Wideband Up-Converting Mixer for Digital Communications Implemented in 130nm CMOS Technology

By Gabriel Patulea

A Thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Applied Science

in

Electrical and Computer Engineering

Carleton University

Ottawa, Ontario

Dedicated to my parents

Abstract

The drive for lower manufacturing cost has led to the use of CMOS technology for highly integrated radios dedicated to the consumer market. Based on the observation that the wireless devices cover a wide range of frequencies from 700MHz to almost 6GHz, this work proposes the implementation of a wide band up-converting mixer that can be used as a building block in any device that operates in this frequency range. The intent is to cover more than a decade for the output modulated carrier without any internal band switching.

This thesis presents the design methodology and silicon implementation of a 0.13μm CMOS up-conversion mixer for wireless applications. The design approach is simulation based while maintaining the appropriate reference to the theory.

The novelty of this work mainly consists in the implementation of the wide band digital phase shifter on the same chip with the analog mixer/multiplier cells. The design methodology allows a cost effective implementation of a fully integrated solution that requires 1.2x1.55mm of silicon and it can be used in any wireless device that transmits digitally modulated data.

The following components were designed, simulated then integrated on chip: wideband digital phase shifter, two Gilbert cells for the I and Q paths, wideband active balun and biasing cells.

The modular design and individual testing of each sub-block allowed for effective integration and debugging in the initial development phase.

At the end of this exercise, one of the fabricated silicon chips was successfully tested in the lab. Detailed measurement and test results are provided in Chapter 5

Acknowledgement

This thesis would not have been possible without the support of a number of people. The author would like to take this opportunity to thank everyone that helped in his journey towards completing this work.

I would like to express my gratitude to my supervisor, Professor Calvin Plett, for his continuous support and guidance throughout my Masters Degree Thesis elaboration. His assistance while writing this thesis has proven to be invaluable.

I would like to express my gratitude to Professor Dan Ionescu from University of Ottawa for his support and encouragement throughout the journey towards achieving my goals.

Special thanks to my former colleague Diana Gradinaru for sharing her knowledge and providing precious suggestions in the early stages of my design.

I would like to acknowledge and thank my fellow graduate students for sharing their knowledge and offering their advice, whenever I had technical issues; Kimia Ansari, Jerry Lam, William Che Knisely and Thomas Pepler. They were always available and willing to give me their attention and time.

I would like to thank Scott Bruce from computer and network support in the Department of Electronics at Carleton University for his much needed assistance, even in weekends.

Special thanks to the Department of Electronics staff: Anna Lee, Blazenka Power and Sylvie Beekmans for the work they do in the background that no one sees, but it is invaluable in order to keep the department functioning smoothly.

My deepest thanks go to my family for their unconditional love, support and understanding throughout this program.

Last, but not least, I would like to acknowledge the Canadian Microelectronics Corporation (CMC) for their assistance in circuit verification and fabrication.

Table Of Content

List of Tables

List of Figures

Abbreviations

1 dB CP	Input Referred one dB Compression Point
AF	Audio Frequency
AMP	Amplifier
BALUN	Balanced Unbalanced
BB	Baseband
BER	Bit Error Rate
BPF	Band Pass Filter
dB	Decibels
dBm	Power level in dB (decibels) with respect to 1 mW
DET	Detector
DRC	Design Rules Check
DRV	Driver (as amplifier)
DUT	Device Under test
EDGE	Enhanced Data Rate
EDGE	Enhanced Data Rate for GSM Evolution
ESD	Electrostatic Discharge
ETSI	European Telecommunication Standards Institute
FDD	Frequency Division Duplex
FM	Frequency Modulation
FSK	Frequency Shift Keying
GC	Power / Voltage Conversion Gain
GMSK	Gaussian Minimum Shift Keying
GSM	Global System for Mobile communications
HSPA	High Speed Packet Access
IF	Intermediate Frequency
IIP3	Input Referred 3rd Order Intercept Point
IM	Intermodulation Products
IMD	Inter Modulation Distortion
IMG	Image (Related to Image Filter or Image Frequency)
IQ	Inphase Quadrature

LFSR	Linear Feedback Shift Register
LO	Local Oscillator
LTE	Long Term Evolution
LVS	Layout Vs. Schematic
MIMO	Multiple Input – Multiple Output
MIX	Mixer
MOS	Metal Oxide Semiconductor
MTBF	Mean Time Between Failures
OFDMA	Orthogonal Frequency Division Multiple Access
PA	Power Amplifier
PAC	Periodic AC analysis
PNoise	Periodic Noise
PSK	Phase Shift Keying
PSP	Periodic S Parameters
PSS	Periodic Steady State
QPAC	Quasi Periodic AC Analysis
QPSS	Quasi Periodic Steady State
RF	Radio Frequency
RFIC	Radio Frequency Integrated Circuit
Rx	Receiver
SBS	Side Band Suppression
SC-DSB	Suppressed Carrier - Double Side Band
SCR	Silicon Controlled Rectifier
SDF	Spectral Density Function
SNR	Signal to Noise Ratio
SSB	Single Side Band (Related to the type of modulation)
SSB NF	Single Side Band Noise Figure
TA	Test Attenuator
TDD	Time Division Duplex
TPG	Test Pattern Generator
Tx	Transmitter
UMTS	Universal Mobile Telecommunications System
WCDMA	Wideband Code Division Multiple Access

WiFi	Synonymous to WLAN
WLAN	Wireless Local Area Network

1 Introduction

The last few decades have shown an unprecedented growth in digital wireless communication systems, almost exclusively fuelled by the widespread adoption of cellular mobile radio systems and wireless local area networks (WLAN). With the deployment of 3G cellular technologies since the years 2000 and more recently High Speed Packet Access (HSPA), there has been a dramatic growth globally in both the number of subscribers and the volume of data traffic transported by these sophisticated wireless systems. Complementing this growth in demand, there has been a consistent advancement of the communication theories, enabling technologies and standards that cover these new, complex, wireless, multiple access, RF systems. Currently the commercial deployment of the Long Term Evolution (LTE) cellular mobile radio system is underway. LTE is an all IP based radio access network that uses MIMO/OFDMA technology to achieve peak data rates up to 300Mbit/s. Currently, the standardization of LTE-Advanced is underway which plans data rates up to 1 Gbit/s for low mobility users. Concurrently, 5G wireless communication systems, such as massive MIMO, energy-efficient communications, cognitive radio networks, and visible light communications are being studied and solutions proposed for the constantly increasing user data traffic.

The high demand and rapid pace of development of spectrally efficient, and now energy efficient, wireless communication systems opens new opportunities for research. The wireless devices designed these days are capable of supporting a diverse range of applications needing high bandwidths and differentiated quality of service. The new opportunities have created a strong and vibrant global market for communications engineers needing specialist skills in wireless systems.

1.1 **Motivation and Objectives**

As the price competition in the communications industry continues, the advantages of highly integrating the RF sub-blocks of wireless communication devices circuits are compelling. Aggressive scaling of CMOS devices usually targets the high density digital components and it does not always benefit the performance of RF circuits. However, the progressively smaller submicron CMOS devices provides an increasingly higher f_T and f_{max}, which enable the realization of RF front-ends being able to perform in the GHz frequency range to be integrated on a single chip. As the devices become smaller, there is a similar desire to reduce the number of passive components on board. This approach will lead to smaller board size and usually lower power consumption. Furthermore, the reliability and controllability of the end products using single chip solutions is improved significantly because the process variation of integrated devices is smaller than that of the discrete devices on board.

The frequency range where portable communication devices are licensed is very broad, starting around 700MHz and ending up to 6GHz. This entire spectrum covers from cellular telephony and Bluetooth devices up to the 5GHz wireless LAN (WiFi). The concept of high level of integration led to the idea of building a modulator device that is able to cover the whole range of UHF frequencies without band switching. The device usage is not restricted to mobile devices only. It can be used as well in access point transmitters, Bluetooth, ZigBee, pico-cell base stations and other wireless technologies by applying the appropriate baseband signal.

Since the I/Q modulation method is used, the main challenge is to generate the in-phase, and quadrature phase signals of the RF carrier over a wide frequency range with sufficient accuracy, such that the signal quality at the mixer/modulator output meets the industry standards.

The objective of this thesis work is to explore and design a wideband up-conversion mixer in 0.13μm CMOS technology. This mixer is designed to cover from 500MHz to 6GHz modulated output frequency while generating the RF I/Q signals on chip. As a result, a single differential local oscillator (LO) signal is required to be applied at the RF input of the chip. The LO frequency needs to be twice the output carrier frequency in order to allow the proper operation of the internal 90˚ phase shifter.

1.2 **Thesis Outline**

This thesis is organized to provide readers with a step-by-step introduction to the mixing process used in a transmitter, its requirements and implementation issues. It starts with a brief presentation of historical modulation methods and attempts to increase the bandwidth efficiency as a natural transition towards the digital modulation process. Then, gradually elements are introduced specific to the digital modulation and circuit implementation.

Chapter 1 starts with an introduction and motivation to this thesis work. It briefly mentions the fast paced evolution of the wireless communication field as well as the necessity of improving the level of component integration in modern devices.

Chapter 2 provides a literature review and details about the theoretical background of modulation methods. It provides the reader with a smooth transition from the historical AM systems towards the I/Q modulation concepts. Chapter 2 ends with an in depth analysis of the impact that the circuit impairments have on the modulated signal quality.

Chapter 3 analyzes various functional blocks available to the designer to implement the signal processing required in an I/Q modulator. There are separate sections that present phase shifters, multipliers and their performance. Advantages and disadvantages of each solution are being scrutinized. The chosen solution for this work is analyzed and presented in greater detail.

Chapter 4 provides information regarding the circuit implementation. It starts with the detailed analysis of the active devices, including the effects of scaling on the transition frequency, transconductance and noise. The passive elements follow. Eventually, the complete schematics of the circuit sub-blocks are shown together with calculations and simulations. The focus is on the CML D type flip-flop as basic block for the wideband phase shifter, the Gilbert cell used as multiplier and the active balun that converts the differential modulated signal to a single ended version that is routed to the chip output.

Gain, noise figure and linearity are part of the circuit simulation and characterization.

Elements of reliability included in the design such as ESD and latch-up protection are discussed.

Chapter 5 presents the test bench setup and the actual measurements taken on a fabricated chip. Where appropriate, a small “Discussions” section is added in order to clarify certain measurement methods and results.

Chapter 6 concludes the current work and it consists of brief description of the achievements and opportunities for improvement. It also outlines the direction of future work that may be required to improve the performance of the design.

2 Literature Review

Mixing is one of the most fundamental operations performed in communication systems. Whether it is performed in the analog or digital domain, the essence is the same. Mixing is actually a multiplication in the time domain that allows the frequency translation of baseband signals to a higher frequency; then the signal can be radiated in space using an antenna.

Mixing is also used in receiving devices. Historically, communication systems have used analog mixing techniques. From the very first heterodyne receiver principle invented by Edwin H Armstrong, mixing is used to translate the high frequency signals from the antenna to an intermediate frequency, where most of the signal processing is taking place. High gain and selectivity are implemented in a fixed frequency amplifier and filter. The fixed intermediate frequency, narrow band amplifier is easier to build and tune, compared to a wideband amplifier.

The block diagram of a traditional SSB communication system is shown in Figure 2.1 below.

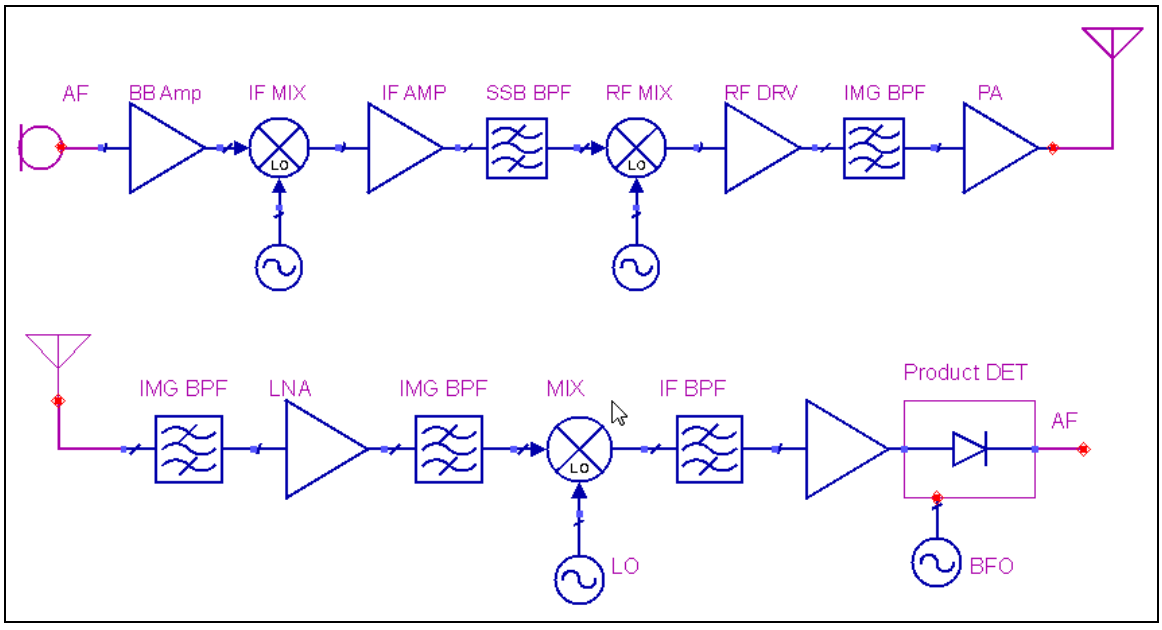

Figure 2.1 Block Diagram of a Traditional SSB Communication System

2.1 Math behind mixing

2.1.1 Simple AM systems

What is called mixing in the radio frequency (RF) field is essentially a multiplication in the time domain. Let's assume we have two signal generators, one of them being operated at the carrier frequency and the other generating the modulating, or the baseband frequency. To keep the demonstration simple, we will assume that the modulating frequency is a sine wave, the same as the carrier frequency. We note the two frequencies f_c and f_m and we make the assumption that f_c is much higher than f_m. The signals from the two generators can be represented as follows:

The carrier signal: $s_c(t) = A_c \cdot \cos(2\pi f_c \cdot t)$

The modulating signal: $s_m(t) = A_m \cdot \cos(2\pi f_m \cdot t)$

The time domain multiplication of the two signals becomes:

$$s_{AM}(t) = A_c \cdot A_m \cdot \cos(2\pi f_c \cdot t) \cdot \cos(2\pi f_m \cdot t)$$

Where the AM subscript stands for Amplitude Modulation.

After applying simple trigonometric identities, we get:

$$s_{AM}(t) = \frac{A_c \cdot A_m}{2} \cdot \left(\cos(2\pi(f_c - fm) \cdot t) + \cos(2\pi(f_c + fm) \cdot t)\right) \quad (1)$$

It can be noted that the multiplication process generated two frequencies that were not present in either of the carrier or modulating signal. One way to interpret the process is that the low frequency modulating signal was translated around the high frequency carrier spectrum.

Although the calculation was shown using a single, discrete sine wave for the modulating (or baseband) signal, it can be easily shown that a wider spectrum of the modulating signal would be translated entirely around the carrier frequency.

It is worth to mention that the general shape of the modulating spectrum is preserved after the frequency translation only in this case of amplitude modulation. Frequency, or angle modulation is governed by different rules and it is outside the scope of this work.

The simple calculation above shows the *suppressed carrier* version of the amplitude modulation. Historically, the AM transmitters used for broadcasting were required to transmit the carrier as well. At the expense of the power transmitted on the carrier frequency, the mass produced receivers were simple and low cost. Modeling of the broadcast AM modulation requires a DC component to be added to the modulating signal, as follows:

$$s_{C\_AM}(t) = A_m \cdot (1 + \cos(2\pi f_m \cdot t)) \cdot A_c \cos(2\pi f_c \cdot t) \text{ and further}$$

$$s_{C\_AM}(t) = A_c \cdot A_m \cdot \left[\cos(2\pi f_c \cdot t) + \frac{(\cos(2\pi(f_c - fm) \cdot t) + \cos(2\pi(f_c + fm) \cdot t))}{2} \right] \quad (2)$$

In this case, the carrier component $\cos(2\pi f_c \cdot t)$ is present at the output. The main disadvantage of the carrier AM is that half of the available power at the TX output is wasted in the carrier and only one quarter in each side band. It should be noted that the carrier alone is a deterministic signal and it does not carry any information, therefore, in an efficient communication system it should not be transmitted. This is, however, the price paid for the simplicity and low cost at the AM receiver end.

The time domain signals and their associated spectra for an AM signal are shown in Figure 2.2 below.

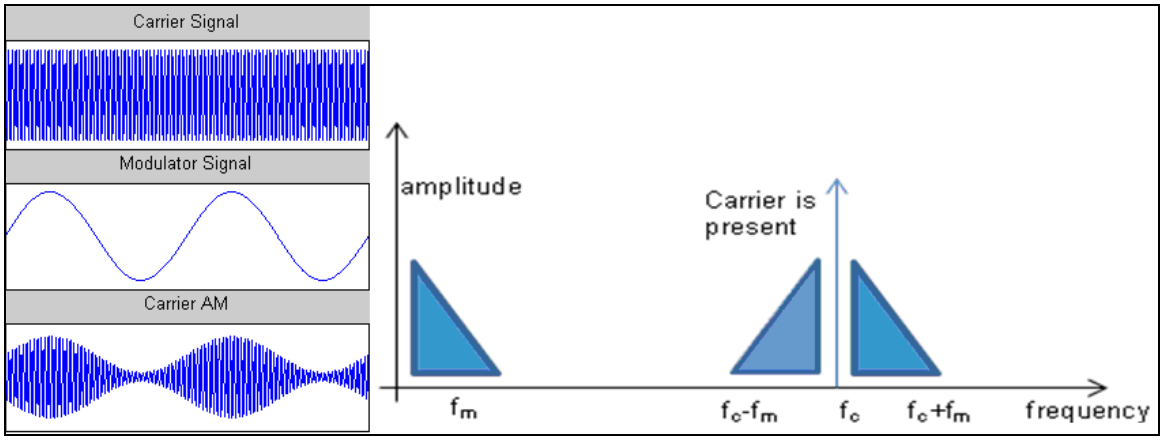

Figure 2.2 Amplitude Modulated signal and its associated spectrum

2.1.2 Single Side Band (SSB) AM; a step towards spectral efficiency

It can be shown that only one of the sidebands is required in order to completely recover the original information from the baseband. In the case of carrier AM, the available power that can be delivered by the transmitter's final stage needs to be shared between the main carrier and the two side bands. Estimating the power in each spectral component, we notice that only ¼ of the total power is used to transmit one of the sidebands. If only a sideband would be transmitted, then the whole available TX power could be used to carry information. Hence, this leads to the idea of single sideband modulation where only a single replica of the original baseband spectrum would be generated and amplified; no carrier and no additional sideband would be transmitted.

2.1.3 SSB Demodulation

A demodulation example will be shown as a simple justification that all the baseband information can be recovered from one of the sidebands. Let's assume that the upper sideband (USB) is available from the example above. We can write the USB signal as:

$$s_{USB}(t) = \frac{A_c \cdot A_m}{2} \cdot \cos(2\pi(f_c + fm) \cdot t)$$

By multiplying the USB signal with a local replica of the carrier frequency, generated at the receiver end, we can write:

$$s_{DEM}(t) = \cos(2\pi \cdot f_c \cdot t) \cdot \cos(2\pi(f_c + fm) \cdot t) = \frac{1}{2} \cdot \left(\cos(2\pi \cdot f_m \cdot t) + \cos(2\pi \cdot 2f_c \cdot t)\right) \text{ and}$$

$$s_{DEM\_FILT}(t) = \frac{1}{2} \cdot \cos(2\pi \cdot f_m \cdot t)$$

The demodulated signal contains the original modulating signal and a high frequency component that is twice the carrier frequency. The $2f_c$ component can be easily filtered out based on the fact that its frequency is much higher than the baseband signal. The am-

plitude of the signal was intentionally omitted in order to illustrate the point that the baseband frequency signal is recovered.

2.2 Classic methods for generating Single Side Band signals

2.2.1 Filtering Method

The filtering method is schematically shown in Figure 2.3 below.

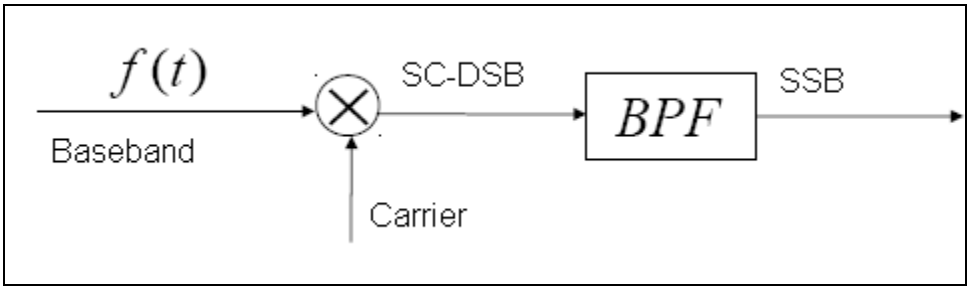

Figure 2.3 Filtering method for generating SSB signal

The Baseband signal is multiplied in time domain with the carrier signal. After multiplication the resulting signal becomes Suppressed Carrier – Double Side band (SC-DSB). Upon filtering using a narrow band pass filter (BPF), one of the side bands is being removed and the signal contains only one of the side bands.

Usually the Single Side Band (SSB) signal generation takes place at some conveniently chosen intermediate frequency. A second mixer and image rejection filter may be used in order to generate the RF signal that is applied to the antenna.

The main disadvantage of the method is the requirement of the BPF. This filter is usually crystal based and integration and miniaturization are difficult.

2.2.2 Phasing Method (Hartley Modulator)

At the expense of adding circuit complexity, the phasing method no longer requires the large crystal based BPF.

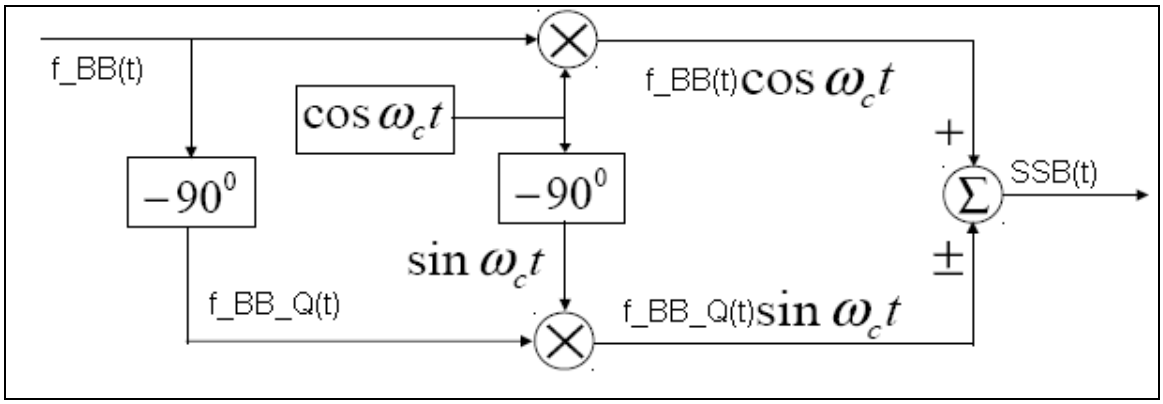

Figure 2.4 Phasing Method for generating SSB

The phasing method for generating SSB signals relies on the availability of two different SC-DSB sets of signals and adding or subtracting them together. Depending on the addition or subtraction performed at the end of the processing chain, the result is either the USB or the LSB respectively (see Figure 2.4).

One of the advantages of the phasing method is that there is no need for additional mixing (heterodyning) to generate the SSB signal at the RF frequency.

Another advantage is that high Q, narrow band filters are not required in the process.

However, it is difficult in practice to maintain an accurate phase relationship across the bandwidth of the modulating signal. The extent to which the sideband suppression is dependent on the phase and amplitude relationship between the signals will be investigated in section 2.5.

2.2.3 Weaver Method

The Weaver method for SSB signal generation is schematically shown in Figure 2.5 below.

