noise in the feedforward-linearized mixer can also be compensated at the baseband using analog techniques such as chopping.

As further refinement, a control system that adaptively adjust the bias current and LO swings for the two switching quads can also be incorporated into the receiver design. One application of the control system would be to allow the feedforward operation be turned off when immunity to interference is not required, this will help to reduce the overall power consumption of the receiver and extend the battery life.

Reference


Summary

Title: Low-voltage, Low-power, Feedforward-compensated Active Mixers with Improved Linearity
By: Su-Tam Lim

This thesis presents circuit topologies that employ feedforward-compensation for linearization. Besides having a significant lower distortion profile, these topologies exhibit other attractive factors such as low-voltage operation, broadband operation, and robustness to process variation.

Chapter 2 introduces basic terminology, sources of distortion, and the various type of distortion. These information provided an understanding to how distortion are quantified, how distortion can corrupt the desired signal, and how distortion is generated in transistors and basic amplifier stages. Following the introductory discussion, distortion generated by the Gilbert multiplier, is then presented and discussed in Chapter 3. With the aid of simulation results, the discussions in Chapter 3 show how distortion is generated in the switching quad stage, the input transconductor stage, and the output stages. Existing linearization techniques that are commonly utilized to reduce distortion in radio frequency (RF) circuits are also presented and reviewed in Chapter 3. The analyses describe the limitation of these techniques thereby making them unattractive for applications where high linearity along with low-voltage operation, broadband operation, and robustness to process variation are required.

Chapter 4 presents the circuit topology that utilizes feedforward-compensation to reduce the distortion generated at the output port. Two parallel amplifiers with different local feedback were designed to realize the condition for distortion cancellation. Simulation results showed that this topology offers improvement in linearity and third-order harmonic distortion compared to existing techniques, while requiring only a low bias current. Simulation results also show that this topology exhibits broadband operation, -3 dB bandwidth that is equal to 80% of the f2, and robustness to process variation. The results indicate that this circuit topology is suitable for the input stage of a highly-linear wideband Gilbert mixer.

The next two chapters describe how feedforward-linearization can be used to reduce the distortion generated by the switching quad of a Gilbert mixer. The BJT version is detailed in Chapter 5 while the CMOS variant is discussed in Chapter 6. In both versions, the appropriate selection of LO swing, bias current, and degeneration resistance of two modified switching quads operating in parallel create the conditions for distortion cancellation and higher linearity at the output. Simulation and measured results demonstrate that the feedforward mechanism can suppress intermodulation distortion and improved the linearity, IIP3, by at least 6-7dB.

Chapter 7 lists the conclusions and recommendations of this work. The most important conclusion is that feedforward-linearization technique can be utilized in the design of active mixers with low distortion. This linearization technique introduces a new design approach allowing an RF designer to design a higher linearity mixer with lower power consumption than was possible previously.
Samenvatting

Titel: Low-voltage, Low-power, Feedforward-compensated Active Mixers with Improved Linearity
Door: Su-Tam Lim


Hoofdstuk 2 introduceert basisinformatie zoals terminologie, bron van vervorming, en het diverse type van vervorming. Een detailbespreking over hoe de verschillende types van vervorming worden gekwantificeerd, hoe zij het gewenste signaal bederven, en hoe zij in activie transistors worden geproduceerd en de kriegen ook worden voorgesteld. Na de presentatie van de basisinformatie, werden de mechanisma waarin de vervormingen door de het meest meest gebruikte actieve mixer worden geproduceerd, multiplicator Gilbert, voorgesteld in Hoofdstuk 3. Samen simulatie resultaten, tomen de besprekingen hoe de vervorming in het stadium van de omschakelingsvierling, het stadium van inputtransconductor, en de outputstadia wordt geproduceerd. De bestaande linearisatie technieken werden die algemeen in de ontwikkeling van kriegen worden gebruikt de hoge van de lineariteit radio frequentie (rf) ook voorgesteld en werden herzien in Hoofdstuk 3. De analyses beschrijven ook de beperking van deze technieken die daardoor hen maken voor toepassingen onaantreldcelijk waar de hoge lineariteit samen met zwakstroomverrichting, de breedbandverrichting, en de robuustheid om variatie te verwerken worden vereist.

Hoofdstuk 4 presenteert het circuit topologie die feedforward-compensatie voor de verstoring wordt opgewekt bij de uitgang te verminderen gebruikt. Twee parallelle versterkers met verschillende lokale feedback waren bedoeld om de toestand van vervorming annulering te realiseren. Simulatie resultaten toonden aan dat deze topologie bieden aanzienlijk verbetering van de lineariteit, derde-orde harmonische vervorming, ten opzichte van andere bestaande technieken, terwijl er slechts een lage bias stroom. Simulatie resultaten toonden ook aan dat deze topologie vertoont breedband werking, 3dB bandbreedte die gelijk is aan 80% van de FT, en robuustheid te verwerken variatie. De resultaten toonden aan dat dit circuit topologie leent steeds in de input voor een zeer lineaire wideband Gilbert mixer.

De volgende twee hoofdstukken beschrijven hoe feedforward-linearisatie kan worden gebruikt om de vervorming gegenereerd door het schakelen van een quad Gilbert mixer te verminderen. De BJT versie is opgenomen in hoofdstuk 5, terwijl de CMOS-variant werd besproken in hoofdstuk 6. In beide versies, de juiste selectie in LO swing, de bias stroom en weerstand tegen degeneratie twee gewijzigde schakelen quads die parallel gemaakt het scenario voor de annulering vervorming en hogere lineairiteit bij de uitgang. Simulatie en gemeten resultaten toonden aan dat het stuursignaal mechanisme intermodulatie vervorming te onderdrukken en een verbetering van de lineariteit, IIP₃, door ten minste 6-7dB.

Hoofdstuk 7 bevat de conclusies en aanbevelingen van dit werk. De belangrijkste conclusie is dat feedforward-linearisatie techniek kan worden toegepast in het ontwerp van actieve mixers met lage vervorming. Deze linearisatie techniek introduceert een nieuw design aanpak, waarbij de RF-ontwerper hoge lineariteit mixer met een laag stroomverbruik ontwerp.

Acknowledgements

The work described in this thesis could not have been possible without the contribution and support from many people. I would like to take this opportunity to express my sincere gratitude.

First, I would like to thank my thesis supervisor and promoter, Prof. John Long, for his guidance, encouragement, and the periodic prodding. I will always appreciate the time and patience Prof. Long spent in supervising my work.

Next, I would like to thank my parents for their encouragements during the time when things look bleak and supported my decision, thereby allowing me to complete this dissertation. I would also like to express my appreciation and thank to my wife for her patience and encouragement.

I would also like to thank my fellow graduate students for their valuable and stimulating discussions and an attractive working environment. A special thanks to R. Koo, M. Reiha, and W.L. Chan.

The completion of this thesis would not be possible without the support from the staff of the group. I would like to thank M. de Vosger, W. Straver, L. van Schie, and A. Frehe.

Finally, I would like to thank NXP Semiconductors (previously Philips) for their financial support.