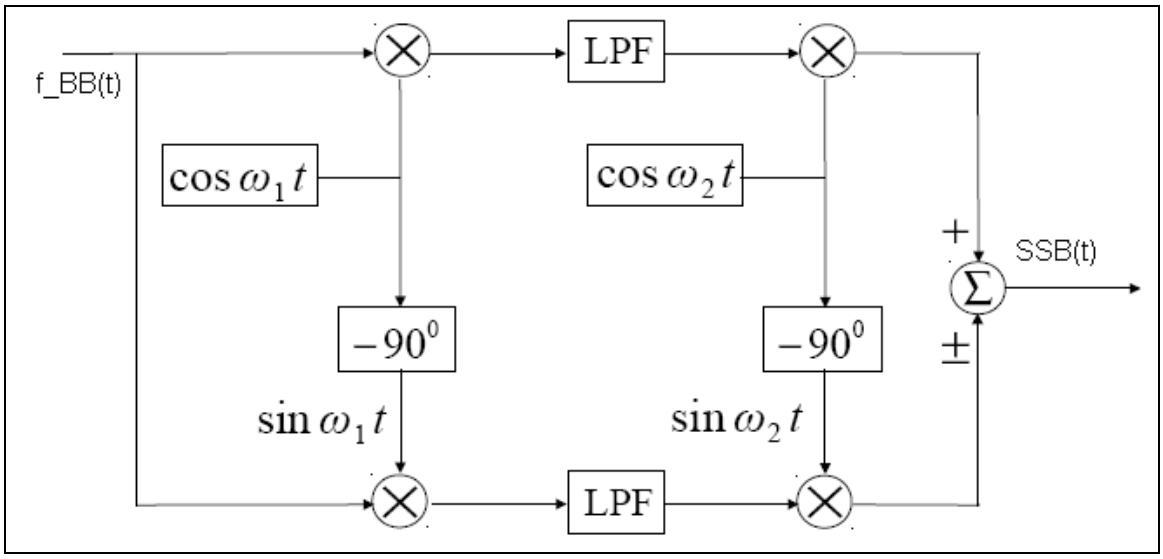

Figure 2.5 Weaver Method for generating SSB signals

Weaver`s method of SSB generation, is to some extent similar to the phasing method. Its proper operation depends upon phase cancellation of the two SC-DSB signals. Its improvement over the filtering and phasing methods are:

- It does not require wideband phase shifting networks (Hilbert transformer), like the phasing method of SSB generation.
- There is no need for the sharp cut-off filter, operating far from baseband, as does the filtering method of SSB generation.
- Unwanted components due to circuit impairments that are not fully removed (by phasing, as in the phasing method, or by imperfect filtering, as in the filter method) fall inside the SSB channel itself, therefore they do not cause interference to the adjacent channels.

It is important to note that the second pair of quadrature multipliers should, ideally, be DC coupled. In practice DC coupling is more often avoided, since it may result in DC-offset problems. If DC coupling is not achievable, there can be a small gap in the translated baseband spectrum located at one half of the baseband signal bandwidth.

The Weaver method is easily implemented in digital modulators for SSB generation.

2.3 **Transition towards digital modulation**

While the AM modulation methods shown above were used for many years in broadcasting equipment, they were still performed in the analog domain and the performance level became too low for communication systems where bandwidth and power efficiency are important.

The fundamental tradeoff in communication systems is that simple hardware can be used in transmitters and receivers to communicate information. But simple methods require a lot of spectrum which limits the number of users of the spectrum. Alternatively, more complex, bandwidth efficient transmitters and receivers can be used to transmit the same information over less bandwidth. There is a price to pay for this. More and more spectrally efficient transmission techniques require complex hardware. Complex hardware is difficult to design, test, and build. This tradeoff exists whether communication is over air or wire, analog or digital.

The RF spectrum is a limited resource and it must be shared between its users. The modern society is addicted to communications and every day there are more users for that spectrum as demand for communications services increases. Digital modulation schemes provide a cost effective option to convey a much larger amount of information compared to the analog modulation schemes.

2.4 **Fundamental Principles of Complex Modulation**

Let's assume that there are two arbitrary base band signal sources i(t) and q(t) and we are interested to translate them around a high carrier frequency denoted ω_c in order to transmit them to a receiving point. For this purpose, a high frequency signal source is required as well. For the high frequency carrier we assume that both in phase - $\cos(\omega_c t)$ - and in quadrature – $j \cdot \sin(\omega_c t)$ components are available.

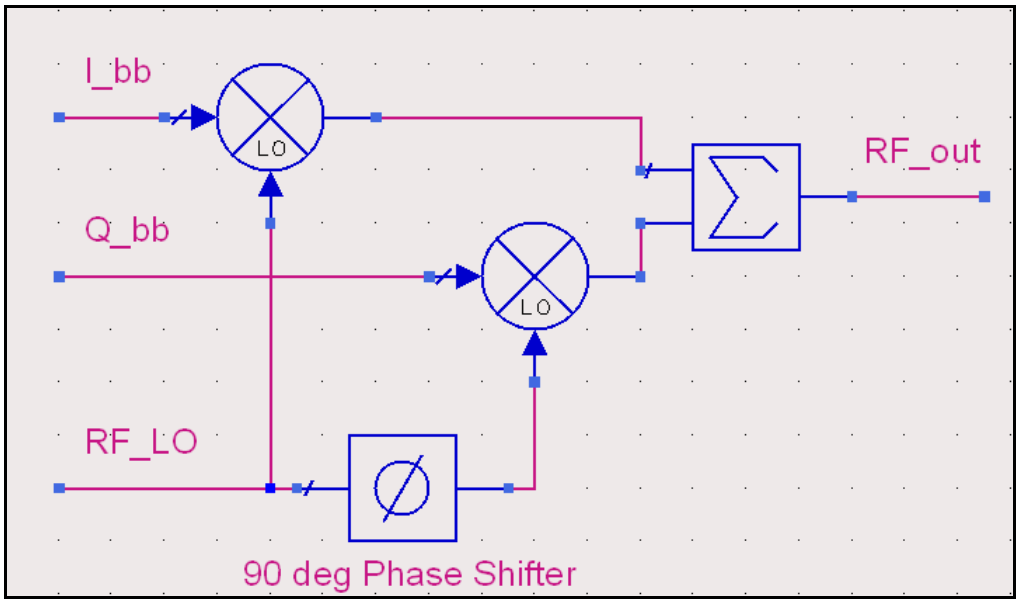

Figure 2.6 System diagram of complex modulation principle

Multiplying the "in phase" RF carrier with i(t) and the "in quadrature" RF carrier with q(t) and then combining them together results in:

$$s(t) = i(t) \cdot \cos(\omega_c t) + j \cdot q(t) \cdot \sin(\omega_c t) \tag{3}$$

This is the signal that will be fed to the transmitting antenna and received at the destination point.

At the receiver end, the signal is split and it follows two paths. Similar to the transmit side, at the receive point one path is multiplied with the in phase and the other with the in quadrature local replica of the carrier signal. Its frequency must be ω_c as well. Mathematically, the two multiplications taking place inside the receiver can be described as follows:

In phase:

$$\begin{aligned} r_i(t) &= s(t) \cdot \cos(\omega_c t) = \left[i(t) \cdot \cos(\omega_c t) + j \cdot q(t) \cdot \sin(\omega_c t) \right] \cdot \cos(\omega_c t) = \\ &= i(t) \cdot \cos^2(\omega_c t) + j \cdot q(t) \cdot \sin(\omega_c t) \cdot \cos(\omega_c t) = \\ &= 0.5 \cdot i(t) + 0.5 \cdot \cos(2\omega_c t) + j \cdot 0.5 \cdot q(t) \cdot \left(\sin(2\omega_c t) + \sin(0) \right) \end{aligned} \tag{4}$$

Quadrature phase:

$$\begin{aligned} r_q(t) &= s(t)\cdot j\cdot\sin(\omega_c t) = \left[i(t)\cdot\cos(\omega_c t) + j\cdot q(t)\cdot\sin(\omega_c t)\right]\cdot j\cdot\sin(\omega_c t) = \\ &= i(t)\cdot\cos(\omega_c t)\cdot j\cdot\sin(\omega_c t) - q(t)\cdot\sin^2(\omega_c t) = \\ &= 0.5\cdot i(t)\cdot\left(\sin(2\omega_c t) + \sin(0)\right) - 0.5\cdot q(t)\cdot\left(1-\cos(2\omega_c t)\right) \end{aligned} \quad (5)$$

Inspecting the in phase and in quadrature expressions at the receive end, we note that there are both high frequency components as well as the original baseband signals i(t) and q(t). By applying low pass filtering on both received signals, we get:

$$r_i(t) = 0.5\cdot i(t) \quad and \quad r_q(t) = -0.5\cdot q(t) \quad (6)$$

The derivation above shows that the initial baseband signals that were used for the IQ modulation can be recovered at the receive point. While the derivation was done using arbitrary baseband signals i(t) and q(t), a special case can be devised in order to gain a better insight of the performance requirements of the complex modulation. In the next section we will take a look at the special case where the baseband signal is a complex sine wave defined by both in phase and in quadrature components.

2.5 **IQ modulation using a complex sine wave in the baseband**

In cases where the baseband signals are no longer arbitrary and they are defined as sine waves as follows:

$$i(t) = \cos(\omega_B t) \quad and \quad q(t) = \sin(\omega_B t) \quad (7)$$

The transmitted signal becomes:

$$s(t) = \cos(\omega_B t)\cdot\cos(\omega_c t) + j\cdot\sin(\omega_B t)\cdot\sin(\omega_c t) \quad (8)$$

Let's assume now that the baseband signals are not exactly matched in amplitude and in phase. We want to find out the impact of the amplitude and phase mismatch on the sideband rejection. To derive the rejection, we assume that the in phase signal cumulates both the amplitude and the phase errors. The baseband signals are written as:

$$i(t) = A\cdot\cos(\omega_B t + \phi) \quad and \quad q(t) = \sin(\omega_B t) \quad (9)$$

where we noted A=the ratio between the magnitude of the in phase and in quadrature signal and Φ is the phase error from an ideal 0˚ phase. Using the notations above, the RF signal, s(t), generated by the modulator becomes:

$$s(t) = A \cdot \cos(\omega_B t + \phi) \cdot \cos(\omega_c t) + j \cdot \sin(\omega_B t) \cdot \sin(\omega_c t) \tag{10}$$

Using the trigonometric identities:

$$\begin{aligned} \sin(\alpha) \cdot \cos(\beta) &= \frac{1}{2} \cdot \sin(\alpha - \beta) + \frac{1}{2} \cdot \sin(\alpha + \beta) \\ \cos(\alpha) \cdot \cos(\beta) &= \frac{1}{2} \cdot \cos(\alpha - \beta) + \frac{1}{2} \cdot \cos(\alpha + \beta) \\ \cos(\alpha + \beta) &= \cos(\alpha) \cdot \cos(\beta) - \sin(\alpha) \cdot \sin(\beta) \end{aligned} \tag{11}$$

The RF signal equation can be re-written such that the USB and the LSB are clearly expressed.

$$\begin{aligned} USB(t) &= \frac{1}{2} \cdot A \cdot \cos((\omega_c + \omega_B)t) \cdot \cos(\phi) - \frac{1}{2} \cdot A \cdot \sin((\omega_c + \omega_B)t) \cdot \sin(\phi) - \frac{1}{2} \cdot \cos((\omega_c + \omega_B)t) \\ LSB(t) &= \frac{1}{2} \cdot A \cdot \cos((\omega_c - \omega_B)t) \cdot \cos(\phi) - \frac{1}{2} \cdot A \cdot \sin((\omega_c - \omega_B)t) \cdot \sin(\phi) + \frac{1}{2} \cdot \cos((\omega_c - \omega_B)t) \end{aligned} \tag{12}$$

The equations need to be converted in envelope-phase format because the sideband suppression is the ratio between the magnitude of the desired sideband and the magnitude of the unwanted sideband. In general, a complex number, c, expressed in real-imaginary format can be converted in magnitude-phase format based on the following equations:

$$\begin{aligned} &c = a + j \cdot b \quad \textit{in real} - \textit{imaginary format} \\ &c = r \cdot \cos(\varphi) \quad \textit{where } r = \sqrt{a^2 + b^2} \textit{ and } \varphi = \arctan\left(\frac{b}{a}\right) \end{aligned} \tag{13}$$

Applying (13) to the USB and LSB signals, we get the envelopes of the two sidebands:

$$\begin{aligned} USB_{mag} &= \sqrt{\left(\frac{1}{2} \cdot A \cdot \cos(\phi) - \frac{1}{2}\right)^2 + \left(-\frac{1}{2} \cdot A \cdot \sin(\phi)\right)^2} = \sqrt{\frac{1}{4} \cdot A^2 - \frac{1}{2} \cdot A \cdot \cos(\phi) + \frac{1}{4}} \\ LSB_{mag} &= \sqrt{\left(\frac{1}{2} \cdot A \cdot \cos(\phi) + \frac{1}{2}\right)^2 + \left(-\frac{1}{2} \cdot A \cdot \sin(\phi)\right)^2} = \sqrt{\frac{1}{4} \cdot A^2 + \frac{1}{2} \cdot A \cdot \cos(\phi) + \frac{1}{4}} \end{aligned} \tag{14}$$

The ratio of the two envelopes can be written as:

$$\frac{USB_{mag}}{LSB_{mag}} = \sqrt{\frac{\frac{1}{4}\cdot A^2 - \frac{1}{2}\cdot A\cdot\cos(\phi) + \frac{1}{4}}{\frac{1}{4}\cdot A^2 + \frac{1}{2}\cdot A\cdot\cos(\phi) + \frac{1}{4}}} = \sqrt{\frac{A^2 - 2\cdot A\cdot\cos(\phi) + 1}{A^2 + 2\cdot A\cdot\cos(\phi) + 1}} \tag{15}$$

And in decibel format:

$$SBS[dB] = 20\cdot\log\left(\frac{USB_{mag}}{LSB_{mag}}\right) = 10\cdot\log\left(\frac{A^2 - 2\cdot A\cdot\cos(\phi) + 1}{A^2 + 2\cdot A\cdot\cos(\phi) + 1}\right) \tag{16}$$

The calculation above shows an estimate of the maximum allowable amplitude and phase impairments for a specific sideband suppression. The following plots are based on (16).

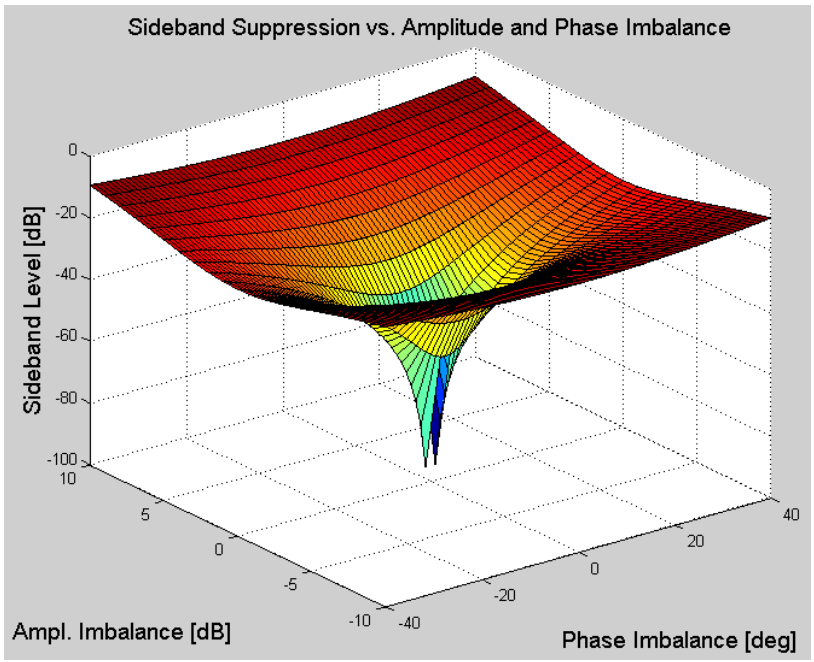

Figure 2.7 3D plot showing sideband level versus Amplitude and Phase Imbalance

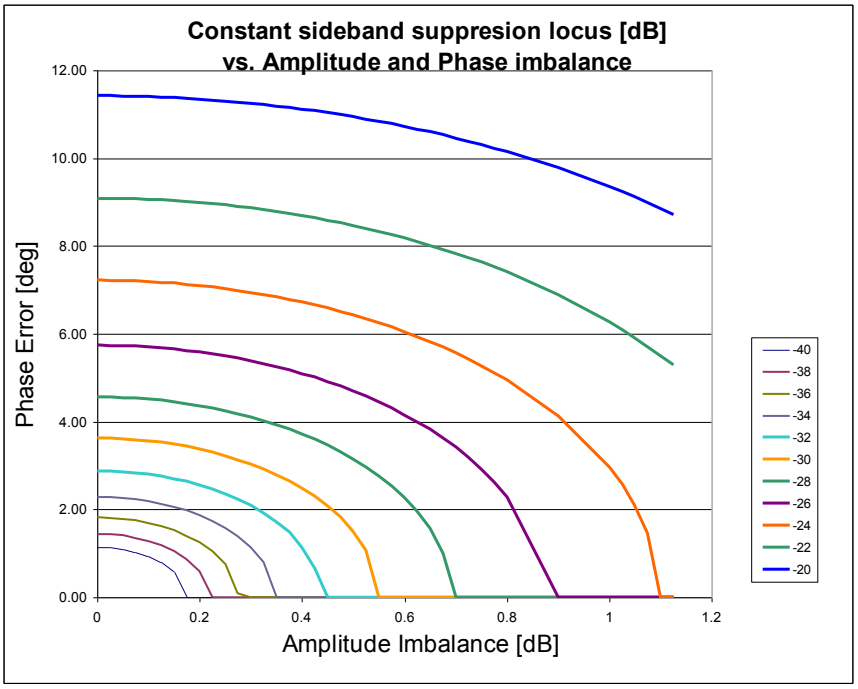

Figure 2.8 Constant sideband suppression circles vs. Amplitude and Phase Imbalance

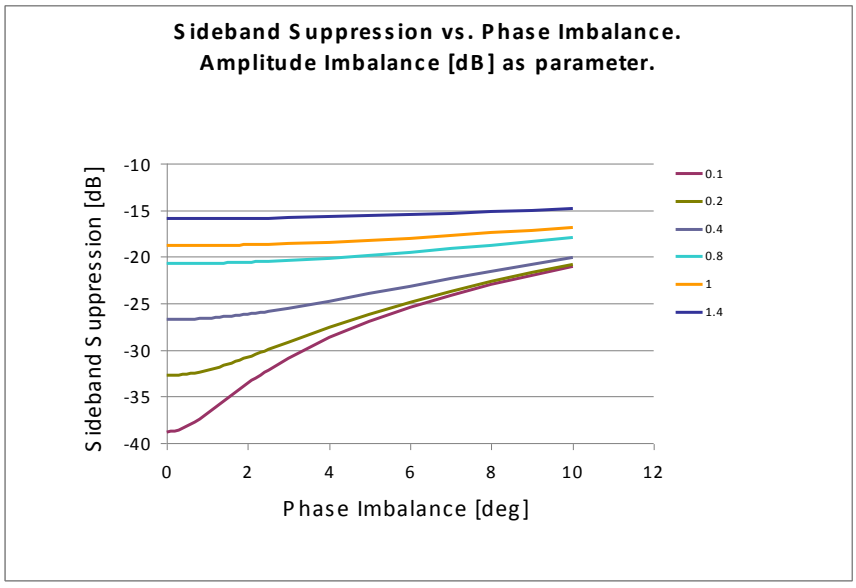

Figure 2.9 Sideband Suppression vs. Phase Imbalance. Amplitude Imbalance [dB] as parameter

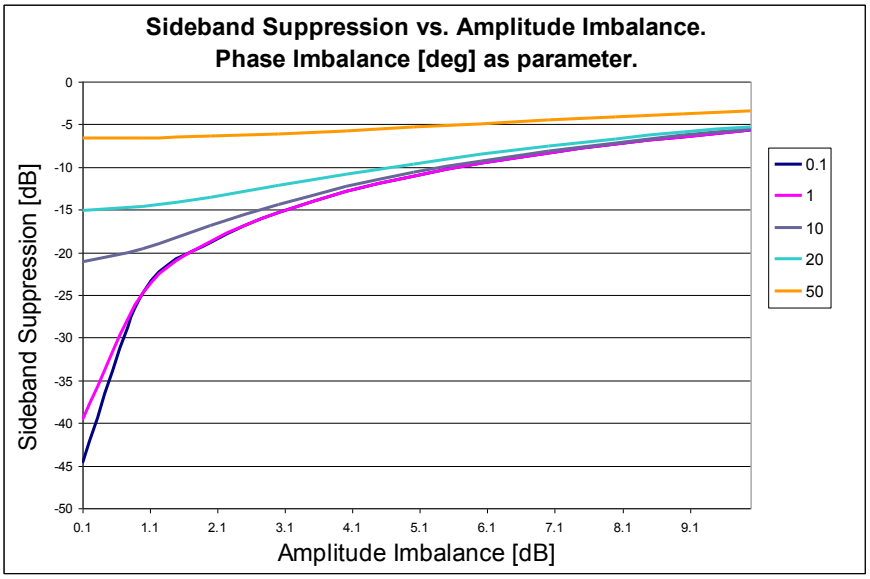

Figure 2.10 Sideband Suppression vs. Amplitude Imbalance. Phase Imbalance [deg] as parameter.

From the plots shown above, one can note the importance of good amplitude and phase balance between the baseband and carrier signals. For instance, assuming an ideal amplitude balance and arbitrarily choosing a sideband suppression of 40dB, the phase error can't be higher than 1.15˚. Conversely, assuming that there is no phase imbalance, the maximum tolerated amplitude ratio error should be less than 0.1dB in order for the sideband suppression to be 40dB or better.

Good sideband suppression is critical especially if the complex modulation method is being used for multilevel (m-ary) data transmission. Insufficient attenuation of the unwanted sideband has a direct impact on the Error Vector Magnitude (EVM), a frequently used measure for characterizing data transmitters.

2.6 **Data transmission using IQ modulation**

A carrier signal is completely defined by its amplitude, frequency and phase. Therefore, there are three opportunities to apply modulation on any of these components alone or

combined. One note though: phase and frequency are inter-related and they can be interpreted as different views of the same signal parameter change.

While analyzing digital modulation, it is convenient to represent the signal as a vector in a polar diagram. The polar diagram allows a fairly intuitive representation of the amplitude and the phase of the signal vector. The magnitude of the vector can be expressed in either absolute or relative value and the phase must be represented relative to a reference signal, usually the carrier. However, it is not uncommon to describe the signal vector by its Cartesian coordinates of in-phase, I, and in-quadrature, Q.

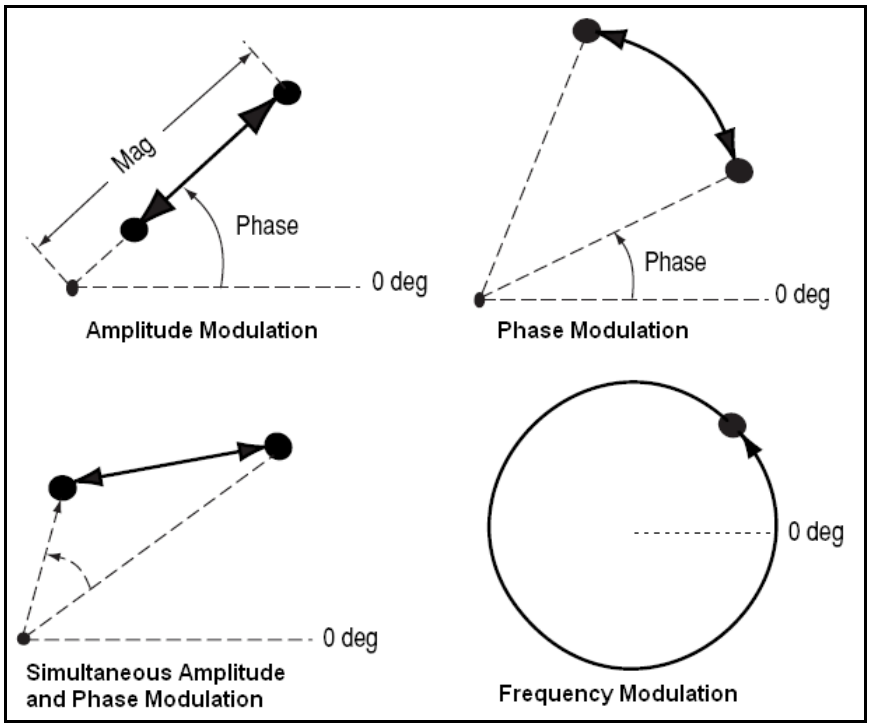

Figure 2.11 Basic representation of amplitude, phase and frequency modulation

Figure 2.11 above shows a graphical representation of each of the three parameters of the signal as well as the meaning of modulating one of the parameters or two parameters combined. In digital communications each possible position of the black dot is interpreted as a symbol that can carry one or more bits of information depending on the type of modulation. The graphical representation of all the defined symbols in a Cartesian IQ plane is also known as the symbol constellation. The symbol constellations for some of

the frequently used modulation schemes used in data communication are shown in Figure 2.12.

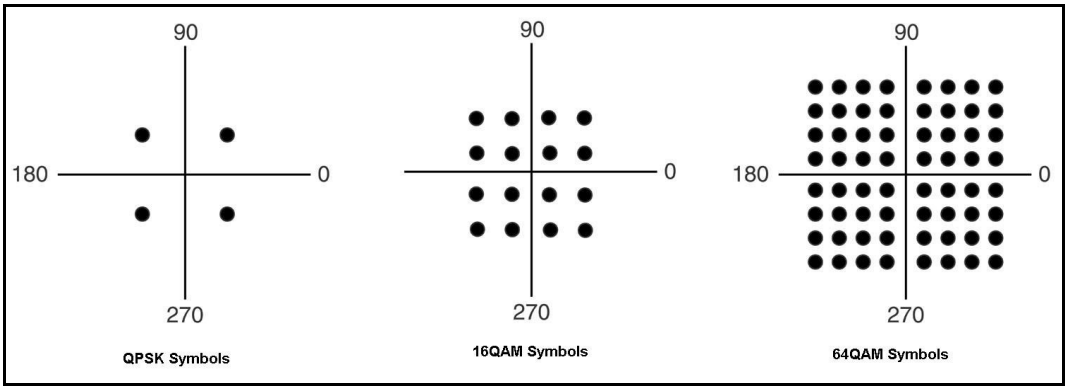

Figure 2.12 Symbol constellations for QPSK, 16QAM and 64QAM

2.7 **I/Q imbalance and noise impact on EVM**

EVM is a widely accepted measure of the quality of a digitally modulated signal. EVM is defined as shown in (17)

$$EVM = \frac{RMS\ magnitude\ of\ observed\ error\ vectors}{Magnitude\ of\ \ farthest\ vector in\ the\ constellation} \times 100$$

$$EVM = \frac{\sqrt{\frac{1}{N} \cdot \sum_{j=1}^{N} \left((I_j - Iideal_j)^2 + (Q_j - Qideal_j)^2 \right)}}{S_{\max}} \times 100 \qquad (17)$$

where we noted:

- I_j, Q_j, $Iideal_j$, and $Qideal_j$ are the coordinates of the received and ideal vectors, respectively, at the jth time interval
- S_{max} is the magnitude of the vector located at the farthest point in the constellation
- N is the total number of analyzed symbols

Figure 2.13 depicts the graphical meaning of an individual error vector measurement. It is quite intuitive that the more symbols are in a constellation, the more susceptible the

transmission is to noise and impairments. As the distance between two adjacent symbols decreases, a higher SNR is required in order to meet the system Bit Error Rate (BER) performance.

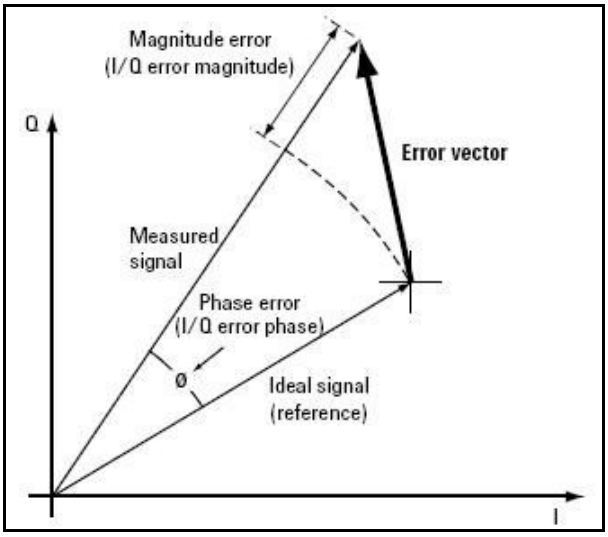

Figure 2.13 Representation of the Error Vector Magnitude

According to reference [1], based on the assumption that either the magnitude or phase imbalance of the IQ modulator is small, an approximate value of the EVM can be determined as follows:

$$EVM \approx \sqrt{\frac{1}{SNR} + 2 \cdot \left(1 - \sqrt{\frac{1}{1+SBS}}\right)} \tag{18}$$

Where we noted:

- SNR - the Signal to Noise Ratio of the digitally modulated signal
- SBS – the Side Band Suppression due to modulator impairments

Figure 2.14 shows the estimated variation of EVM vs. the Side Band Suppression, having the SNR as parameter. The plot shows that both the noise and the modulator impairments must have a low level in order to achieve EVM figures of less than 10%.

For illustration purposes, a few simulations were implemented in Matlab that are meant to show the impact of magnitude and phase imbalance as well as noise on the quality of the modulated signal. The ideal signal is shown as blue dots as a reference, while the red dots represent the distorted signal.

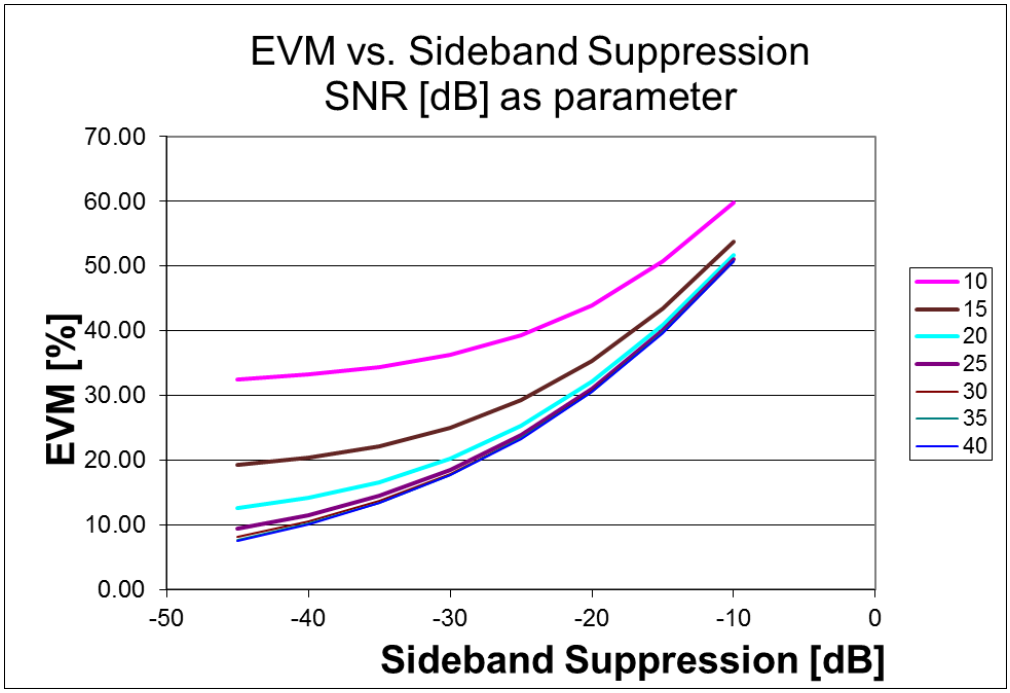

Figure 2.14 EVM vs. Sideband Suppression; SNR as parameter.

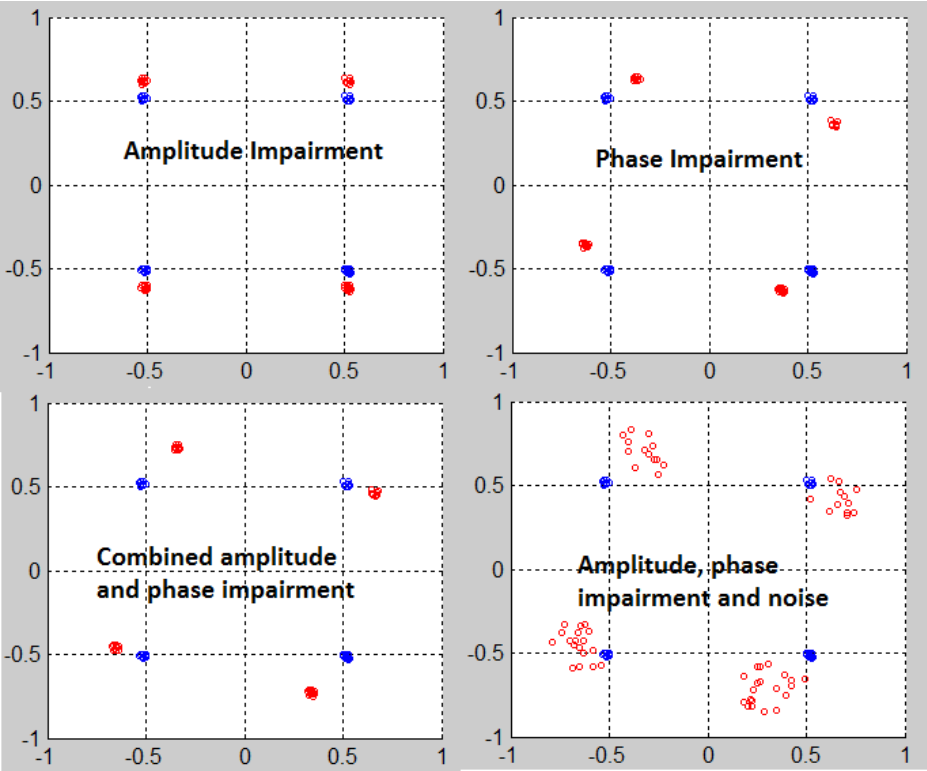

Figure 2.15 Ideal QPSK signal (blue) vs. distorted (red)

The impairment type associated with a specific distortion is shown in the text inside each diagram.

3 Implementation options for the IQ modulator blocks

Referring back to the IQ modulator principle shown in Figure 2.6, we note that there are three functional blocks: the carrier phase shifter, the time domain multipliers and the output combiner. In the next section we'll analyze the implementation options available while emphasizing advantages and disadvantages of each solution presented. While phase shifting is required both in RF and BB, the main focus will be on the RF carrier. Phase shifting in the baseband (BB) is usually performed by a DSP module, implemented either with a dedicated processor or an FPGA. Baseband signal generation is outside the scope of this work.

3.1.1 Phase shifting

3.1.1.1 Passive Low-Pass/High-Pass RC network.

One of the simplest solutions that can be applied for phase shifting is to simultaneously pass the carrier signal through a set of single pole low pass and high pass filters as shown in Figure 3.1.

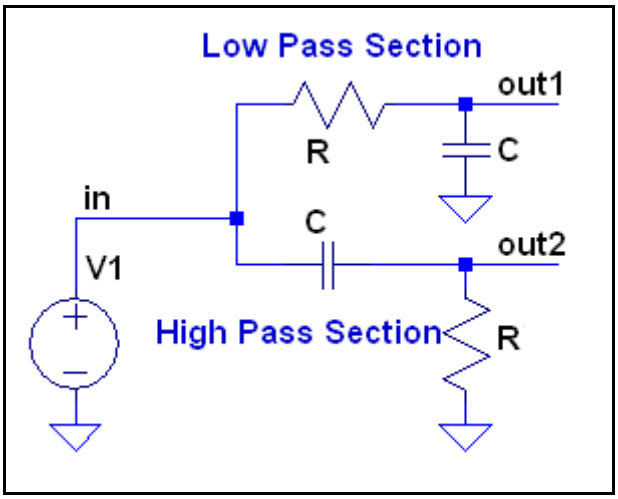

Figure 3.1 Passive Low-Pass/High-Pass phase shifter schematic

Assuming that V1 is a low impedance voltage source, the complex transfer function for the Low Pass section (out1) can be written as:

$$\frac{V(out1)}{V(in)} = \frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} = \frac{1}{1 + j\omega \cdot R \cdot C} \tag{19}$$

The magnitude and phase of the transfer function can be determined from

$$\frac{V(out1)}{V(in)} = \frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} = \frac{1}{1 + j\omega \cdot R \cdot C} \tag{20}$$

$$\left|\frac{V(out1)}{V(in)}\right| = \frac{1}{\sqrt{1 + (\omega \cdot R \cdot C)^2}} \text{ and } \varphi 1 = -a\tan(\omega \cdot R \cdot C) \tag{21}$$

The High Pass filter transfer function corresponding to out2 can be written as:

$$\frac{V(out2)}{V(in)} = \frac{R}{R + \frac{1}{j\omega C}} = \frac{j\omega \cdot R \cdot C}{1 + j\omega \cdot R \cdot C} \tag{22}$$

and the magnitude and phase can be expressed as:

$$\left|\frac{V(out2)}{V(in)}\right| = \frac{\omega \cdot R \cdot C}{\sqrt{1 + (\omega \cdot R \cdot C)^2}} \text{ and } \varphi 2 = a\tan\left(\frac{1}{\omega \cdot R \cdot C}\right) \tag{23}$$

We can evaluate the phase shift between the two outputs by taking the tangent of the difference of each individual phase shift:

$$\tan(\varphi 2 - \varphi 1) = \frac{\tan\varphi 2 - \tan\varphi 1}{1 + \tan\varphi 2 \cdot \tan\varphi 1} = \frac{\frac{1}{\omega \cdot R \cdot C} - (-\omega \cdot R \cdot C)}{1 + \frac{1}{\omega \cdot R \cdot C} \cdot (-\omega \cdot R \cdot C)} = \infty \tag{24}$$

Therefore $\varphi 2 - \varphi 1 = 90\,\text{deg}$, independent of the frequency applied at the input.

Looking at the amplitude behavior, we notice that signal magnitudes at out1 and out2 are equal when $\omega RC = 1$. However, due to the different filter configurations on each branch, Low Pass vs. High Pass, the signal magnitudes at out1 and out2 have different evolutions and a small frequency or component variation results in a significant level difference. We saw in the previous sections the importance of the phase and amplitude balance.

If we note $\frac{1}{RC} = \omega_0$ and we are interested to find out what is the frequency variation that would result in 1dB level variation, we write:

$$\left|\frac{V(out1)}{V(in)}\right| = \frac{1}{\sqrt{1+\left(\frac{\omega}{\omega_0}\right)^2}} = 10^{\frac{-3.5}{20}} \quad \Rightarrow \omega = 1.11 \cdot \omega_0 \tag{25}$$

Equation (25) shows that 11% frequency variation would result in 1dB amplitude impairment. 1dB variation represents the cumulated contributions of 0.5dB from the LP and the HP filter sections.

Advantages:

- The LP/HP passive filter phase shifter is simple to design and to implement
- The phase shift is 90 deg independent of the frequency applied at its input

Disadvantages:

- Due to the significant amplitude variation, this phase shifter may be appropriate for narrow bandwidth of 20% or less.
- On chip component tolerances may result in performance degradation.

3.1.1.2 Passive RC series circuit with transformer

Another relatively simple solution for phase shifting consists in a single series RC circuit supplied from the floating secondary of an RF transformer. Electrically, the transformer secondary winding can be assimilated with a voltage source. The voltages at nodes out1 and out2 will now be calculated.

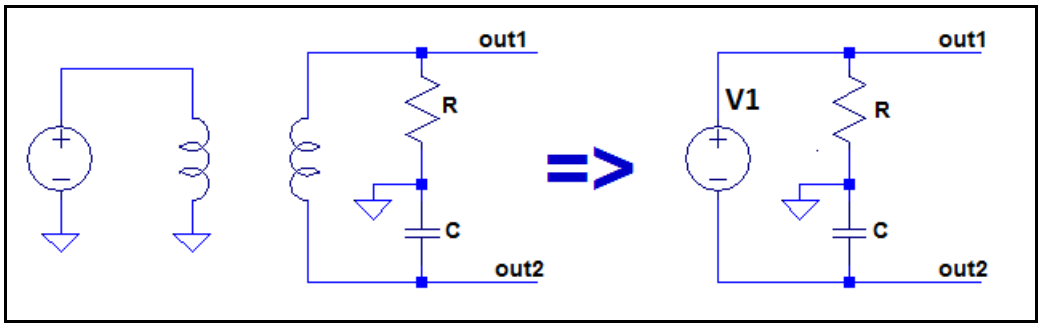

Figure 3.2 Phase shifting using a series RC circuit and a transformer

We start by determining the current through the RC series circuit.

$$I = \frac{V1}{R + \frac{1}{j\omega C}} = \frac{j\omega C}{1 + j\omega \cdot R \cdot C} \tag{26}$$

Therefore, the transfer function between out1, out2 and the input voltage, V1, becomes:

$$\frac{Vout1}{V1} = \frac{R}{R + \frac{1}{j\omega C}} = \frac{j\omega RC}{1 + j\omega \cdot R \cdot C} \quad and$$
$$\frac{Vout2}{V1} = \frac{\frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} = \frac{1}{1 + j\omega \cdot R \cdot C} \tag{27}$$

Since we note that the two equations are the same as the ones derived in the previous section, we can conclude that the performance should be similar to the Low Pass/High Pass circuit. The extra component added in this design may be either an advantage, in case DC blocking is required, but it can further limit the operational bandwidth. Also, transformers are known to take a significant amount of on chip real estate, and contribute to some loss.

3.1.1.3 Polyphase RC network

The polyphase filter is a passive filtering circuit that has a frequency transfer function that is not symmetrical about the zero frequency. This characteristic of the filter has been used in many communication circuits to implement positive frequency conversion for

either transmit or receive, in an attempt to solve the problem of the image signal. Due to the symmetric layout of the circuit, the polyphase filter is able to produce equally balanced quadrature signals at its output. Figure 3.3 shows the schematic diagram of a first order RC polyphase phase shifter.

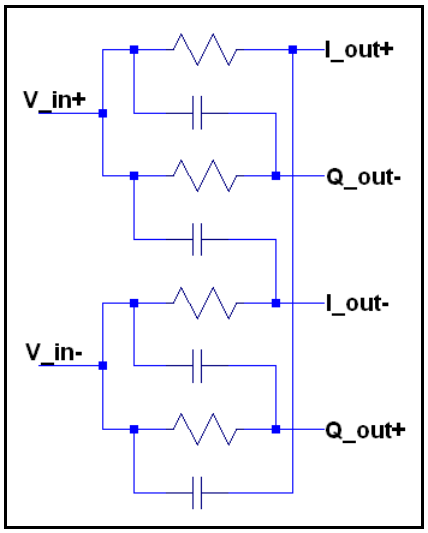

Figure 3.3 First order polyphase passive phase shifter

The operating equations and their derivation of the polyphase circuit do not constitute the scope of the current work. However, it is worth noting that the performance is significantly improved over the simple Low Pass/High Pass configuration. The 90 degrees phase shift is maintained over a wider frequency range. If better precision needs to be maintained, multiple stages can be cascaded. As a rule of thumb, cited from references [6] and [7], for a frequency range of one decade, each stage is expected to provide ~10dB of unwanted sideband attenuation. Increasing the frequency range by one octave requires an additional segment.

Among the disadvantages we can mention:

- The circuit is passive and it attenuates the signal
- A buffer is required at its output in order to drive a mixer cell
- The circuit is sensitive to component variation

- As the required frequency span and accuracy increases, more stages, therefore more area is required on chip

3.1.1.4 Wideband Digital Phase Shifter

Since more advanced technologies allowed the transition frequency of MOS transistors to be extended well above 100GHz, it is natural to attempt a digital state machine design for the phase shifter circuit. The conceptual schematic of such a state machine, including the main waveforms are shown in Figure 3.4.

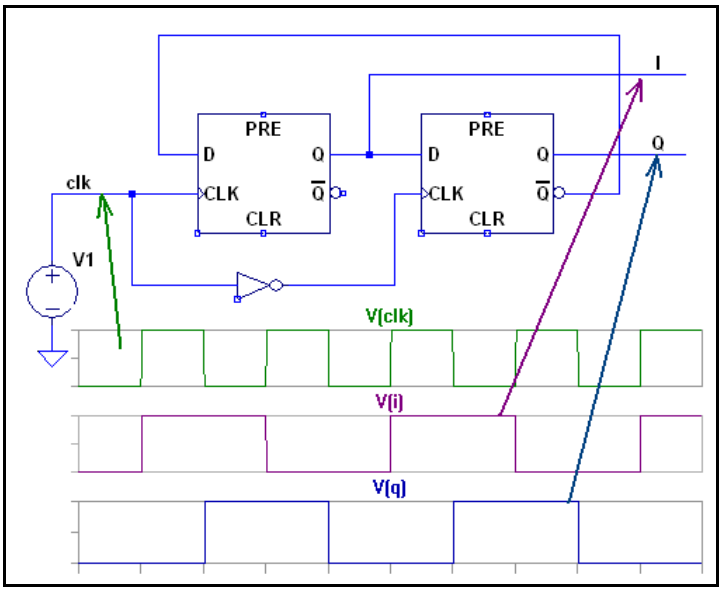

Figure 3.4 Conceptual diagram of a wideband digital phase shifter

The operation mode of the digital phase shifter is described in the next section. Let's assume that the initial state of the circuit is I=0 (In Phase) and Q=0 (Quadrature). At the first rising edge of the clock, the I output rises to logic 1 due to the inverted Q output connected at its D input. The second FF is fed an inverted replica of the clock, therefore at the falling edge of the clock, the Q output will become 1 as well. However, at the third clock edge, which is a rising edge again, the first FF output will become 0 and at the fourth clock edge, which is a falling edge, the state will be back at I=0 and Q=0. This completes a full cycle of the circuit states which repeats itself every two input clock cycles.

With reference to Figure 3.4, we note that the I and Q outputs behave as a clock signal having one half of the input clock frequency. Because the first FF is positive edge triggered and the second FF is negative edge triggered, the state changes every half of the input clock cycle. The result is that the Q output is delayed one quarter cycle behind the I output. The final result is the 90 degrees phase shift between I and Q.

The circuit operation is entirely static. The important benefit of the configuration is that it can operate on a wide range of frequencies. The upper limit is mainly determined by the propagation delay through the gates, while the lowest operating frequency is DC.

The main disadvantage of the circuit, in addition to the need for a high-frequency input clock, is that it may oscillate when driven at low frequencies, if the clock signal slew rate is not high enough to completely and rapidly switch on or off the internal stages. The implication is that a sine wave can be used to drive the clock input at high frequency, however at low frequency the signal needs to be rectangular. More details will be provided in the actual circuit implementation section.

3.1.2 Digital phase shifter implemented in CML technology

While there are many technologies that can be used for the digital divider based phase shifter, the CML option draws attention due to its high speed of operation.

3.1.2.1 The CML inverter gate - the basic building block of the CML logic

The configuration and the operation of the CML inverter gate is similar to the ECL technology. The goal is to achieve the maximum device speed by always operating them either outside of saturation, in the case of bipolar based ECL, or in saturation in the case of MOS based differential pair. Speed optimization is achieved by using the minimum gate length allowed by the technology for all the switching devices, and operation at the optimal current density for highest speed. The following variables are generally considered to be the main knobs that the designer can adjust in order to optimize and find the best trade-offs for the circuit: the drain/load resistance, RL, the MOSFET gate width, W, and the tail current value, IT.

While the focus is on CML gate configuration, for illustration purposes only, a static inverter schematic is shown in Figure 3.5 below. The static gate operation assumes that either one or the other transistors is fully switched on or off. As a result, the propagation delay is increased which results in a reduced maximum operation frequency.

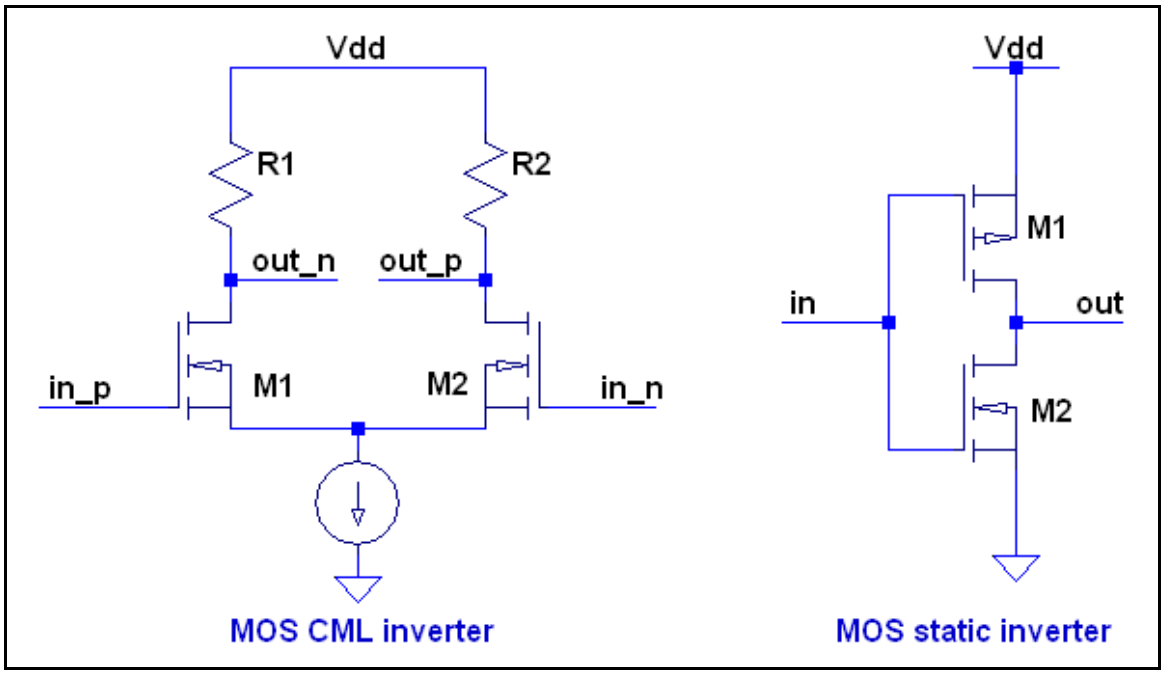

Figure 3.5 MOS CML vs. static inverter schematics

With respect to the CML gate, the logic voltage swing at the inverter output is largely determined by the load resistance and the tail current, which gets switched trough one or the other of the differential pair transistors. We are interested to find out what is the optimum current density that pushes the devices to the maximum operating frequency allowed by a particular technology.

According to reference [2], it appears that keeping the voltage swing to a minimum should improve the delay. The minimum swing should be determined by the minimum voltage required to fully switch the tail current from one transistor to the other within the differential pair. In MOS CML designs, the required voltage swing is strongly dependent on the bias current density of the MOSFETs in the differential pair, but does not depend

on the threshold voltage. The minimum voltage swing required to fully switch the tail current can be expressed as:

$$\Delta V_{\min} = 2 \cdot \left(V_{GS}\left(\frac{I_T}{W}\right) - V_{GS}\left(\frac{I_T}{2W}\right)\right) \tag{28}$$

Experimental measurement data has shown that there is little speed improvement beyond current densities of 0.3-0.4mA/μm which also corresponds to the peak f_T of a NMOS device.

In practice, a gain of 1.5 is considered a typical design starting point that ensures regeneration in the CML latch.

3.1.2.2 The CML latch

The implementation of a MOS CML latch is shown in Figure 3.6. A short description of the operating mode is as follows.

- When ck signal is high and ck_b is low, M5 from the bottom differential pair is in conduction and the current provided by the tail current generator is switched towards the sensing differential pair M1 and M2. The data applied at the D and D_b terminals is transferred (or sampled) to the output.

- The opposite happens when ck_b signal is high and ck is low. M6 from the bottom differential pair is in conduction and the current provided by the tail current generator is switched towards the differential pair M3 and M4 that is part of the actual latching cell. The data that was previously applied at the D and D_b terminals is now latched and continuously applied to the output, irrespective of the state or transitions of the data input terminals.

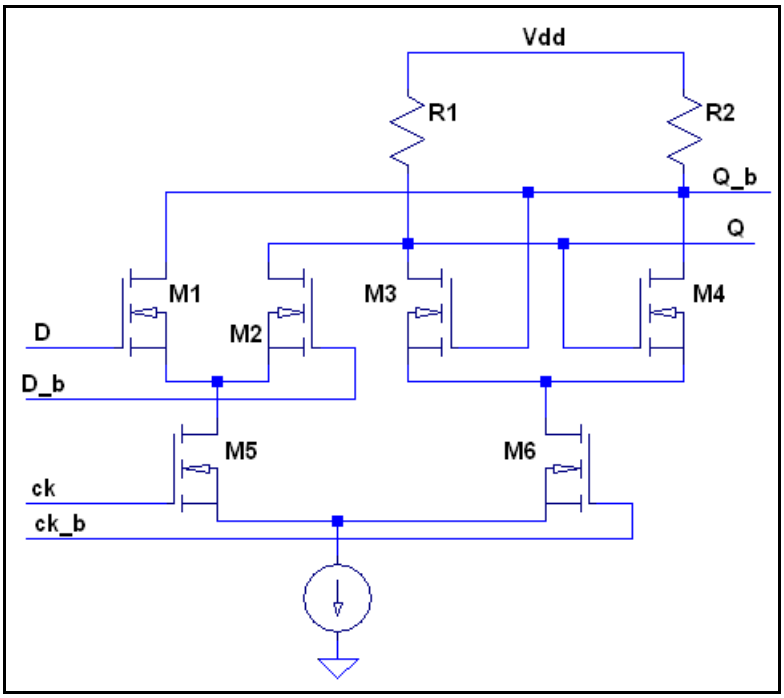

Figure 3.6 MOS CML D latch configuration

Connecting two D latches together in a feedback loop configuration as shown in Figure 3.4 results in a frequency divider that can be used as a wideband phase shifter device.

3.1.2.3 CML divider and phase shifter

The complete implementation of the CML divider – phase shifter is shown in Figure 3.7 below. The diagram follows the conceptual logic diagram shown in Figure 3.4 with the exception that in the CML case, the operation of the circuit is fully differential.

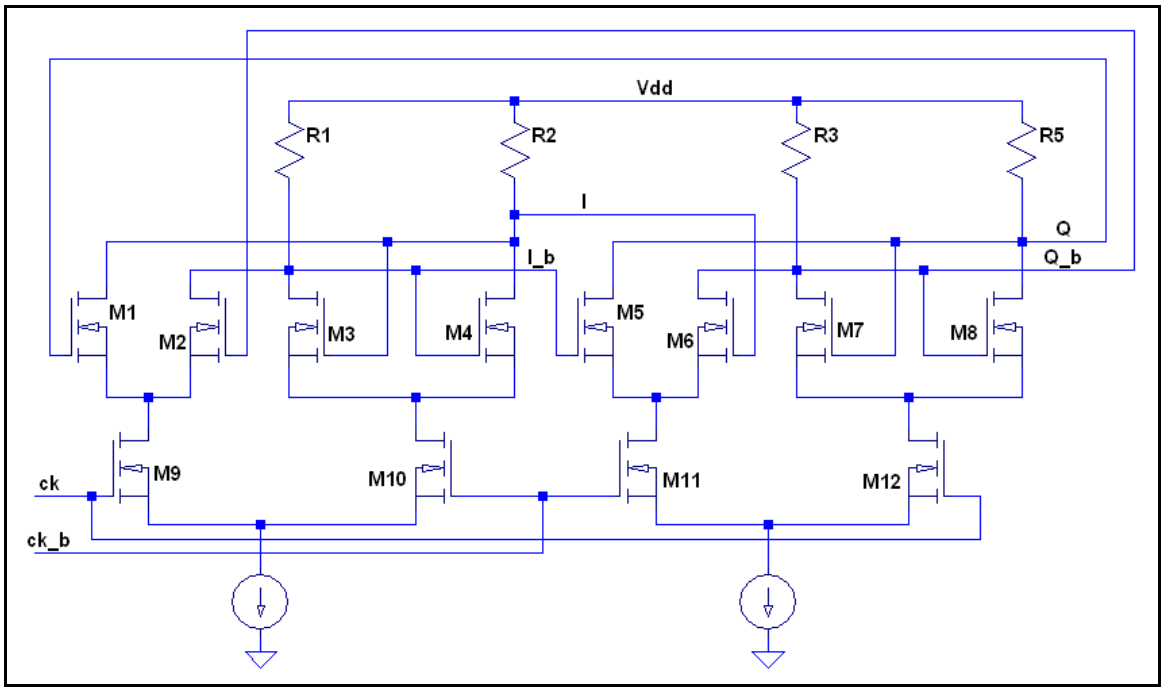

Figure 3.7 Schematic of the CML divider - phase shifter

Let's take now a closer look at the operation of a static divider with CML latches. With reference to Figure 3.7, the first latch consists of MOS transistors M1, M2, M3, M4, M9 and M10. The outputs of the first latch are labeled I and I_b. The second latch is driven by the first and consists of MOS transistors M5, M6, M7, M8, M11 and M12. The outputs Q and Q_b of the second latch are connected back to the data input of the first latch thus forming a feedback loop.

It is worth noting at this point that, in order for the circuit to operate as designed, the two latches need to be triggered on the opposite edges of the clock. This is achieved by cross-connecting the tail transistor gates of the sense and the regeneration sections of the two latches.

At low frequencies, the latches store the sampled data and their state remains unchanged until the next clock edge arrives. As shown before, the loop gain of the positive feedback of each storage cell (pairs M3-M4 and M7-M8) must exceed unity. Under such condition, the output remains constant for a significant period of time and it looks like a square

wave. As the clock frequency increases, the idle time decreases, and the divider would work properly as long as the input pairs M9-M10 and M11-M12 switch the current completely. Further increase of the clock frequency, brings the divider close to a self-resonant state. At this point the divider operates as a two-stage ring oscillator. The regenerative pairs provide sufficient hysteresis such that each latch contributes 90˚ of phase shift, and the required input power becomes very low. It is the point where the clock input sensitivity reaches a maximum, meaning a minimum in terms of the power required to drive the divider. Beyond this frequency, the divider acts again as a driven circuit. As the frequency is further increased, it reaches the bandwidth of the circuit. This is the point where the data input to Q output delay of the latches approaches half of the input clock cycle. As shown in Figure 3.8, Figure 3.9 and Figure 3.10 at multiple frequencies, in such circumstances the timing sequence becomes out of order, and the division process fails irrespective to how large the input power is. The pictures are the result of simulations performed in the design process of the circuit that is the subject of this thesis.

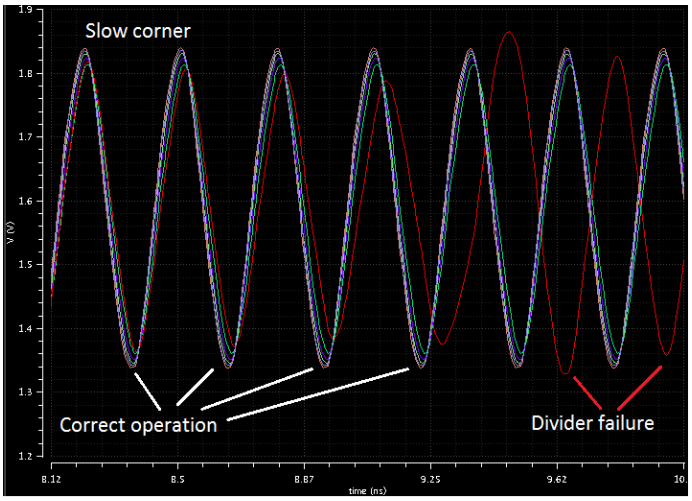

Figure 3.8 Slow corner, Fin=7GHz divider failure while sweeping the input clock power

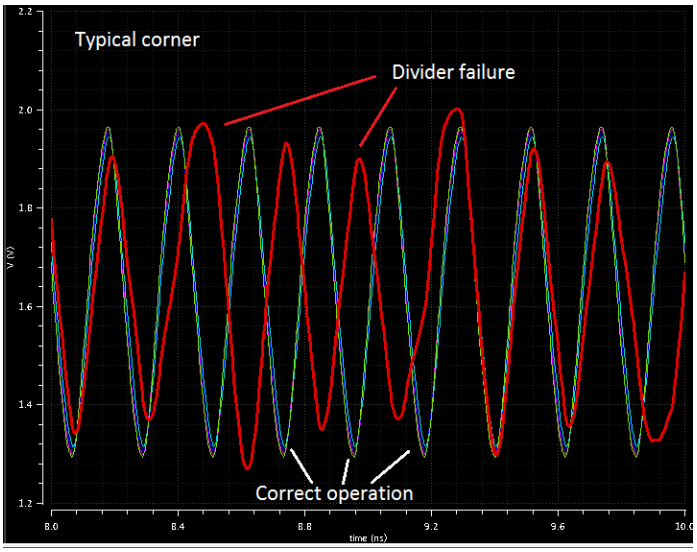

Figure 3.9 Typical corner, Fin=9GHz divider failure while sweeping the input clock power

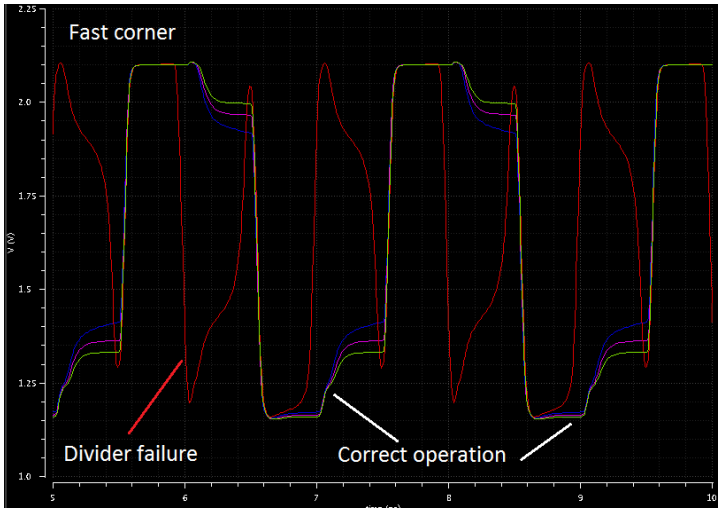

Figure 3.10 Fast corner, Fin=1GHz divider failure while sweeping the input clock power

Figure 3.11 below shows the simulated input sensitivity (i.e., minimum required RF power) as a function of input frequency for the phase shifter section of the up-converting mixer implemented in the 130nm CMOS IBM technology as part of this thesis work. As explained above, the plot shows that the resonant input frequency, where the sensitivity reaches a maximum (minimum input power) is located around 6GHz.

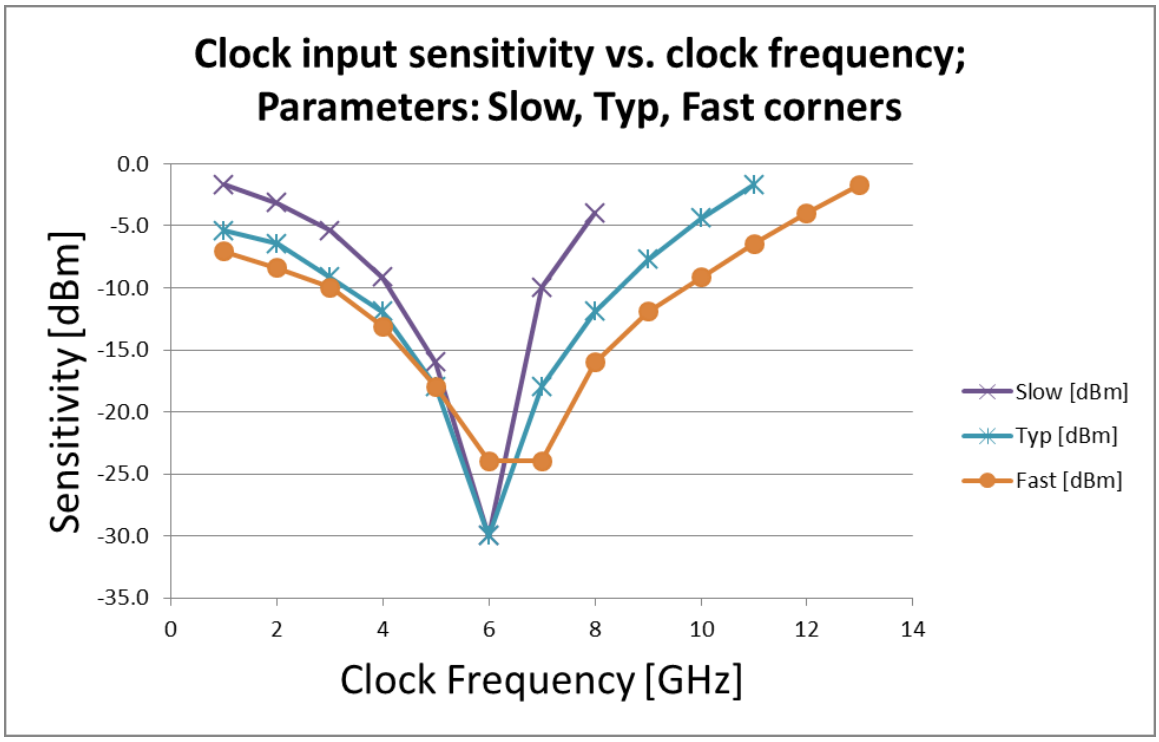

Figure 3.11 Simulated clock input sensitivity vs. frequency

A special particularity of the CML based static divider is that the circuit may not function properly at very low speed if the input is sinusoidal; that is if low slew rate clock signal is being applied to the clock input. If a slow transition signal is applied to the clock input, both latches may be simultaneously turned on, which would allow a complete feedback loop to be established inside the circuit. With reference to Figure 3.7 this would happen when transistors M9, M10 and M11, M12 would each conduct approximately half of their corresponding tail currents. As a consequence, the latches would act as amplifiers with a total loop gain higher than unity, therefore allowing the oscillation to occur at certain frequency where the loop phase shift is a multiple of 2π. Another way of interpreting the phenomenon is to look at the process as ”racing”, a condition well known from the digital circuit design. Fortunately, at low clock frequencies, it is quite easy to generate square-wave signals with sharp transitions (high slew rate), that allow fast switching of the bottom differential pairs of the drivers. In this process, where the pairs are completely

switched on or off, the latches are isolated from each other and an analog loop can't be created.

The maximum frequency at which such a divider can operate can be extended by using inductive peaking, a known technique used in the industry. However, there is a price to pay using inductive peaking. In addition to the increased layout area, the stability of the circuit may be degraded. In this particular design, inductive peaking was not used in an attempt to minimize the layout area as much as possible, while preserving the circuit stability.

In a real life application, however, the sensitivity may not be an issue if the oscillator driving the phase shifter is on chip and it has sufficient signal magnitude to drive the clock inputs of the shifter.

3.1.3 Multiplier cell

Multiplication is the fundamental time domain operation that is required for frequency translation. It is essentially a nonlinear process that takes two or more signals as an input and generates, among other components, the time domain product of the input signal. While nonlinearity is the main characteristic of the mixing device, it will be shown later that linearity is still important. Quite paradoxical! More details in the next chapters.

3.1.3.1 Single diode as a multiplier.

The simplest nonlinear device that can perform the multiplication of two waveforms is the diode. Figure 3.12 below shows the generic I(V) curve of a general purpose diode.

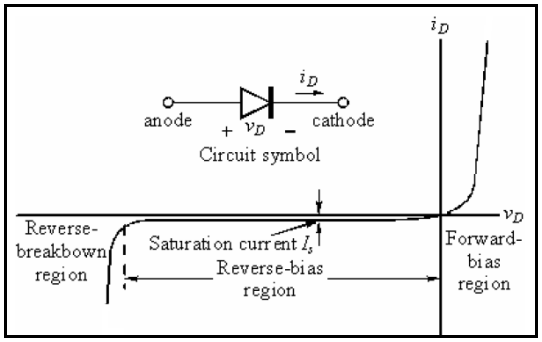

Figure 3.12 I/V curve of a general purpose diode

Mathematically, knowing the voltage applied at the diode terminals, the current that passes through the diode can be expressed as:

$$i_D = I_s \cdot \left(e^{\left(\frac{V_D}{n \cdot V_T}\right)} - 1 \right) \tag{29}$$

Where the following notations were used:

- i_D - the diode current
- I_s - the saturation current of the device
- V_D - the voltage applied at the diode terminals
- V_T - the threshold voltage of the diode semiconductor material; typically V_T=25mV for silicon
- n – an arbitrary constant between 1 … 2 that depends on the fabrication process.

In order to analyze the nonlinear behavior of the diode, we need to take a look at the Taylor series expansion of the exponential characteristic of the diode and simultaneously apply two sine waves at its terminals.

The Taylor expansion of the exponential function is:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots \tag{30}$$

Assuming we apply the sum of two sine waves at the diode terminals and we only look at the first three terms of the Taylor series expansion, we get:

$$\begin{aligned} V_D &= \cos(\omega_1 t) + \cos(\omega_2 t) \\ e^{V_D} &= e^{(\cos(\omega_1 t) + \cos(\omega_2 t))} \\ e^{V_D} &\approx 1 + \left(\cos(\omega_1 t) + \cos(\omega_2 t)\right) + 0.5 \cdot \left(\cos(\omega_1 t) + \cos(\omega_2 t)\right)^2 \end{aligned} \tag{31}$$

After a few manipulations, the e^x expression becomes:

$$e^{V_D} \approx 1 + \left(\cos(\omega_1 t) + \cos(\omega_2 t)\right) + 0.5 \cdot \cos^2(\omega_1 t) + \cos^2(\omega_2 t) + \\ + \cos(\omega_1 t) \cdot \cos(\omega_2 t) \tag{32}$$

The last term of (32) contains the product of the two tones that were initially applied to the diode as a sum of components.

The derivation above can be thought more like a proof of concept; however the applicability is limited due to a number of factors:

- The multiplication must be done in the first quadrant only, if we want to take advantage of the exponential characteristic.
- While this may be useful in ultra-high frequency multiplication applications, it is impractical for digital modulation where the baseband signal voltages pass through all four quadrants.
- The conversion gain is usually negative
- There are many undesired products being generated which may fall in band. They are usually difficult to filter and they may lead to increased signal distortion

3.1.3.2 Diode ring multiplier.

Another option for frequency multiplication that is popular in the industry is the diode ring multiplier or mixer. A typical schematic of the diode ring mixer is shown in Figure 3.13 below.

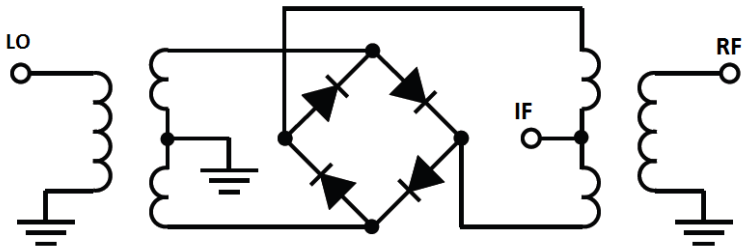

Figure 3.13 Diode Ring multiplier

The operation of the diode ring mixer relies on the basic assumption that the voltage applied at the LO port is large compared to the diode threshold, and that the signal applied

at either the RF or the IF port is small. Under this condition the ring acts a switch. During the positive half-period of the LO sine wave, two of the diodes are reverse biased and the other two diodes are forward biased and saturated. On the negative half-period the roles are swapped. The diodes that were saturated become reverse biased and the previously reverse biased are driven into saturation.

For the small RF signal, the diodes are open circuit when reverse biased, and present small resistances when forward biased. As a result, the IF signal switches between $+V_{RF}(t)$ and $-V_{RF}(t)$ depending on the sign of the V_{LO} (t). This operation is equivalent to multiplying $V_{RF}(t)$ by a square wave of amplitude ±1 that takes the sign of $V_{LO}(t)$. In most practical applications, it is sufficient to describe the frequency conversion mechanism as the product between $V_{RF}(t)$ and the fundamental frequency term of the Fourier expansion of the square wave. More accurate models can be devised that account for the higher-order Fourier terms, and for the dynamic resistance and capacitance of the diodes.

3.1.3.3 Differential Pair – Single Balanced Mixer

The Differential Pair – Single Balanced Mixer is an improvement compared to the previously described mixing methods. The following analysis will be made with reference to Figure 3.14 below.

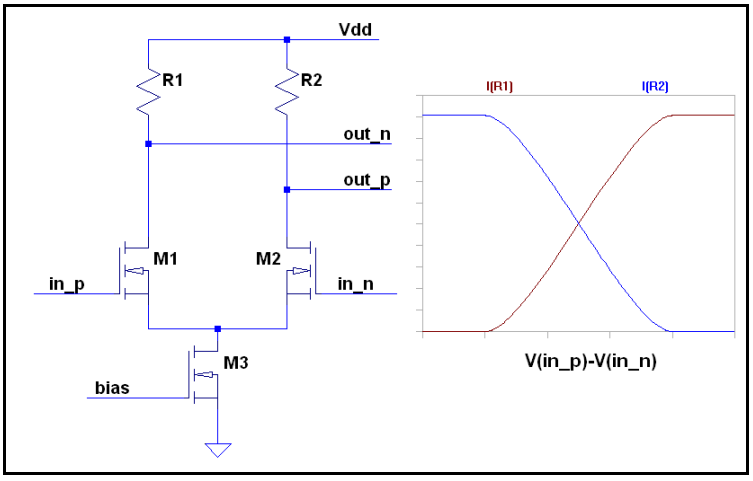

Figure 3.14 Differential Pair and its transfer function

We'll start the analysis by writing the following equations:

$$\begin{aligned} i_D &= \frac{1}{2}\cdot\mu\cdot C_{ox}\cdot\left(\frac{W}{L}\right)\cdot\left(v_{GS}-V_T\right)^2 && \text{device current vs. } V_{gs} \\ I_0 &= I_{M3} = i_{D1}+i_{D2} && \text{bias or tail current} \\ Vin &= V(in\_p)-V(in\_n) && \text{differential input voltage} \end{aligned} \tag{33}$$

The following notations were used:

- i_{D1}, i_{D2} – drain current through transistors M1 and M2
- W – width of the MOS transistor
- L – length of the MOS transistor
- γ – parameter related to the fraction of the gate not pinched-off; typically 2/3 for long channel
- C_{OX} – gate unit capacitance
- μ= – surface electron mobility

Some of the parameters above are dependent on the technology used and the size of the transistors. At this point in the analysis, actual numerical values are not relevant.

Solving for ID1 as a function of Vin we get:

$$i_{D1} = \frac{I_0}{2} + \frac{g_m\cdot v_{in}}{2}\cdot\sqrt{1-\frac{\mu\cdot C_{ox}}{4\cdot I_0}\cdot\left(\frac{W}{L}\right)\cdot {v_2}^2} \tag{34}$$

where we noted:

$$g_m = \sqrt{\mu\cdot C_{ox}\cdot\frac{W}{L}\cdot I_0} = K\cdot\sqrt{I_0} \tag{35}$$

Since I_0 is dependent on the bias voltage, we can recognize the product term $g_m\cdot V_{in}$ in (35) that performs the multiplication between V_{in} and V_{bias}

While the differential pair seen as a single balanced mixer performs the product between the two voltages applied at its inputs, it has some limitations in terms of voltage levels. The bias voltage must be permanently positive in order to maintain proper bias of the tail current generator, M3.

Let's assume that the LO voltage is applied at the gates of M1 and M2 and the baseband signal is superimposed on the bias voltage. If the output is taken differentially between out_p and out_n then the baseband component is rejected. However, it can be shown that beside the desired output frequency, other resulting components from mixing in the single balanced mixer are harmonics of the LO frequency. This is known as LO feed-through and it is an undesirable output component [10].

3.1.3.4 Double Balanced Mixer – Gilbert Cell

The most commonly used mixer configuration in RF IC design is the Double balanced mixer, also known as the Gilbert cell. The Gilbert cell is a combination of two single balanced mixers. Due to its double balanced structure, a double balanced mixer rejects both RF and LO feed-through. The Gilbert cell is the choice for the design and its behavior will be thoroughly analyzed in the next section.

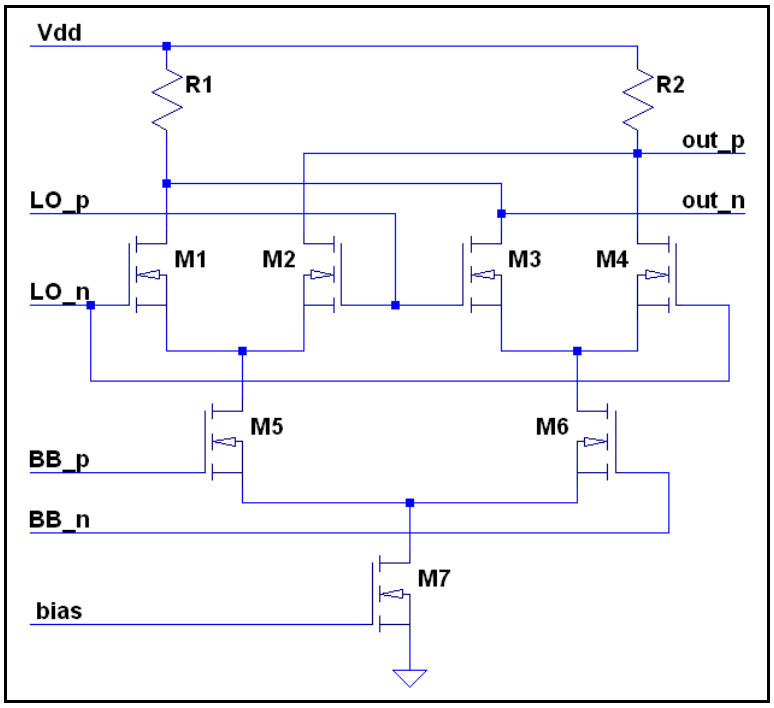

Figure 3.15 Gilbert Cell schematic diagram implemented with MOS transistors

From a functional standpoint, a Gilbert mixer consists of a voltage-current converter (M5, M6), two current switching blocks (M1, M2 and M3, M4), and a current-voltage converter (R1, R2). Figure 3.15 shows an implementation of the double balanced mixer. The two switching blocks operate in opposite polarity since the LO_p and LO_n signals are applied 180˚ out of phase. The V-I converter (M5, M6) generates two current inputs to the tail nodes of the switching blocks, which are also 180˚ out of phase.

The operation of the Gilbert cell can be explained as follows: transistors.M5 and M6 form a differential pair that performs a voltage to current conversion for the baseband input signal, (BB_p, BB_n). The drain currents generated from M5, M6 are then switched through the top two differential pairs formed by transistors M1, M2 and M3, M4. To simplify the analysis, let's assume that the LO_p and LO_n are square wave voltages large enough to completely switch on or off one of the top differential pairs. As shown it is assumed to switch transistor M2, M3 completely on when it is high and off when it is low. We further assume that the baseband currents passing through M5, M6 remain ap-

proximately constant during the LO period. While the LO=LO_p-LO_n voltage is switching, there are two states of the circuit.

First LO half cycle: V(LO)=LO_p-LO_n > 0

If we apply LO such that LO>0 and LO_p > LO_n, this means that M2 and M3 are completely on, with their drain current equal to each tail current, which in turn equals I(M5) and I(M6), respectively. M1 and M4 are completely off and their drain currents are zero. The positive output current of the mixer, V(out_p), through the cross-coupled connection, is the sum of the drain currents injected by M2 and M4 and hence I(R2)=I(M2)+I(M4)= I(M2)+0 (i.e, it equals the current from the tail node of M2). The negative output current of the mixer, V(out_n), again through the cross-coupled connection, is the sum of the drain currents injected by M1 and M3 and hence I(R1)=I(M1)+I(M3)= 0+I(M3) (i.e, it equals the current from the tail node of M3).

Second LO half cycle: V(LO)=LO_p-LO_n < 0

In the second LO half cycle, V(LO) switches and exactly the opposite occurs. Since V(LO) is now negative and LO_p < LO_n, this means that M2, M3 are now completely turned off. On the other hand, M1 and M4 are now completely turned on. Since the pair of "on" transistors has switched, the current that got routed to the output has also been swapped around. Repeating the analysis shown for the first half of the LO cycle, the positive output current of the mixer, V(out_p), is still the sum of drain currents from M2 and M4. However, since the current is now from the tail node of M4 as opposed to being from the tail node of M2, then I(R2)=I(M2)+I(M4)= 0+I(M4) (i.e, it equals the current from the tail node of M4).

3.1.4 Mixer Specification Parameters

A couple of mixer parameters are widely used in the industry and allow the characterization of mixers performance. In the next section we will look at: gain, linearity, noise, isolation and a figure of merit proposed in reference [2]. Since the final design of this work uses the Gilbert Cell, all the analysis will focus on this type of mixer.

3.1.4.1 Mixer Gain

The conversion gain of a mixer represents the small signal transfer function from the baseband or IF input to the RF output in an up-converting mixing device, or from the RF input to the IF or baseband output in a down-converting topology. The gain can be defined both in terms of power or voltage. Defining the gain in terms of power makes it easier to accurately measure the circuit stand-alone at the RF and LO ports using standard 50Ω test equipment. It is not uncommon for some publications to describe CMOS mixers in terms of voltage conversion gain. One reason for this practice is that voltage is displayed by low-frequency test equipment with high input impedance. Another reason is that the mixers may drive a circuit that has high impedance, instead of 50Ω.

The conversion gain of a mixer is always smaller than that of the corresponding amplifier using the same transistor topology, usually a cascade amplifier. This is primarily due to the fact that one of the signals produced by the mixer is usually discarded, resulting in a 3dB power loss. The net result is a factor of two in power gain reduction. Secondly, the mixer transconductance is modulated by the large LO signal, resulting in an average transconductance value in the mixer that is smaller than the peak transconductance of an amplifier using the same configuration.

Since the upper transistors are switching, the output of the mixer can be seen as a sine wave multiplied by a square wave having +1 and -1 as high and low levels. The sine wave originates from the baseband signal amplified by the M5, M6 transistor pair. In general, we can write the mixer gain as:

$$\frac{Vo}{Vin} = V(LO) \cdot gm \cdot Rd \tag{36}$$

where we noted:

Vo – the differential output voltage, V(out_p)-V(out_n) in our case

Vin - the differential input voltage, V(bb_p)-V(bb_n)

V(LO) – the LO voltage that switches on and off the upper transistor quad

gm – the transconductance of the lower differential pair operated as a linear amplifier

Rd – the drain load resistance, R1 and R2 in out case

Let's take a closer look at the LO voltage. As explained above, it is a square wave alternating from +1 to -1. The Fourier analysis of the square wave shows that:

$$V(LO) = \frac{4}{\pi} \cdot \sin(\omega_{LO} \cdot t) + \frac{4}{3\pi} \cdot \sin(\omega_{LO} \cdot t) + \frac{4}{5\pi} \cdot \sin(\omega_{LO} \cdot t) + \ldots \tag{37}$$

Multiplication in the time domain is equivalent with convolution in the frequency domain, hence, the baseband signal will be translated around each harmonic component of the LO. Since we are interested in the fundamental component located at the LO frequency, we can ignore the harmonics and assuming a sinusoidal baseband input, the output voltage becomes:

$$\begin{aligned} Vo = V(LO) \cdot Vin \cdot gm \cdot Rd = \frac{4}{\pi} \cdot \sin(\omega_{LO} \cdot t) \cdot Vbb \cdot \sin(\omega_{bb} \cdot t) \cdot gm \cdot Rd = \\ = \frac{4}{\pi} \cdot \frac{1}{2} \cdot Vbb \cdot \left(\sin(\omega_{LO} + \omega_{bb}) \cdot t) + \sin(\omega_{LO} - \omega_{bb}) \cdot t)\right) \cdot gm \cdot Rd \end{aligned} \tag{38}$$

Only one of the side bands is usually selected for further processing, the mixer gain, G_{mix}, becomes:

$$G_{mix} = \frac{Vo}{Vbb} = \frac{2}{\pi} \cdot gm \cdot Rd \tag{39}$$

The above equation shows that the Gilbert Cell mixer gain is lower that the equivalent linear cascade amplifier configuration by $20 \cdot \log\left(\frac{2}{\pi}\right) = -3.92dB$.

A couple of notes are worth making with regard to the advantages of switching at the LO frequencies [1]:

- If the LO is operated at low levels, then the amplitude of the output depends on the amplitude of the LO signal. The highest gain is obtained at high LO amplitude. For large LO drive, the upper quad switches and no further gain increases is possible. While tuning the LO over its specified range, the amplitude may change slightly. If the upper quad isn't driven into switching, this may have an impact on

the output power of the transmitter in case of the up-converting mixer, or sensitivity, in case of the receiver down-converting mixer. If the amplitude of the LO signal is higher than the minimum level required to saturate the quad, then small level variations are irrelevant to the system performance.

- Low sensitivity to LO level variation is especially important for zero IF up-converting mixers or for image-reject mixers. It has been shown in the previous section related to the theoretical basis for IQ modulation that both phase and amplitude matching is important for a good signal quality at the mixer output. By driving the upper quad into switching, amplitude matching on the LO side has less impact on the output signal.

- Noise performance of the Gilbert cell is also minimized with large LO amplitude. With large LO, the upper quad transistors are alternately switched between completely off and fully on. When off, the transistor contributes no noise, and when fully on, the switching transistor behaves as the upper transistor in a cascade linear amplifier. It has been shown in the literature [1], [2], [10] that the upper transistor of a cascade configuration does not significantly contribute to the overall noise figure of the circuit.

- Although port matching has not been explicitly discussed in this section, it is worth mentioning that conversion gain and noise figure optimization can be achieved by complex conjugate terminations implemented at the RF, LO, and IF/BB ports at the RF, LO, and IF/BB frequencies, respectively.

- Resistive mixers implemented with diodes and transistors exhibit conversion loss, whereas active mixers implemented with transistors can exhibit conversion gain.

3.1.4.2 Mixer Linearity

Mixer are nonlinear devices by definition, hence it may seem paradoxical to mention mixer linearity. However, looking closer at the mixer operation we realize that there is a desired and an undesired nonlinearity. The desired nonlinearity stems from the switching nature of the upper transistors quad. A purely linear device would not be able to generate

the mixing products since the spectral components at the output follow the input spectrum. On the other hand, although we would like the bottom differential pair (see M5 and M6 in Figure 3.15) to operate linearly, there are inherent nonlinear effects.

Similar to power amplifiers or low noise amplifiers, there are two main parameters that are used in characterizing the linearity of mixer: the 1dB compression point, P1dB, and the third order intercept point, IP3. The third order intercept point can be referred to the input or to the output of the device, in which case they are noted as IIP3 or OIP3 respectively.

A nonlinear circuit transfer function, in case the circuit has no memory, can be expressed by its Taylor power series expansion.

$$y(t) = k_0 + k_1 \cdot x(t) + k_2 \cdot x^2(t) + k_3 \cdot x^3(t) + \ldots \tag{40}$$

where we noted:

y(t) – the output of the system

x(t) – the stimulus applied to the system input

k_i – coefficients of the power series expansion

In order to perfectly model the nonlinear behavior of the system, an infinite number of expansion terms need to be taken into consideration. However, this is not possible, and for the purpose of the analysis, only the terms up to the third power will be considered as being dominant.

If a single sinusoidal signal, $x(t)=A\cdot\cos(\omega_1 t)$, is applied to the input of the circuit, the output becomes:

$$\begin{aligned} y(t) &= k_0 + k_1 \cdot A \cdot \cos(\omega_1 t) + k_2 \cdot A^2 \cdot \cos^2(\omega_1 t) + k_3 \cdot A^3 \cdot \cos^3(\omega_1 t) + \ldots \\ &= k_0 + \frac{k_2 \cdot A^2}{2} + \left(k_1 \cdot A + \frac{3 \cdot k^3 \cdot A^3}{4} \right) \cdot \cos(\omega_1 t) + \frac{k_2 \cdot A^2}{2} \cdot \cos(2\omega_1 t) + \\ &+ \frac{k_3 \cdot A^3}{4} \cdot \cos(3\omega_1 t) + \ldots \end{aligned} \tag{41}$$

One can note that, although a single frequency has been applied at the circuit input, multiple harmonic components have occurred at the output. They have been generated due to the nonlinear behavior of the circuit. Typically, in this case, the harmonic components can be filtered out, but this scenario, where only one frequency is applied to the circuit, is rarely seen in practice. A more useful analysis can show what happens if two tones are simultaneously applied to the circuit. Let's assume that the stimulus is now $x(t)= A\cdot\cos(\omega_1 t)+ B\cdot\cos(\omega_2 t)$. The output becomes:

$$y(t) = k_0 + k_1 \cdot (A\cdot\cos(\omega_1 t) + B\cdot\cos(\omega_2 t)) + k_2 \cdot (A\cdot\cos(\omega_1 t) + B\cdot\cos(\omega_2 t))^2 + \\ + k_3 \cdot (A\cdot\cos(\omega_1 t) + B\cdot\cos(\omega_2 t))^3 + ... \tag{42}$$

Although the derivation of the amplitudes of the resulting frequency components is quite laborious and it is beyond the scope of this work, a table containing the amplitudes of the first, second and third order components is provided below [1].

Table 1 IMD products magnitudes

Frequency component	Distortion Order	Distortion amplitude
DC	Second	$k0+\frac{k2}{2}\cdot\left(A^2+B^2\right)$
ω1	Desired	$k1\cdot A+k3\cdot A\cdot\left(\frac{3}{4}\cdot A^2+\frac{3}{2}B^2\right)$
ω2	Desired	$k1\cdot B+k3\cdot B\cdot\left(\frac{3}{4}\cdot B^2+\frac{3}{2}A^2\right)$
2·ω1	Second	$\frac{k_2\cdot A^2}{2}$
2·ω2	Second	$\frac{k_2\cdot B^2}{2}$
ω1 ± ω2	Second	$k_2\cdot A\cdot B$
ω2 ± ω1	Second	$k_2\cdot A\cdot B$
3·ω1	Third	$\frac{k_3\cdot A^3}{4}$
3·ω2	Third	$\frac{k_3\cdot B^3}{4}$
2·ω1 ± ω2	Third	$\frac{3\cdot k_3\cdot A^2\cdot B}{4}$
2·ω2 ± ω1	Third	$\frac{3\cdot k_3\cdot A\cdot B^2}{4}$

Two figures of merit are widely used in the industry for linearity characterization. They are the 1dB gain compression point (P1dB) and the 3rd order intercept point (IP3). They will be briefly discussed in the following sections.

- **1dB Gain Compression Point, P1dB**

When electronic circuits are required to process small signals, their gain is usually constant over a wide range of signal magnitudes and the harmonic components generated in the circuit are negligible. However, if the input signal is increased beyond a certain limit, the circuit begins to saturate, the level of harmonic components increase significantly and the gain starts dropping.

By definition, the 1dB output compression point is the output power at which the circuit gain drops by 1dB with regard to the small signal gain. The 1dB compression point can be referred to the input as well. The graphical illustration of the phenomenon is shown in Figure 3.16 below. The plot shows using solid red line Pout(Pin) of a hypothetical device, the ideal linear characteristic using the blue dotted line and the gain variation vs. the input power using the green dotted line. It is noticeable that the gain is fairly constant for input powers of -20dBm or less. The gain starts dropping considerably due to device saturation for input powers above -15dBm.

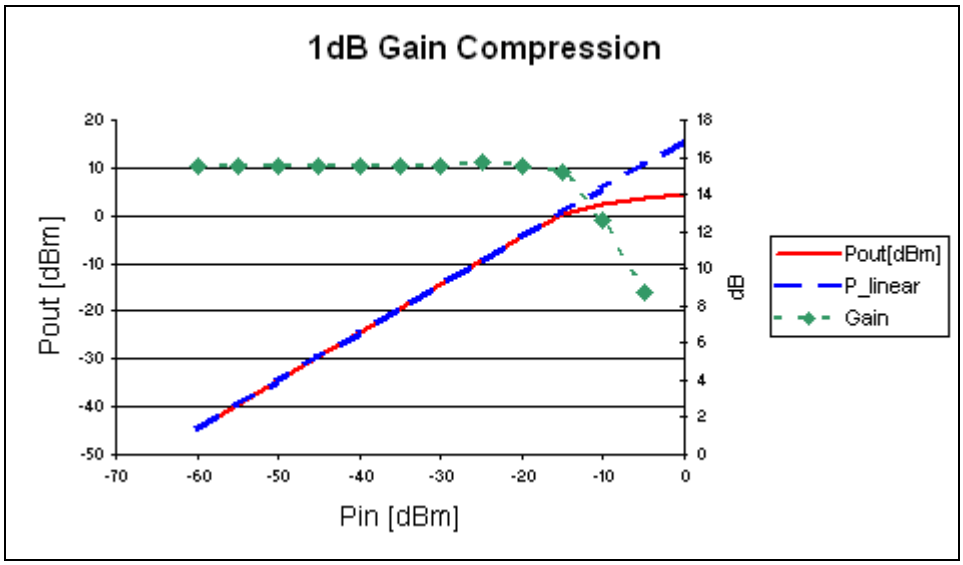

Figure 3.16 1dB compression point example

For modulation schemes where the information is stored in the phase of the modulated signal, this may not be an issue. For instance, in GSM where GMSK modulation is being used, the amplifiers can operate close to their saturation output power, where the efficiency is high, with no penalty on the BER.

On the other hand, when LTE or WCDMA modulation schemes are being used, operating the amplifiers close to the 1dB compression point, often results in EVM degradation as well as spectral re-growth. Several techniques such as power back-off and pre-distortion can be used to in order to find the best tradeoff between signal quality and amplifier efficiency.

- **Intermodulation Distortion and Third Order Intercept Point**

Another way of characterizing the linearity of a system is to look at its intermodulation distortion (IMD) and the third order intercept point (IP3).

Non-linear devices stimulated with multiple tones generate mixing products that arise as linear combinations of the input tones. The mixing products are known as intermodulation (IM) products. Intermodulation products generated in an amplifier, mixer or communication equipment can cause problems, because they represent spurious signals that can

fall in the desired band and alter the quality of the signals. Even-order intermodulation products usually occur at frequencies far from the signals that generate them and they typically are of little concern because they can be eliminated by filtering. The third-order intermodulation products are usually the ones of greatest concern. They occur at 2f1-f2 and 2f2-f1, they are the strongest of all odd order products, and they are close to the signals that generated them. This makes them difficult to eliminate or attenuate by filters.

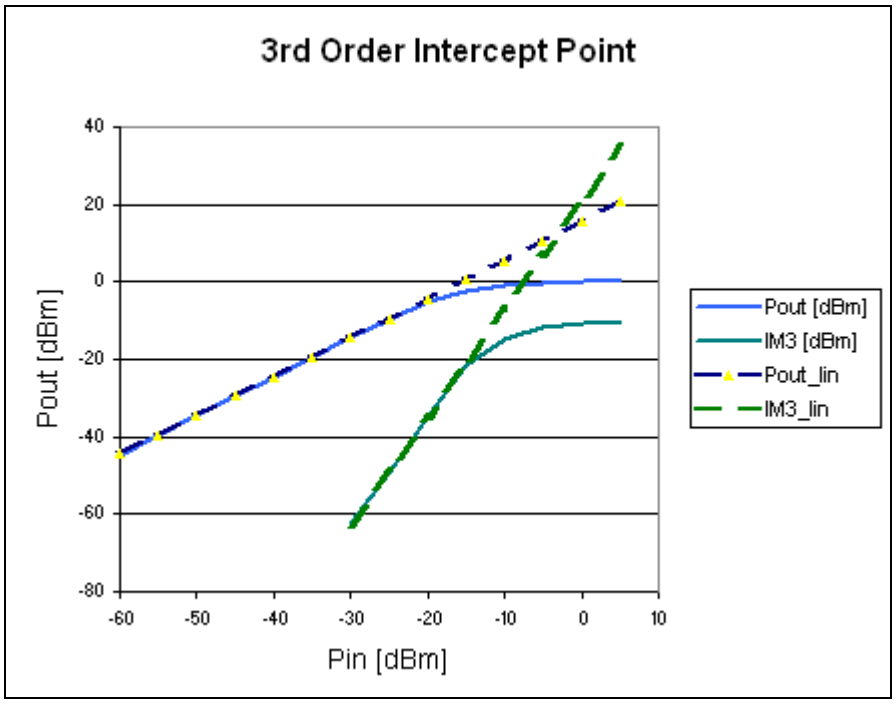

Figure 3.17 3rd order intercept point example

A widely used method of characterizing the linearity performance of a circuit and to determine the intermodulation product levels is to perform a two-tone test. This test requires that two CW carriers of equal amplitude and separated in frequency by a small amount, are simultaneously applied at the circuit input. The powers of the resulting output intermodulation components are measured and plotted on a graph. Such plot is shown in Figure 3.17 above.

One can see that the slope of the third order IM product is three times the slope of the fundamental tones. This is also predicted by the fundamental and third order tone magnitudes shown in Table 1 above, by setting B=A (the fundamental magnitudes) and calculating the ratios between the IMD3 and the fundamental as follows:

$$\text{Fundamental Mag} = k1 \cdot A + k3 \cdot A \cdot \left(\frac{3}{4} \cdot A^2 + \frac{3}{2} B^2 \right) = k1 \cdot A + k3 \cdot \frac{9}{4} \cdot A^3 \tag{43}$$

Extracting the linear component only results in Linear Mag= $k1 \cdot A$

On the other hand, the 3rd order intermodulation product can be written as

$$\text{IM3 Mag} = k3 \cdot \frac{3}{4} \cdot A^3 \tag{44}$$

For small input signal magnitudes, A, the linear term rises linearly, while the IM3 rises three times faster, due to the proportionality to the cube of the input level. As a result, a fictitious point can be defined where the IMD level becomes equal to the fundamental tone level. By definition, this intersection is called the third order intercept point and it can be calculated by setting

$$\frac{IM3\_Mag}{Fundamental\_Mag} = \frac{\frac{3}{4} \cdot k3 \cdot A_{IP3}^3}{k1 \cdot A_{IP3}} = 1 \tag{45}$$

$$\text{And further} \quad A_{IP3} = 2 \cdot \sqrt{\frac{k1}{3 \cdot k3}} \tag{46}$$

Equation (46) provides the input voltage magnitude at the third order intercept point.

The third order intercept point is an important measure of the linearity performance of a mixer or amplifier. In a lab experiment, after measuring the input power (P_i) and the IMD level (P3 in this example), the third order intercept point referenced to the input of the device can be determined as follows:

$$IIP3 = P_i + 0.5 \cdot (P1 - P3) \tag{47}$$

If we take into account that OIP3=IIP3+G and P1=P_i+G, then the output referenced third order intercept point or the OIP3 can be calculated as:

$$OIP3 = P1 + 0.5 \cdot (P1 - P3) \tag{48}$$

Conversely, given the IP3, the amount of IMD generated by the device can be determined at a specific input signal level as follows:

$$P3 = 3 \cdot P1 - 2 \cdot OIP3 \tag{49}$$

At the system level it is important to know the individual component requirements for linearity and noise as well as many other parameters. The formula used for multi-stage system linearity analysis is shown below.

$$\frac{1}{IP3_{sys}} = \frac{1}{IP3_N} + \frac{1}{IP3_{N-1} \cdot G_N} + \frac{1}{IP3_{N-2} \cdot G_{N-1}} + \ldots \tag{50}$$

Where $IP3_N$, $IP3_{N-1}$, $IP3_{N-2}$ … and G_N, G_{N-1} … are the intercept points and the gains of stages N, N-1, N-2 … counted starting from the system output.

More accurate results are usually being obtained using software system simulators.

3.1.4.3 Mixer Noise

The noise contribution is an important parameter of the mixer block and it needs to be taken into account as part of the system analysis.

In general, the noise factor (F) of a signal processing block is defined as the power ratio between the signal to noise ratio at the input and the signal to noise ratio at the output of the block.

$$F = \frac{SNR_{IN}}{SNR_{OUT}} \tag{51}$$

Expressed in dB, the noise factor becomes the noise figure (NF).

$$NF = 10 \cdot \log(F) \tag{52}$$

Mixer noise figure can be specified as either double sideband (DSB) or single sideband (SSB). SSB noise figure is typically used for the mixers that are part of a heterodyne transmitters or receivers. In such cases the input signal is contained in one sideband and the other sideband is removed by an image rejection filter. DSB noise figure is applied to the mixers which are part of direct conversion configurations. Single sideband and double sideband noise figures are related to each other by the following equation:

$$NF_{SSB} \cong NF_{DSB} + 3dB \tag{53}$$

Multiple noise sources and types can be identified in a mixer cell. The most important contributors will be shown below.

- **MOS transistor noise**

MOS transistors are present in a significant number in the Gilbert cell based mixer. Hence it is natural to evaluate the noise sources in the transistor.

There are two dominant noise sources in the MOS transistor: the thermal noise and the flicker noise.

The noise generated inside a MOS transistor can be modeled by means of two current generators, one associated with the gate resistance and the other associated with the drain resistance, and a voltage source as shown in Figure 3.18 below.

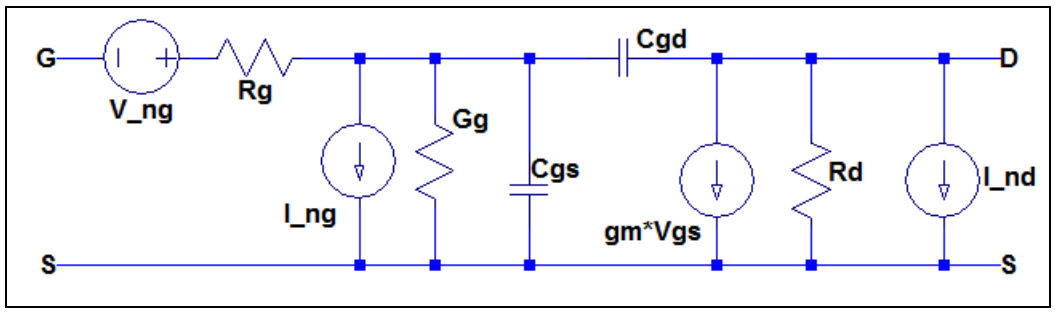

Figure 3.18 MOS transistor equivalent circuit showing the noise sources

The series voltage source is associated with the flicker noise and its power spectral density can be written as [15]:

$$\frac{V_{\_ng}^2}{\Delta f} = \frac{K_f}{W \cdot L \cdot C_{ox} \cdot f} \tag{54}$$

Where we noted:

W - the width of the transistor

L – the length of the transistor

Cox - the gate unit capacitance

F – the frequency

Kf – a constant related to the fabrication process

Δf – the noise bandwidth

The thermal noise is associated with two current sources:
The drain and the gate current noise generators can be written as:

$$i_{\_nd}^2 = 4 \cdot K \cdot T \cdot \gamma \cdot g_m \cdot \Delta f \tag{55}$$

$$i_{\_ng}^2 = 4 \cdot K \cdot T \cdot \delta \cdot g_g \cdot \Delta f \tag{56}$$

where we noted $g_g = \frac{\omega^2 \cdot C_{gs}^2}{5 \cdot g_m}$

Having the same origin, the two current sources are correlated with a correlation coefficient:

$$c = \frac{i_{ng} \cdot i_{nd}}{\sqrt{i_{ng}^2 \cdot i_{nd}^2}} \tag{57}$$

A couple of observations can be made with regard to the gate resistance. In low frequency designs it is often assumed that the MOS devices have extremely high impedance, close to infinity for all practical reasons. However, in RF, the impedances are neither in-

finity, nor zero. The gate resistance can be calculated based on the material parameters provided by the technology. For large size transistors, the gate length can be large, further affecting the f_{max} of the transistor. Improvements can be made by splitting the gate into multiple fingers and by providing gate contacts at both end of the gate poly silicon. Connecting the gate at both sides can provide a further reduction of the gate resistance by a factor of 4, compared to the single contact gate [1]. It is expected that reducing the gate resistance should also improve the noise performance of the MOS device.

It is worth noting that the evaluation shown above still underestimates the noise generated in the MOS device. Noise modeling in the MOS transistors is still an area of active research.

- **Noise Analysis in Gilbert Cell Based Mixers**

As expected from the MOS transistor analysis, the MOS based mixer is affected by the two dominant types of noises: the flicker noise and the white noise. Due to its 1/f frequency dependence, it is expected that the flicker noise has more impact in low IF or even zero IF architectures.

The Gilbert cell mixer consists of an input transconductance stage, a switching stage and the output load. All transistors in the circuit are contributing to the overall noise of the mixer block. References [15] and [16] discuss in detail both the flicker and the white noise issues in high frequency MOS active mixers. A brief description of the two noise types is provided below.

  - Flicker Noise:

    It is also known as 1/f noise due to its power spectral density shape. Using PMOS loads is a technique that allows the minimization of the load noise contribution. This is possible due to the fact that the K_f constant used in (54) is smaller for PMOS devices. An alternative is to use polysilicon resistors as load resistors. The poly silicon material is known to be free from 1/f noise.

For zero or low IF architectures, the transconductance stage transistors only contribute white noise after frequency translation. The flicker noise in these transistors is up-converted to ω_{LO} and to its odd harmonics.

The switching stage in active mixers significantly contributes to the flicker noise. This noise can be minimized by ensuring sharp transitions on the LO signal fed to the mixer. It has also been observed a reduced noise level with lower LO frequencies or with devices having higher f_T [16].

- White Noise:

In Gilbert cell based double balanced mixers the switching stage contributes with white noise to the output when both switches are ON, which should happen only for a small part of every cycle. The noise added by the switching transistors can be written as [16]:

$$\overline{v_n^2} = 8 \cdot K \cdot T \cdot R_L \cdot \left(1 + \gamma \cdot \frac{R_L \cdot I}{\pi \cdot A} + \gamma \cdot \frac{g_m \cdot R_L}{2} \right) \tag{58}$$

where A is the amplitude of the LO signal.

Inspecting (58) above, one can determine that the bias current, the load resistance and the transistor transconductance, all contribute to the increase of the noise level of the mixer output. Conversely, a high LO amplitude is expected to reduce the output noise.

Similar to system linearity, system noise is an important parameter that relies on individual component performance. The formula that allows quick system noise factor estimation is:

$$F_{sys} = F1 + \frac{F2 - 1}{G1} + \frac{F3 - 1}{G1 \cdot G2} + \frac{F4 - 1}{G1 \cdot G2 \cdot G3} + \ldots \tag{59}$$

Equation (59) is also known as the Friis formula for system noise estimation, where F1, F2, F3 … and G1, G2, G3 … are the noise factors and the gains of stages 1, 2, 3, … counted starting from the system input.

3.1.4.4 Mixer Port to Port Isolation

Poor port to port isolation should primarily address the LO input due to its high level compared to the signal levels applied to the other ports. LO leakage towards the baseband input ports may result in DC offset for a receiving mixer. In a transmit mixer, insufficient isolation usually results in spurs that can be detected in band or out of band depending on the system configuration.

The remnants of the LO are known as LO feed-through. LO feed-through will occur from the LO port to the output port due to parasitic capacitance, power supply coupling, differential pair imbalance, on chip layout and so on. LO feed-through that is co-located with a carrier or present at the location of an out-of-channel measurement will have effects similar to those experienced with baseband carrier images. LO feed-through can be compensated or minimized by adjusting I and Q offset, and I/Q quadrature skew.

4 Circuit Implementation

The wideband up-converting mixer is designed in 130nm CMOS technology provided by IBM. This chapter describes the steps that were followed for the circuit implementation starting with passive and component characterization, transistors scaling, and pre/post layout system simulations.

4.1 MOS transistor characterization

The design starts with the characterization of a MOS transistor from the 130nm device library that will be used throughout the up-converter design. In order to start the iterations, a 20µm transistor was included in a test bench and simulated using SpectreRF. The schematic of the test bench can be seen in Figure 4.1 below.

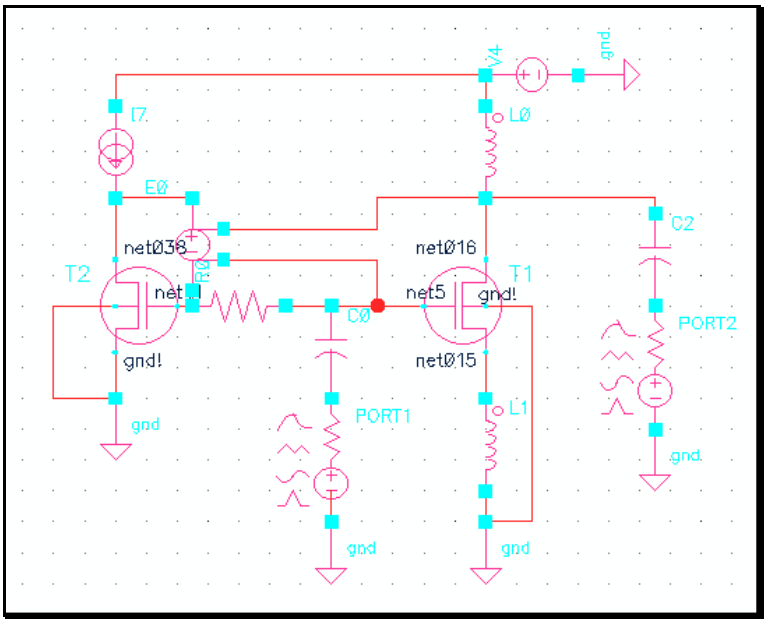

Figure 4.1 Schematic of MOS transistor test bench

Transistor T1 is the device under test (DUT). Transistor T2 together with the voltage controlled voltage source (VCVS) regulates the bias current of T1. Note that during transistor characterization the inductor L1 placed in series with T1 source is typically set to 0.

4.1.1 Gate resistance, R_G

A 20µm transistor using a single gate connected at both ends is expected to have a gate resistance of:

$$R_{SINGLE\_GATE} = \frac{1}{12} \cdot \rho_s \cdot \frac{W}{L} = \frac{1}{12} \cdot 7.6 \left[\frac{\Omega}{sq} \right] \cdot \frac{20[\mu m]}{0.12[\mu m]} = 105[\Omega] \tag{60}$$

However, this figure is too high to be used in an RF transistor. Instead, the gate was split in 20 gate fingers of 1µm each. This results in a reduction factor of 20x20=400 due to the width reduction as well as the parallel connection of the gates. The actual predicted gate resistance would then become:

$$R_G = \frac{R_{SINGLE\_GATE}}{20 \times 20} = \frac{105}{400} = 0.26\,[\Omega] \tag{61}$$

This is acceptable for the transistor and it is expected that the gate resistance should not be a significant factor in the maximum frequency, f_{MAX}, and in the noise figure, NF, of the transistor.

4.1.2 Gate to source capacitance, C_{GS}

C_{GS} is calculated based on the total gate surface and the unit capacitance of the gate above the silicon oxide insulator, weighted by a factor γ that is associated with the fraction of the gate not pinched-off. Typically, γ=2/3 for long channel transistors. For the 20µm transistor, however, in this calculation γ≈1 due to the short channel:

$$C_{GS} = \gamma \cdot W \cdot L \cdot C_{ox} = 1 \cdot 20\,[\mu m] \cdot 0.12\,[\mu m] \cdot 10\,[\frac{fF}{(\mu m)^2}] = 24\,[fF] \tag{62}$$

4.1.3 Transconductance, g_m

The transconductance at I_{DS}=2mA can be estimated as follows:

$$g_m = \sqrt{2 \cdot \mu \cdot C_{ox} \cdot \frac{W}{L} \cdot I_{DS}} = \\ = \sqrt{2 \cdot 500\,[\frac{cm^2}{v \cdot s}] \cdot 1\,[\frac{\mu F}{cm^2}] \cdot \frac{20[\mu m]}{0.12[\mu m]} \cdot 2\,[mA]} = 18.25\,[mA/V] \tag{63}$$

Note: In the simulations section ahead, it will be shown that in this particular case, the calculated g_m is a reasonably good fit. The actual g_m may sometimes be lower than the figure predicted by (63). Discrepancies could be the result of the reduction in electron mobility caused by the electric field as well as velocity saturation. It is generally recommended to perform device characterization and lab validation, and not relying solely on simple calculations.

4.1.4 Transition frequency, f_T

The transition frequency, f_T, is defined as the frequency at which the current gain becomes unity.

$$f_T = \frac{g_m}{2 \cdot \pi \cdot C_{GS}} = \frac{18.25[mA/v]}{2 \cdot \pi \cdot 24[fF]} = 121\,[GHz] \tag{64}$$

The result is a good fit compared to the simulation results shown in Figure 4.3.

4.1.5 Noise Factor, F, and Noise Figure, NF

The Noise Factor of a single 20μm transistor operated at 2mA and 5GHz can be calculated using a simple model as follows:

$$F = 1 + \frac{R_G + R_S}{R_S} + \frac{\gamma}{g_m \cdot R_S} + \frac{\delta \cdot g_m \cdot R_S}{5} \cdot \left(\frac{\omega}{\omega_T}\right) =$$

$$= 1 + \frac{0.26 + 8.3}{50} + \frac{0.67}{18.25 \cdot 10^{-3} \cdot 50} + \frac{1.33 \cdot 18.25 \cdot 10^{-3} \cdot 50}{5} \cdot \left(\frac{5}{67}\right) =$$

$$= 1 + 0.18 + 9.63 \cdot 10^{-3} = 1.19$$

The noise factor translates to a noise figure:

$$NF = 10 \cdot \log(F) = 0.75dB \tag{65}$$

The equation above is valid for a hypothetical common source amplifier using a single transistor. However, the Gilbert cell that is the subject of the design does not have matching components meant for noise optimization. The Gilbert cell mixer seen as a system requires a different approach for noise evaluation.

While noise performance is an important parameter in mixer characterization, in a transmit mixer, due to higher signal levels available compared to a receiving device, it is expected to be somehow easier to meet signal to noise ratio requirements.

4.1.6 Simulated results for several MOS device sizes

Aside from the calculated parameters for the 20um MOS transistor, a set of simulation were performed for the devices used in the implementation. The results are shown in Figure 4.2 and Figure 4.3 below.

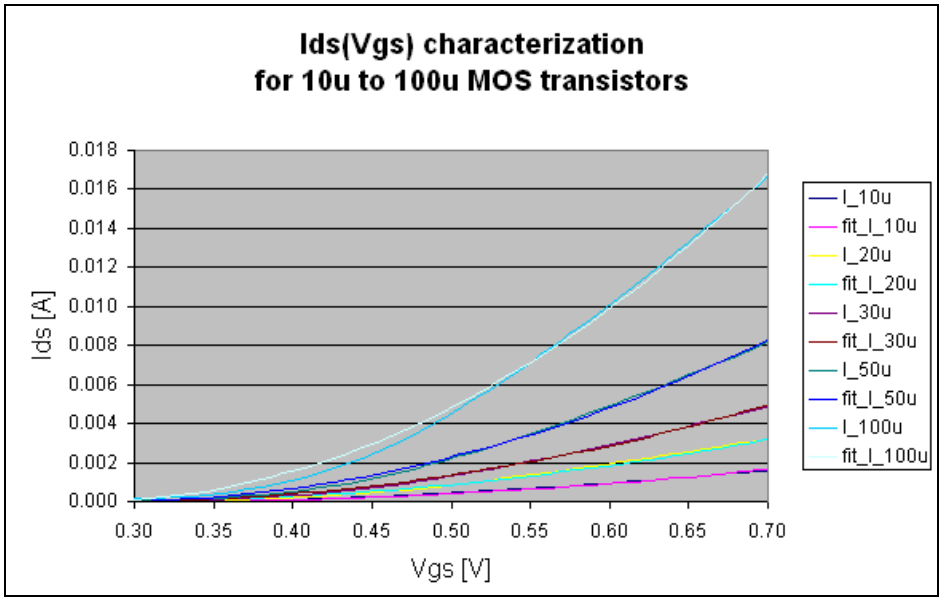

Figure 4.2 Ids vs. Vgs for several MOS device widths.

The characterization was used to extract the gm and Vt parameters used in the transfer function.

Table 2 g_m and v_T parameters determined based on curve fitting the simulated results

MOS size [um]	10	20	30	50	100
g_m [mA/V]	11	18	28	47	91
v_T [V]	0.312	0.28	0.28	0.28	0.27

It is worth noting that the maximum fT for any device is related to a particular current density as shown in reference [2]. Devices of different sizes achieve the maximum fT at different drain currents. The common parameter, though, is the current density which is approximately constant for a particular technology.

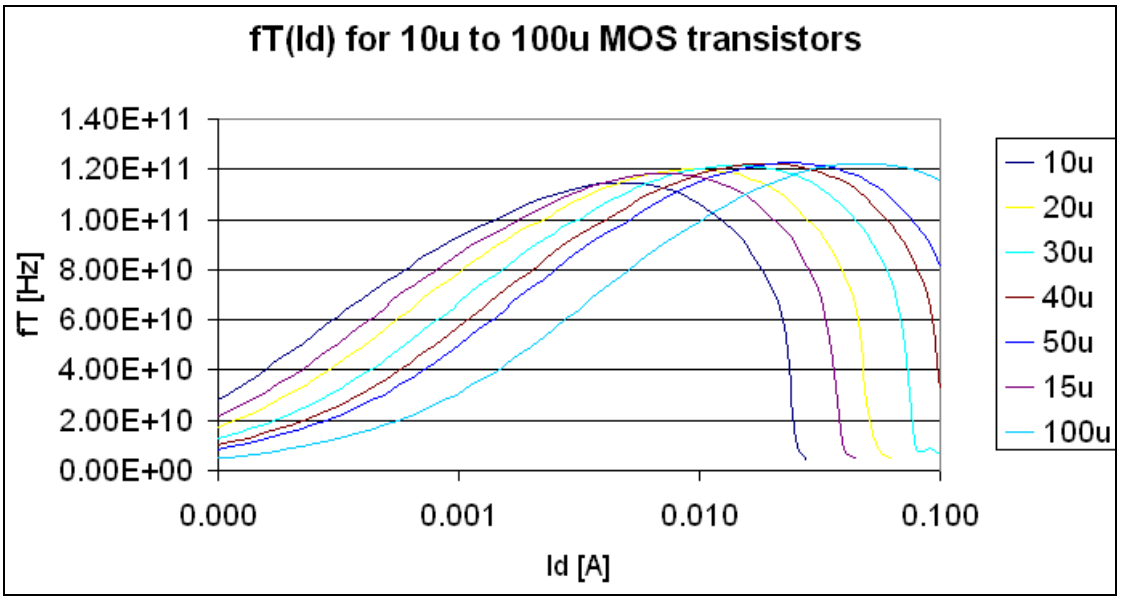

Figure 4.3 fT vs. Id characterization for several MOS device widths

An analysis of transistor sizes from 10um to 100um is shown in Table 3 below.

Table 3 Average Ids density for achieving the maximum fT

MOS size [um]	fT [Ghz]	Id [A]	current density [mA/um]
10	114.70	0.005	0.50
15	118.40	0.007	0.47
20	120.10	0.010	0.50
30	121.70	0.014	0.47
40	122.30	0.018	0.44
50	122.50	0.025	0.50
100	122.40	0.050	0.50
		Average Density [mA/um]	0.48

Table 3 above shows that for the 130nm process used in the design, the average current density required for the devices to achieve maximum fT is around 0.48mA/um. The average current density found above is consistent with the 0.3-0.4mA/um suggested in reference [2]. While some of the devices may not be operated at their optimal current density, it is good design practice to understand the limits of the technology and how to adjust the design for optimization purposes.

4.2 **Passive devices characterization**

The inductors provided in the 130nm IBM library may be connected with either terminal to the AC ground of the circuit. Due to the asymmetrical nature of the geometry, it is expected that small differences between inductance, L, and Q factors are shown for each inductor end is connected to the ground.

In the following simulations, L and Q were plotted for the same inductor in two configurations of ground connection.

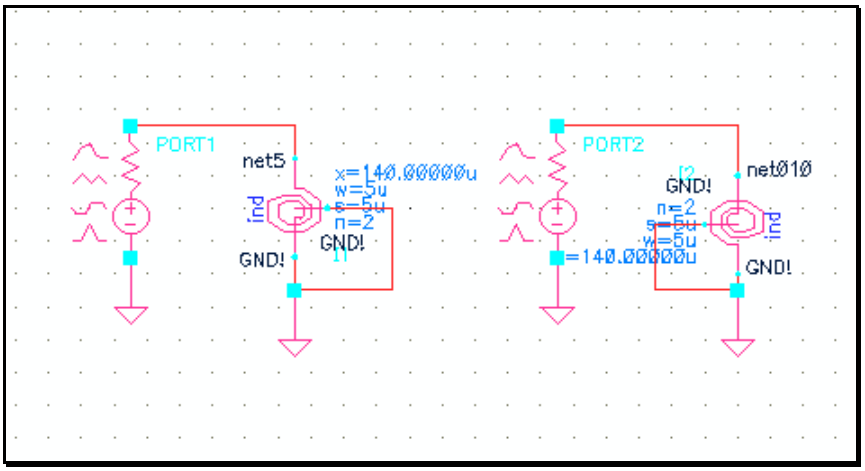

Figure 4.4 Inductor characterization setup

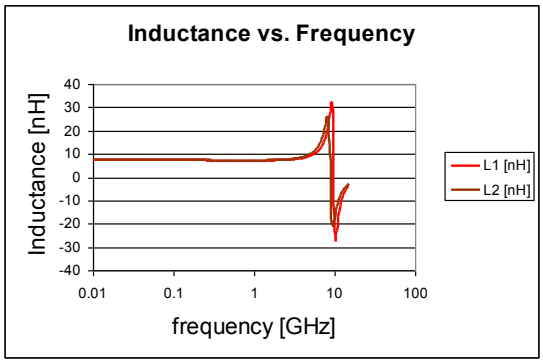

Figure 4.5 Simulated 8nH inductor

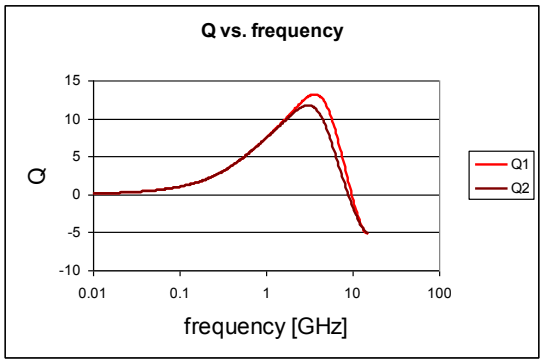

Figure 4.6 Simulated Q factor

In the pictures above, the simulated inductance and the Q factor can be analyzed. There are a few observations relevant to the design that can be made from the plots:

- The frequency range of interest, 0.5 to 6GHz, is far from the inductor's self resonant frequency.
- The actual inductance value does not depend significantly on the inductor ground connection
- At the designed operating frequency, the Q factor is also independent of the ground connection

- The slight differences, mostly at higher frequencies, are expected and are determined by the different parasitic capacitance in the two analyzed cases

We can conclude that the design behavior should not change significantly depending on the inductor AC ground connection.

Inductors are used in the design for matching between the multiplier output and the active balun input.

4.3 **Mixer Design**

In the following section, the detailed design procedure is presented.

4.3.1 Wideband 90˚ Phase Shifter

As shown in the previous chapters, the wideband 90˚ phase shifter consists of two cascaded CML D flip-flops. The flip-flops are configured as a "divide by two" counter that outputs precise quadrature signals. Due to the chosen configuration, the output frequency is one half of the input clock frequency.

4.3.1.1 D flip-flop

The D flip-flop schematic is shown in Figure 4.7 below. The initial design attempt was to operate the top transistors close to their maximum fT allowed by the technology. The voltage swing at the output needs to be large enough to allow the next stage to be fully switched on one side or the other. Setting a tail current of 4mA and using 200Ω load re-

sistors results in 800mV voltage swing. This is enough to completely switch on/off the MOS transistors.

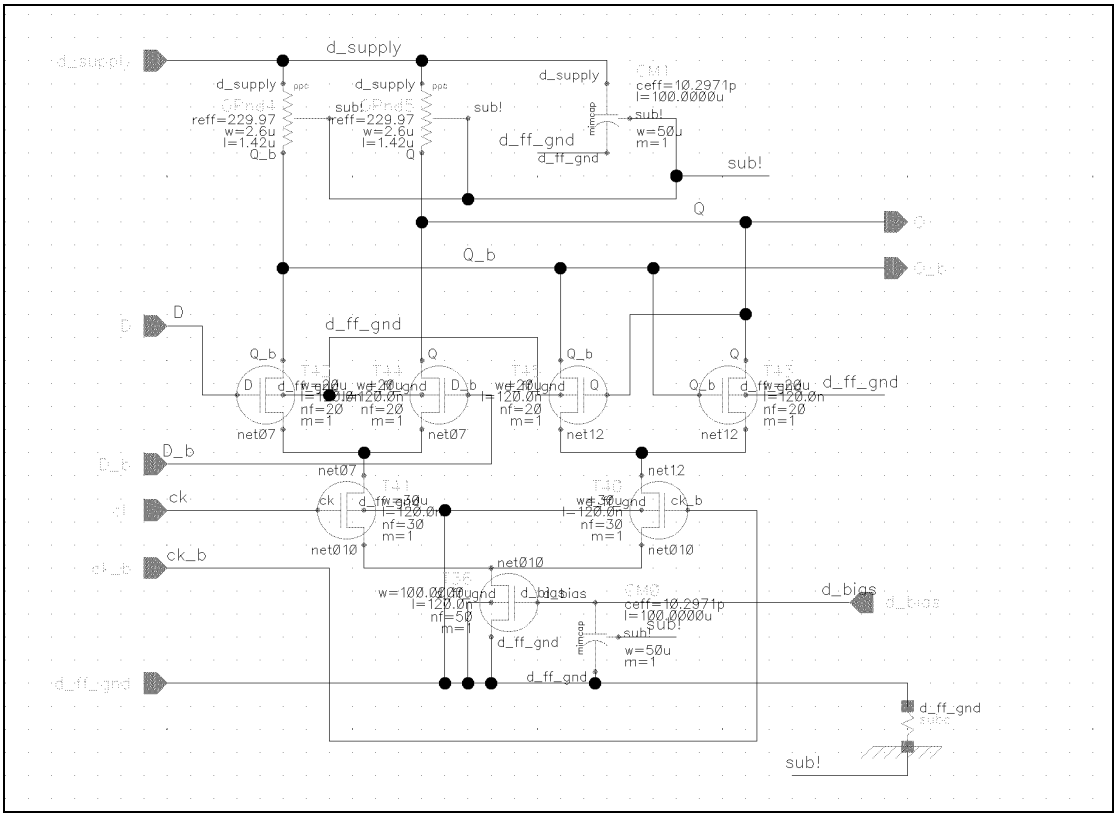

Figure 4.7 CML flip-flop used in the digital phase shifter

4.3.1.2 Buffered phase shifter schematic

It was determined through simulations that the phase shifter is sensitive to the capacitive loading at its outputs. High capacitance loads at its outputs significantly degrades the performance of the circuit in terms of maximum operating frequency. The phase shifter needs to simultaneously drive two Gilbert cells having a combined input capacitance on the LO side inputs of 50fF. Simulations have shown that the maximum capacitance tolerated by the D flip-flops is around 16fF. In order to overcome the fan-out limitation, a buffer amplifier stage is inserted between the digital circuit and the Gilbert cells. The gain of the buffer stage is low, but the main benefit is the reduced capacitance connected at the divider output.

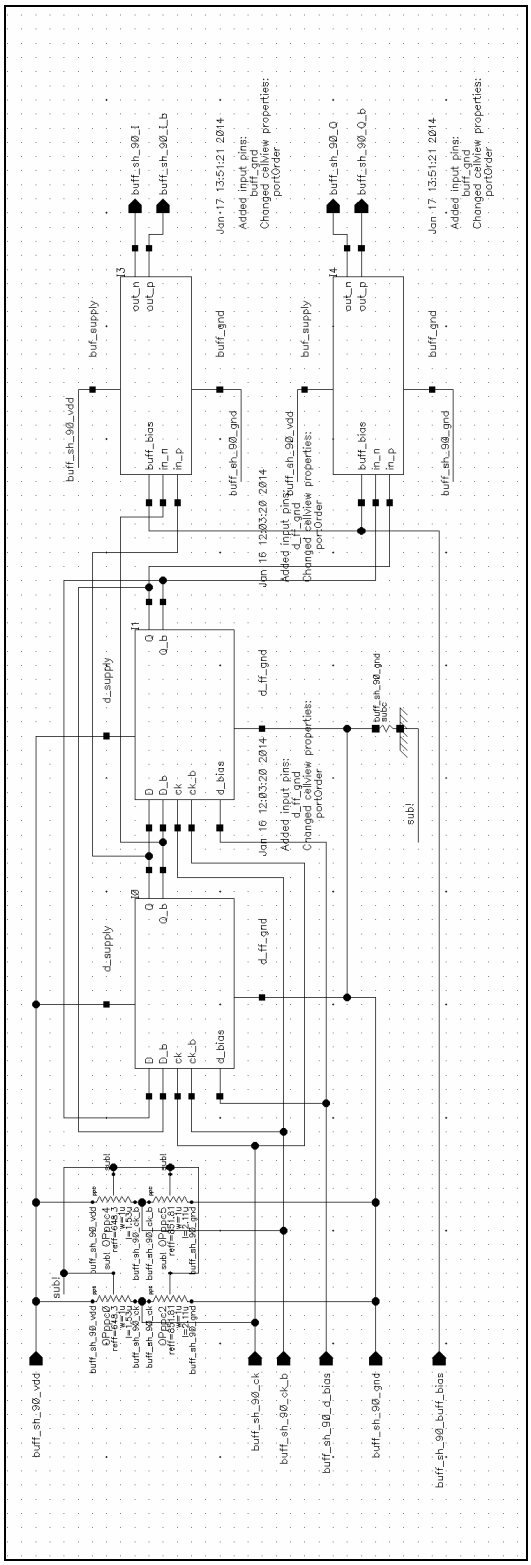

Figure 4.8 Complete schematic of the digital wideband phase shifter

4.3.2 Multiplier Topology and Bias Design

The chosen topology for the mixer multiplier is the Gilbert cell also known as double balanced mixer. It consists of the upper quad operated in switching mode and the lower differential stage that linearly converts the input RF voltage into current. Some of the advantages of the Gilbert cell are:

- Provided the magnitude of the LO applied to the upper quad is high enough to drive the transistors into switching, small LO level variations don't affect the mixer gain
- Once the upper quad is switched, the mixer behaves as a cascode amplifier
- The double balanced configuration reduces the LO/RF feed-through

The curve fit equation determined for the 100µm transistor is: I_{DS} [mA]=91*(V_{GS}-0.27)^2. The equation is generally valid for low current density. Operating at lower current density is acceptable for the mixer cells that operate at half the frequency required for the CML divider. The 100µm transistor will be used for the tail current generator. A 2mA current through the tail current source will be generated when V_{GS_tail} = 0.44V. Data is taken from simulation.

For proper operation, the MOS transistors need to bias in the saturation region. Therefore, the drain-source voltage should meet the condition:

$$V_{DS} \geq V_{GS} - V_T + V_{Swing} \tag{66}$$

This results in:

$V_{DS} \geq 0.44 - 0.27 + 0 = 0.17\,V$ for the current source. V_{Swing} is virtually 0 in the source of the RF transistor which is also connected to the tail current source drain.

For the RF transistors: $V_{DS} \geq 0.48 - 0.28 + 0.055 = 0.255\,V$

Where, for the RF transistors (30µm), 0.48V is the V_{GS} biased at I_D=1mA, V_T=0.28V and 0.055V is the 1dB compression voltage of the Gilbert cell.

Further, the minimum bias voltages for the RF and the switching transistors become:

$V_{G\_RF}=V_{GS\_RF}+V_{DS\_crt\_source}=0.44+0.17=0.61V$.

$V_{G\_SW}=V_{DS\_crt\_source}+V_{DS\_RF}+V_{GS\_SW}=0.17+0.255+0.5=0.925V$

For safety margin, the minimum bias voltages calculated above will be rounded up to 0.7V and 1V respectively. Later in the design, the saturation condition will be checked for the switching transistors.

For the RF transistors, the gate bias voltage is generated by a resistive divider. However, the divider is chosen to have high enough impedance since the intent is to drive the RF pair externally with the baseband signal, which may include DC offset adjustment for carrier feed-through optimization. There is no resistive bias attached to the switching pair since it will be driven from the 90˚ phase shifter circuit. It will be shown later on that the level generated by the phase shifter is high enough to completely switch on and off the switching pair.

4.3.3 Mixer Gain

The mixer gain can be expressed as:

$$A_{MIX} = \frac{2}{\pi} \cdot g_m \cdot R_D \tag{67}$$

Designing for a mixer gain $A_{MIX} = 1$ (or 0dB) and solving for R_D we obtain:

$$R_D = \frac{A_{MIX} \cdot \pi}{2 \cdot g_m} = \frac{1 \cdot \pi}{2 \cdot 4.63\,[mA/V]} = 339\,\Omega \tag{68}$$

In practice, with an output load resistance R_L, then R_D in (68) is replaced with R_{DL}, the equivalent impedance of the drain and load resistors in parallel, single ended.

$g_m = 2*I_{DS}/(V_{GS}-V_T)=4.6mA/V$

At this point we can check the saturation of the switching transistors. Based on a drain current I_{DS}=0.5mA, the voltage drop across R_D, $V_{RD}=I_{DS}*R_D$=0.165V, which leaves a voltage $V_{DS\_SW}=V_{DD}-V_{RD}-V_{D\_RF}$=1.8-0.165-1=0.635V $\geq$ $V_{GS}-V_T$ = 0.44-0.28=0.16V. Therefore the upper quad transistors operate in the saturation region.

4.3.4 Mixer Noise Figure, NF

The noise factor of a MOS transistors Gilbert cell based mixer measured in a 50Ω system can be expressed as [1], [26]:

$$F = 1 + \frac{R_D}{A_{MIX}^2 \cdot R_S} + \frac{2 \cdot \gamma}{g_m \cdot R_S} = 1 + \frac{339}{1 \cdot 50} + \frac{2 \cdot 1}{4.6mA/V \cdot 50} = 16.47$$
$$\Rightarrow NF = 10 \cdot \log(F) = 12.1dB \quad (69)$$

The following notations were used in Equation (69) above:

R_D – the Gilbert cell drain resistor
A – voltage gain
R_S – system impedance. 50 Ohm in this case.
$\gamma = 1$ – parameter related to the fraction of the gate not pinched-off
g_m – transistor transconductance

Note: It is expected that the mixer noise will be slightly higher than predicted using simple calculation formulas.

4.3.5 Mixer Linearity

The 1dB compression and the IIP3 input referred voltages of the mixer will be calculated for the RF differential pair [1], [26].

$$V_{1dB} = V_L = \frac{I_0}{g_m} = \frac{1mA}{4.6mA/V} = 217mV \tag{70}$$

$$V_{IIP3} = 3.266 \cdot V_L = 3.266 \cdot \frac{I_0}{g_m} = 3.266 \cdot \frac{1mA}{4.6mA/V} = 710mV \tag{71}$$

The following notations were used in Equations (70) and (71) above:

I_0 – the Gilbert cell tail current
g_m – transistor transconductance
VL – the input voltage at which I_0 has been completely switched through only one of the top transistor pairs and no further output voltage increase is possible

4.3.6 Gilbert Cell LO Input Level

The LO input level should be chosen in such a way that the upper quad of transistors is working as a switch. There are multiple benefits of switching the upper quad:

- Minimization of the noise contribution of the cascode transistors
- Maximization of the mixer gain
- Gain independence of the LO magnitude. This is particularly useful in image reject mixers

However, care should be exercised since increasing the LO input level will also increase the LO power leakage in the chip or even damage the switching transistors.

Simulations of the Gilbert cell gain vs. LO level will determine if the phase shifter output is high enough to optimally operate the mixer top quad in switching mode.

4.4 **Active Balun**

One of the challenges of the design is the implementation of a wideband balun that converts the differential output from the Gilbert cells to the single ended output of the device. The single ended output was chosen to simplify interconnects for initial testing, but it may also be the right choice if the mixer is required to drive a PA or directly an antenna. A passive solution using transformers is difficult to implement mainly because of the inherent limited bandwidth of such devices. Although compensation options might be available, they are not likely to cover from 500MHz to 6GHz; this is more than a decade. Other disadvantages of transformer based passive baluns are the insertion loss and the area that inductors and compensation capacitors require on a chip.

An interesting solution is suggested in reference [17]. Two MOS transistors, one in common-drain and the other in common-source configuration are connected as shown in Figure 4.9 below. Assuming that the transistors are being operated in their linear region, the superposition principle can be applied to determine the circuit transfer function. For brevity it is also assumed that the MOS transistors M1 and M2 are identical and their parasitic capacitances can be neglected. Note that in the process used, the source and drain of M1 could be connected together to avoid body effect, but only if triple well transistors were used.

If V(in_n)=0 and the RF voltage is applied at in_p only, one can write:

$$V_{in\_p} = V_{GS} + V_{out} \text{ and } V_{out} = g_m \cdot V_{GS} \cdot R_L = g_m \cdot R_L \left(V_{in\_p} - V_{out} \right) \tag{72}$$

Which results in:

$$V_{out} = V_{in\_p} \cdot \frac{g_m \cdot R_L}{1 + g_m \cdot R_L} \tag{73}$$

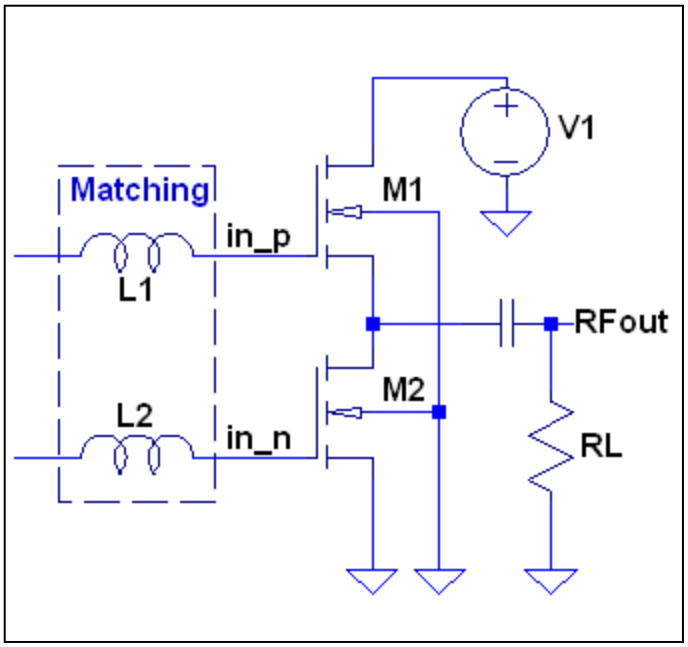

Figure 4.9 Active balun simplified schematic.

Alternatively, if V(in_p) = 0 and the RF voltage is applied at in_n only, one can write:

$$V_{out} = -g_m \cdot V_{GS} \cdot (R_L \parallel g_m) = -V_{in\_n} \cdot g_m \cdot \frac{R_L}{1 + g_m \cdot R_L} \tag{74}$$

Combining Equations (73) and (74) together, we get the output voltage as:

$$V_{out} = V_{in\_p} \cdot \frac{g_m \cdot R_L}{1 + g_m \cdot R_L} - V_{in\_n} \cdot g_m \cdot \frac{R_L}{1 + g_m \cdot R_L} \tag{75}$$

with the assumptions that:

- the two MOS transistors are identical, therefore $g_{m1}=g_{m2}=g_m$
- the drain impedance presented by M2 to M1 is much higher than RL
- $R_L = \frac{1}{g_m}$

If $V_{in\_p}= -V_{in\_n}=A$, meaning that the active balun is stimulated with two signals 180 degrees out of phase, then the output voltage V_{out} becomes:

$$V_{out} = V_{in\_p} \cdot \frac{1}{2} - V_{in\_n} \cdot \frac{1}{2} = A \cdot \frac{1}{2} - (-A) \cdot \frac{1}{2} = A \tag{76}$$

If V_{in_p}= V_{in_n}= A, meaning that the active balun is stimulated with two in phase signals, then the output voltage V_{out} becomes:

$$V_{out} = V_{in\_p} \cdot \frac{1}{2} - V_{in\_n} \cdot \frac{1}{2} = A \cdot \frac{1}{2} - A \cdot \frac{1}{2} = 0 \tag{77}$$

Equations (76) and (77) show that the active balun converts differential signals applied at its inputs into a single ended voltage, while rejecting the in phase input signals. The in-phase signals rejection behavior of the balun is useful for reducing the local oscillator feed-through of the Gilbert cells. The calculation is based on the assumption that the two MOS devices are identical. In practice, due to process variations, differences between the devices are expected. As a result, the signal rejection when in_p and in_n are stimulated with in-phase voltages should have a finite figure.

Based on the device characterization, and the requirement that $R_L = \frac{1}{g_m}$, the output balun is initially implemented using 30 μm devices operated at 3.5mA. Further optimization will be based on simulations.

In order to extend the top end of the operating frequency range, inductive peaking is used between the Gilbert cells and the input of the active balun circuit. The two inductors extend the operating range beyond 6GHz by acting as a low Q matching circuit.

4.5 **Bias Cells**

Most of the structures used in the design are differential; therefore they require a tail current generator for biasing. The simplest method to provide the biasing current is the current mirror. The current mirror is essentially a current amplifier. The characteristics of an ideal current amplifier are:

- Output current linearly related to the input current, Iout = A*Iin
- Input resistance is zero (Ideal current receiver)

- Output resistance is infinity (Ideal current source)
- There is no dependence on the power supply or temperature

While the characteristics above are ideal and impossible to meet in practice, there are circuit configurations that approximate the current generator behavior with sufficient accuracy. The current mirror will be briefly described in the following section.

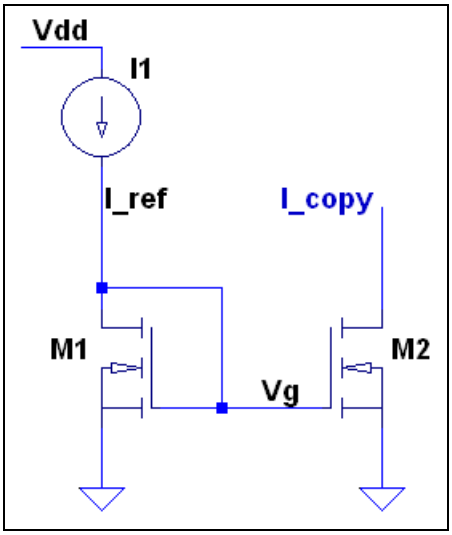

Figure 4.10 Schematic diagram of a simple current mirror

With reference to Figure 4.10, MOS transistor M1 generates the gate voltage Vg according to the following equation:

$$I\_ref = \frac{1}{2} \cdot \mu \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_{M1} (Vg - V_{TH1})^2 \cdot (1 + \lambda \cdot Vg) \tag{78}$$

And since λ is a small factor [22], its contribution can be neglected and

$$Vg = \sqrt{\frac{2 \cdot I\_ref}{\mu \cdot Cox \cdot \left(\frac{W}{L}\right)_{M1}}} + V_{TH1} \tag{79}$$

On the other hand, the output transistor, M2, generates its drain current according to the following equation:

$$I\_copy = \frac{1}{2} \cdot \mu \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_{M2} \left(Vg - V_{TH2}\right)^2 \tag{80}$$

The ratio between the copy (or the output current) and the reference (or input current) can be written as:

$$\frac{I\_copy}{I\_ref} = \frac{\frac{1}{2} \cdot \mu \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_{M2} \left(Vg - V_{TH2}\right)^2}{\frac{1}{2} \cdot \mu \cdot C_{ox} \cdot \left(\frac{W}{L}\right)_{M1} \left(Vg - V_{TH1}\right)^2} = \frac{\left(\frac{W}{L}\right)_{M2}}{\left(\frac{W}{L}\right)_{M1}} \tag{81}$$

And finally:

$$I\_copy = \frac{\left(\frac{W}{L}\right)_{M2}}{\left(\frac{W}{L}\right)_{M1}} \cdot I\_ref \tag{82}$$

Where the threshold voltages, $V_{TH1,2}$ were assumed to be equal since the transistors are on the sane silicon die.

Equation (82) shows that, if the width and length of the reference and copy transistors are the same, M2 generates the same current as the one supplied to M1. If the two transistors have different sizes, then the current ratio can be adjusted according to the transistor dimensions. This is useful when the supply voltage is relatively low and the tail current generator, M2 in Figure 4.10, can be increased in size in order to generate the same current, but at a reduced V_{DS}.

The design described in this work uses three current mirrors for biasing the CML D flip-flops, the phase shifter output buffer and the Gilbert cell multiplier. While an attempt was made to standardize the bias cell instances, in the optimization phase small bias current variations were required. Hence, an extra feature was added to the bias cells that allowed independent current adjustment both on chip and by externally applying resistors either to V_{DD} or to V_{SS}.

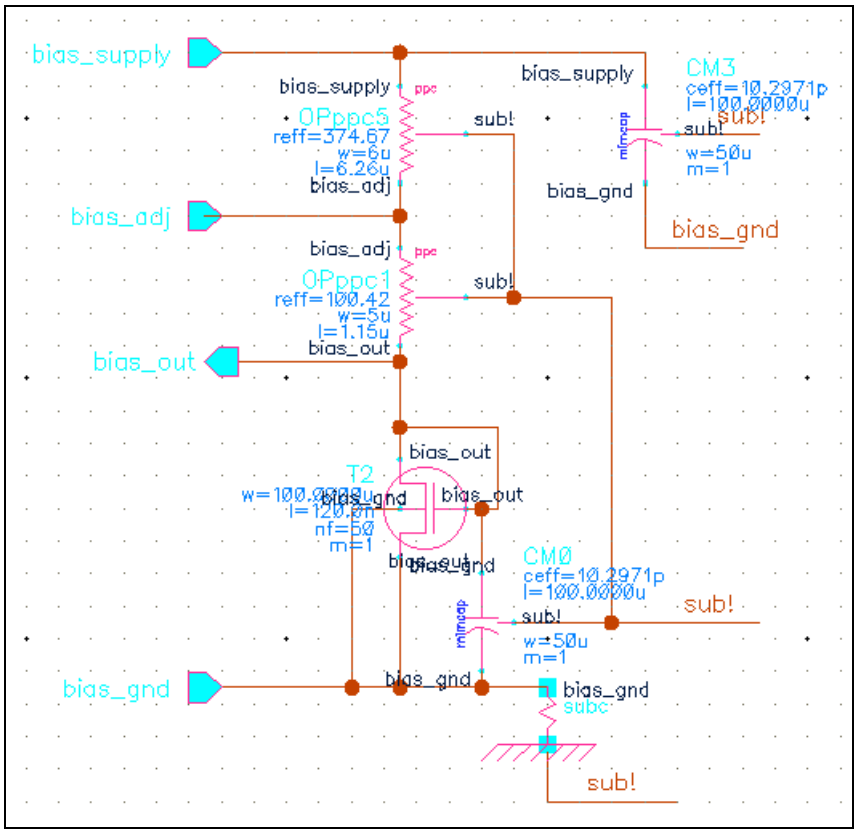

Figure 4.11 Bias Cell used in the design. The "bias_adj" net is connected to an external pad that allows further current adjustments in the testing phase.

As shown in Figure 4.11 above, by connecting resistors between bias_adj and V_{DD} or V_{SS}, the current passing through the reference transistor can be increased or decreased respectively. As a consequence, the voltage generated at "bias_out" can be adjusted, which in turn controls the tail current of their respective circuits.

The size of the transistors used in the current mirrors was increased in order to lower their minimum V_{DS} that allow them to operate in saturation. As well, non-minimum L can be used to increase the output resistance.

The input resistance can be calculated by applying the small signal circuit equivalent model to M1. With reference to Figure 4.10 we can write:

$$I\_ref = \frac{V_g}{r_{DS1}} + V_g \cdot g_{m1} \quad \text{and} \quad r_{in} = \frac{\Delta V_g}{\Delta I\_ref} = \frac{1}{g_{m1}} \parallel r_{DS1} \cong \frac{1}{g_{m1}} \tag{83}$$

As an example, if the reference MOS transistor is operated at a drain current that results in a transconductance gm=5mA/V, the resulting current mirror input impedance is 200Ω.

The output resistance can be determined by applying a similar analysis on M2. But, since the gate voltage, Vg, is a small signal ground, the output resistance is:

$$r_{out} = r_{DS2} = \frac{1}{\lambda \cdot I_D} \tag{84}$$

where λ is the channel length modulation parameter. Reference [22] shows that the calculation of λ from the device structure is quite difficult due to the fact that the field distribution in the drain depletion region is not one-dimensional. As a result, the effective values of λ need to be determined from experimental data. The parameter λ is inversely proportional to the effective channel length and a decreasing function of the doping level in the channel. Typical values of λ are in the range 0.05 V^{-1} to 0.005 V^{-1} resulting in Rout values in the range of tens to hundreds of KΩ.

4.6 Gilbert Cell Simulations

The verification and simulation starts with the Gilbert cell as a standalone module. The bias and operating points using a DC and a transient analysis are being verified. In order to increase the simulation speed, the LO voltage applied to the mixer was exported after the separate simulation of the phase shifter. A Periodic Steady State (PSS) analysis is used for full characterization of the mixer performance. Snapshots of the result plots are taken to illustrate the performance achieved. The plots included in the report show the performance after optimization. Consequently, some differences from the calculations should be expected.

4.6.1 Gilbert Cell Bias and Transient Simulation

The schematic diagram and the DC operating points of the basic Gilbert cell can be seen in Figure 4.12.

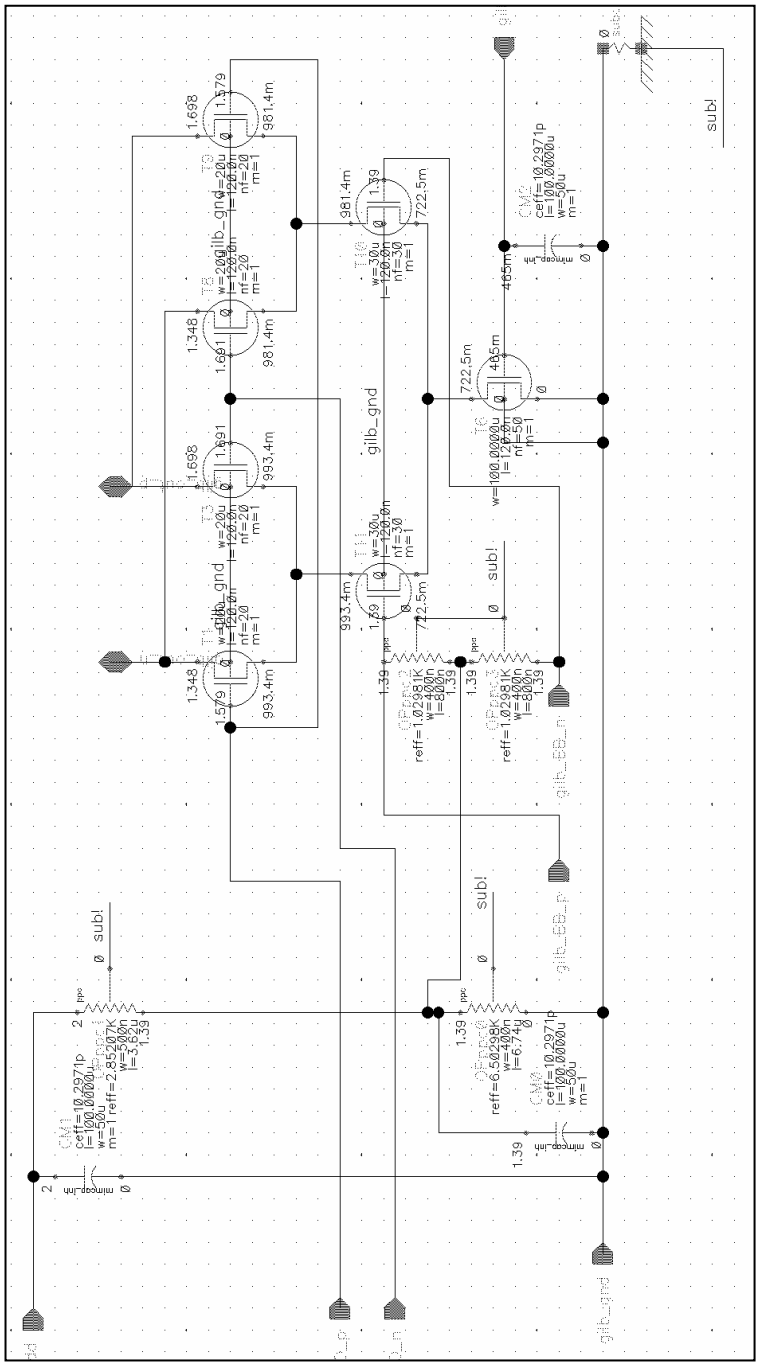

Figure 4.12 Gilbert Cell Schematic and DC bias

The correct bias and operation of the mixer is tested by applying a 10GHz clock at the mixer input and a 20MHz I/Q signal on the baseband inputs. A transient simulation of a full cycle of the baseband 20MHz is shown in Figure 4.13 below. The top trace shows the

LO signal generated from the wideband phase shifter, in the middle, the 20MHz base-band cosine/sine and the bottom trace shows the modulated RF output. The small enve-lope variation is the result of the finite sideband rejection due to circuit impairments.

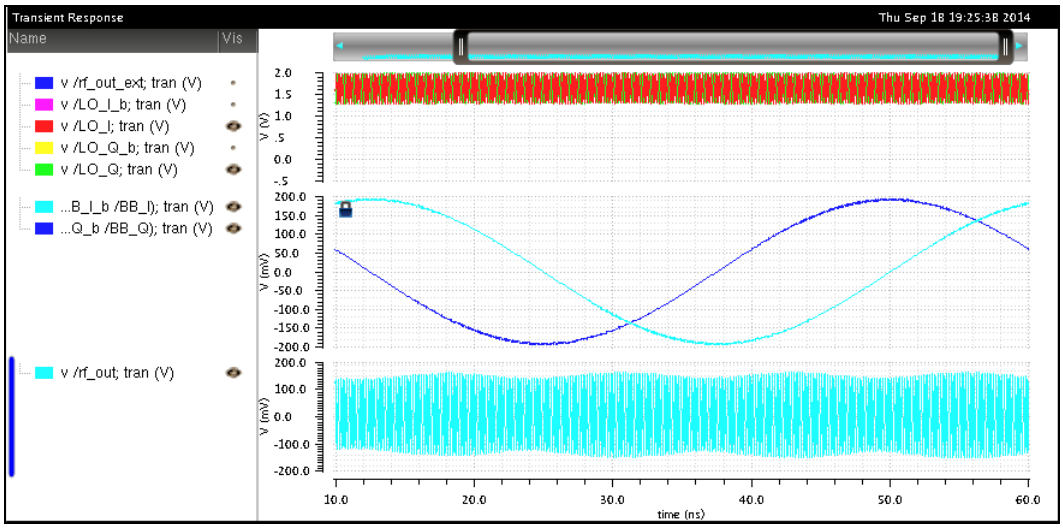

Figure 4.13 Transient simulation of the mixer for a full period of the 20MHz BB signal

A zoomed-in version of the LO, baseband and RF output are also shown in Figure 4.14. It is noticeable the 90˚ phase shift between the LO-I and LO-Q signals generated by the phase shifter.

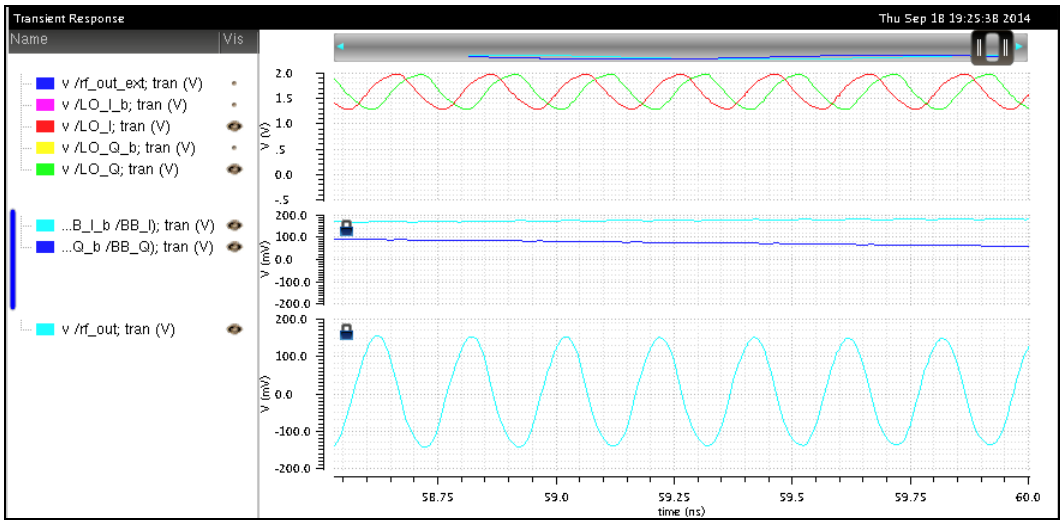

Figure 4.14 Transient simulation. Zoomed-in capture that reveals the LO I/Q signals.

Magnitudes of the undesired side-band rejection will be shown in the following sections using PSS simulations.

Note: Based on the transient plots shown above, one can conclude that the choice of bias voltages maintains the transistors in the saturation region.

4.6.2 Complete Mixer Gain

The simulated gain of the complete I/Q mixer is shown in Figure 4.15 below. Markers are added in order to show the undesired sideband and carrier rejection.

Due to the complexity of the circuit, the multiplier section is simulated separately from the digital phase shifter, otherwise the simulation would take too long or it would reach the maximum amount of available memory. For this purpose, the phase shifter output resulting form separate simulations were stored in several PWL files then used as stimuli for the Gilbert cells. This approach allowed the simulation to complete.

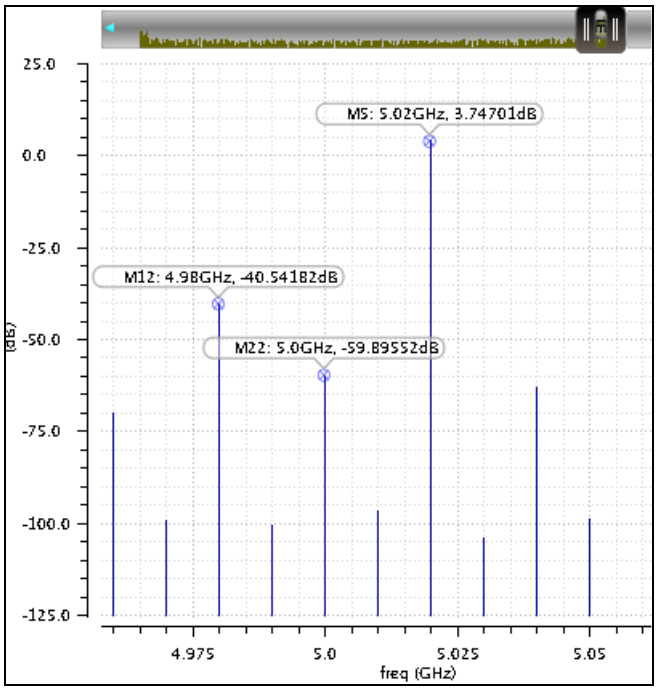

Figure 4.15 Mixer Gain = 3.74dB for f_{out}=5.02GHz, f_{LO}=5GHz, V_{LO}=340mV_p and f_{BB}=20MHz, V_{BB}=100mV_p. Also shown are the unwanted sideband at 4.98GHz and LO rejection.

Notes: The simulated mixer gain is slightly higher than the 0dB gain initially tergeted in section 4.3.3 based on transistors characterization. The difference can be explained as follows:

- The initial calculations took into account a single Gilbert cell. In the actual design, the output power is the result of both I and Q power contributions.
- An additional gain of approximately 6dB is expected: 3 dB from the combined power of the two Gilbert cells and another 3dB from the contribution of the rejected sideband. The energy from the unwanted sideband can't simply vanish. Instead, it constructively adds to the desired sideband, increasing the power twofold or by 3dB (See Equation (12) for justification).
- Combining together the -3.9dB calculated gain with the 6dB explained above results in 2.1dB overall gain which is closer to the simulated 3.7dB

The Mixer Gain vs. LO peak voltage dependence is shown in Figure 4.16 below.

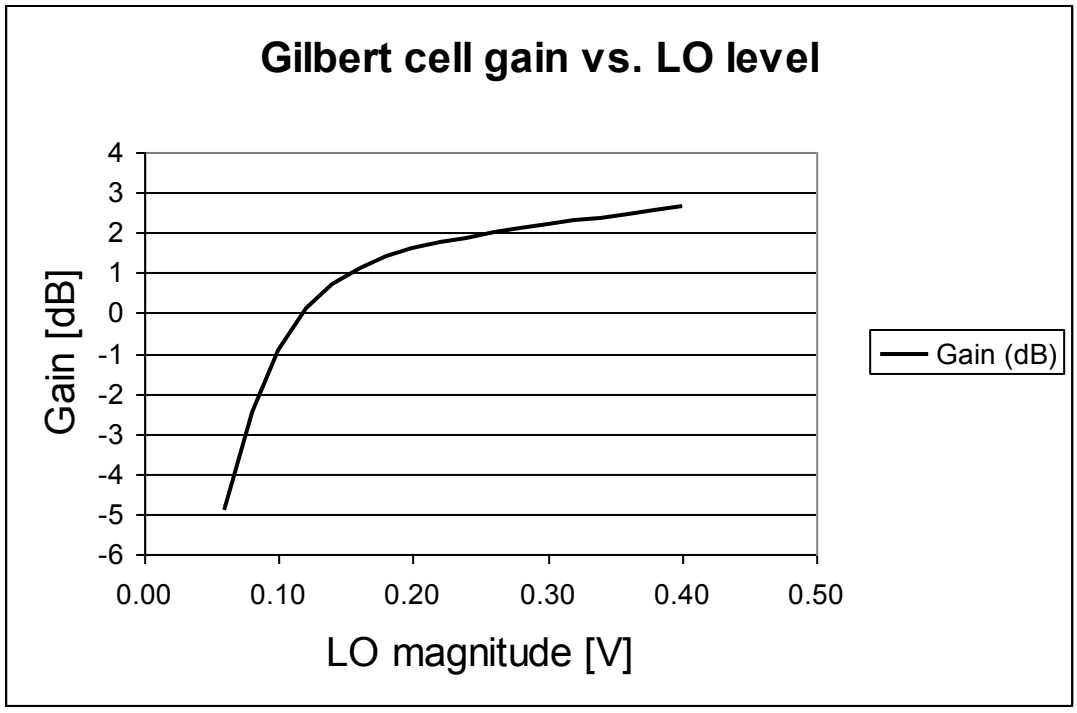

Figure 4.16 Mixer Gain vs. Local Oscillator magnitude. f_{LO}=5GHz, f_{BB}=20MHz, f_{RF}=5.02GHz, V_{BB}=100mV

From the plot it can be determined that the 340mV LO level available from the phase shifter is high enough to drive the upper quads into switching. While further LO magnitude increase may add up to 1dB to the gain, other adverse effects such as carrier leakage may take place.

4.6.3 Complete Mixer Noise Figure

The simulated noise performance of the complete chip is shown in Figure 4.17 and Figure 4.18 below. Both simulations were performed using a 5GHz carrier.

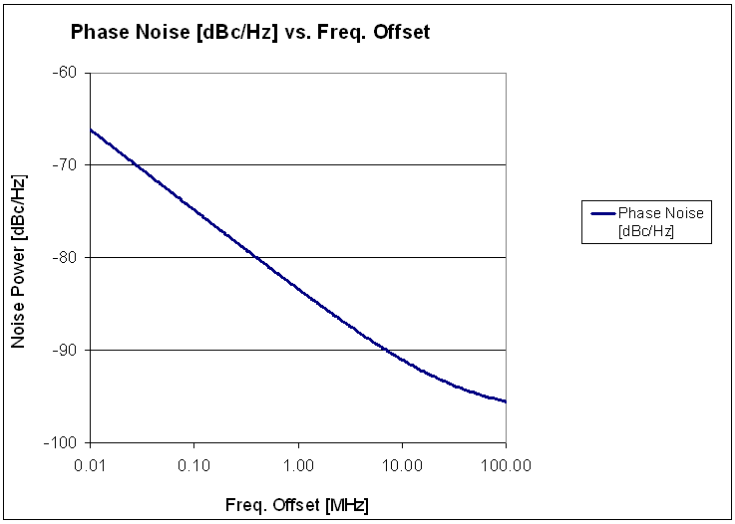

Figure 4.17 Mixer Phase Noise vs. Frequency Offset. Simulated at 5GHz carrier.

Figure 4.17 shows that, as predicted by theory, the noise power density is higher at small offsets from the carrier and decreases as the frequency offset from the carrier becomes higher due to 1/f noise in the input circuit.

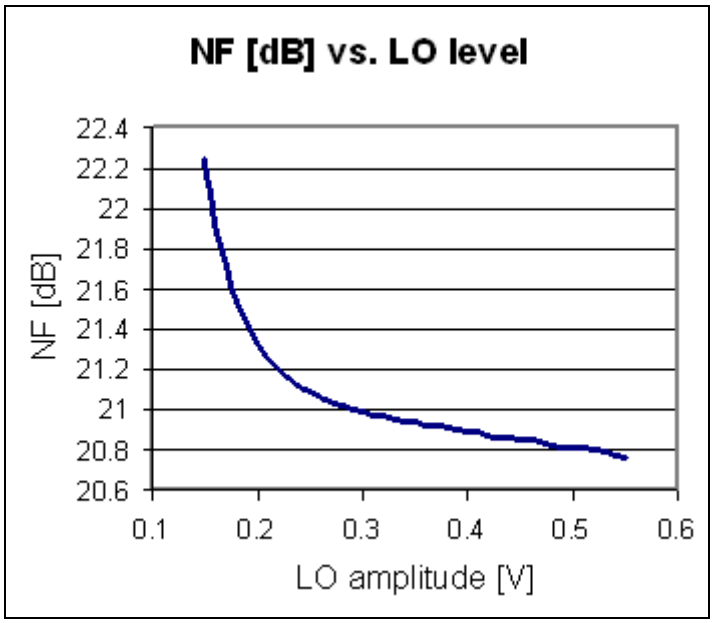

Figure 4.18 Mixer Noise Figure vs. LO level applied at the Gilbert Cell LO Port. Simulated at 5GHz carrier.

In Figure 4.18 one can note that a significant impact on the noise figure has the LO driving level. The LO level should be chosen such that the upper transistors are operating as switches. Note that the noise is higher than predicted as expected considering that 1/f noise was shown to exist in Figure 4.18, but not included in the prediction.

Switching mode allows optimization of both the Noise Figure (see Figure 4.18) and the conversion gain (see Figure 4.16).

The interpretation of Figure 4.18 is that beyond 300-400mV LO amplitude, no significant NF improvement should be expected. Further increasing the driving level may increase the LO leakage and may cause reliability issues to the mixer.

In this design, the LO voltage applied at the two Gilbert cells is provided by the wideband phase shifter. Transient simulations of the phase shifter showed that the signal level is around 340mVp, therefore, the mixer is driven close to its optimal operating point.

4.6.4 Complete Mixer Linearity

For mixer linearity test, a Periodic Steady State (PSS) simulation was setup in SpectreRF. Separate simulations were performed for compression point and for third order intercept point measurements.

4.6.4.1 Compression point simulations

The following parameters were used in the P1dB simulations:

- f_{CLK} = 10GHz, the input frequency applied at the digital phase shifter input
- f_{LO} = 5GHz, the local oscillator frequency generated by the phase shifter and applied to each LO port of the two Gilbert cells
- f_{BB1} = 20MHz, a single baseband sinusoidal component
- p_BB_pwr = a simulation parameter, in dBm, that was swept for linearity testing purposes. The parameter represents the power of a single baseband tone applied at the mixer input.

The input referred P1dB simulated results are shown in Figure 4.19 below.

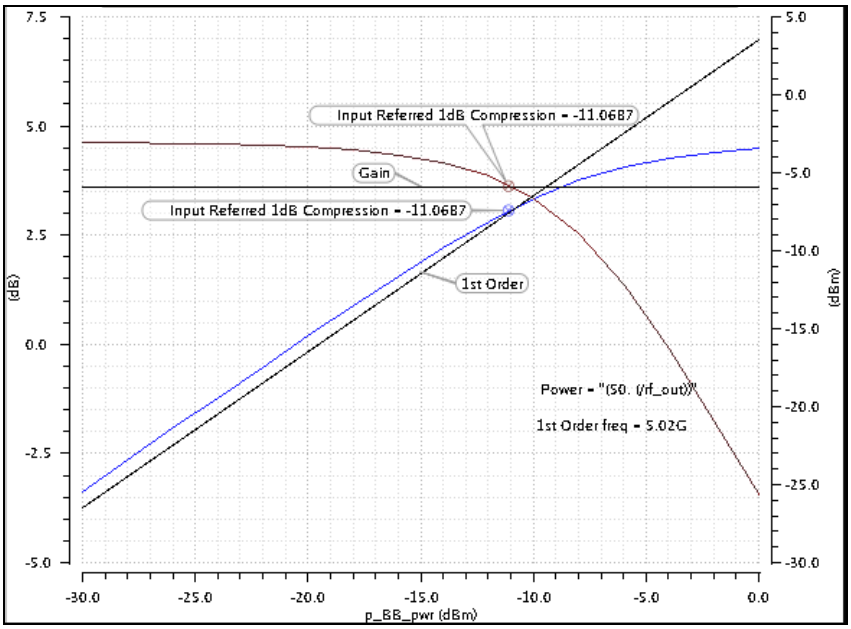

Figure 4.19 Gain compression and input referred P1dB

The following frequency was monitored at the RF output port of the mixer:

- f_{RF1}=5GHz+20MHz=5.02GHz

The plot shows that the input referred 1dB compression point is -11dBm in 50 Ohm. This power level translates into 178mVpp.

4.6.4.2 Third order IMD and Intercept Point simulations

The following parameters were used in the IIP3 simulations:

- f_{CLK} = 10GHz, the input frequency applied at the digital phase shifter input
- f_{LO} = 5GHz, the local oscillator frequency generated by the phase shifter
- f_{BB1} = 20MHz, the first baseband sinusoidal component
- f_{BB2} = 30MHz, the second baseband sinusoidal component
- p_BB_pwr = a simulation parameter, in dBm, that was swept for linearity testing purposes. The parameter represents the power of one of the two tones applied at the input. Both tones had equal powers.

The following frequencies were explored at the RF output port of the mixer:

- f_{RF1}=5GHz+20MHz=5.02GHz, the result of the 20MHz baseband tone
- f_{IF2} = 5GHz+30MHz=5.03GHz, the result of the 30MHz baseband tone
- Left side IMD3 frequency: $f_{IMD3\_left} = 2*f_{RF2}-1*f_{RF1}$= 5.01GHz
- Right side IMD3 frequency: $f_{IMD3\_right} = 2*f_{RF1}-1*f_{RF2}$= 5.04GHz

The fundamental tones and the IMD3 simulated results are shown in Figure 4.20 below

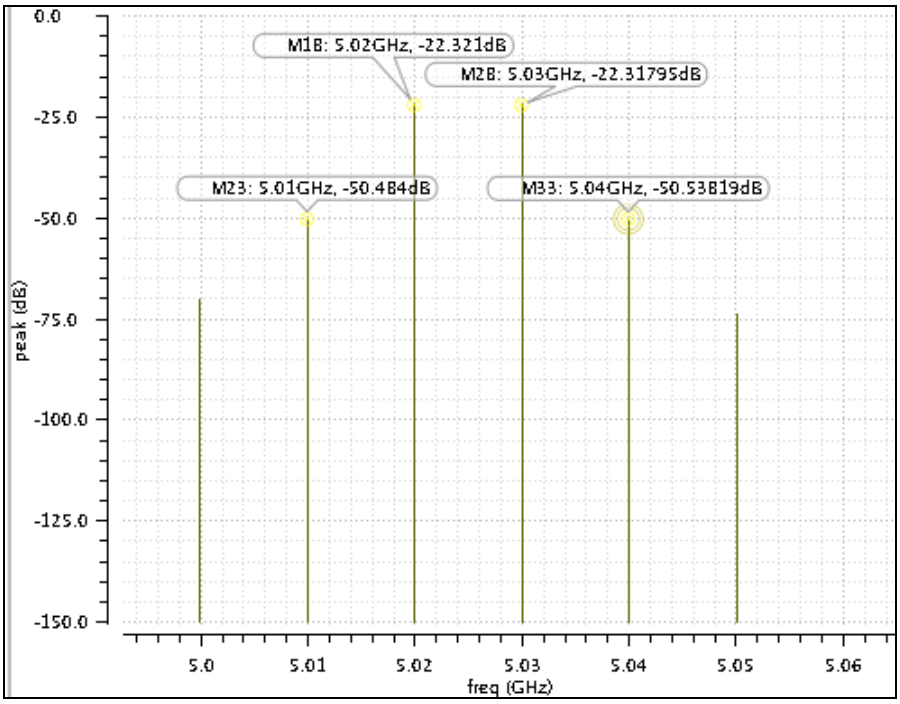

Figure 4.20 Fundamental Tones and Intermodulation Products in 2 tone testing simulation

The third order intercept points are shown separately for the left hand side (Figure 4.21) and the right hand side (Figure 4.22) IMD tones. The results are reasonably balanced, within 0.3dB from each other. Large differences between the left/right IP3 figures should usually be regarded as design issues that point to phase and/or amplitude impairments in the circuit.

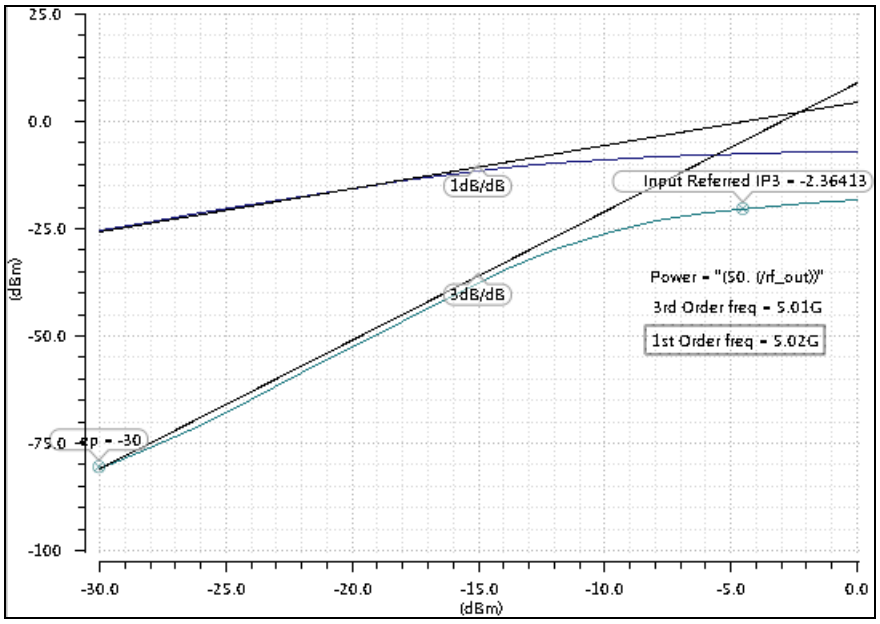

Figure 4.21 Left IMD component input referred third order intercept point

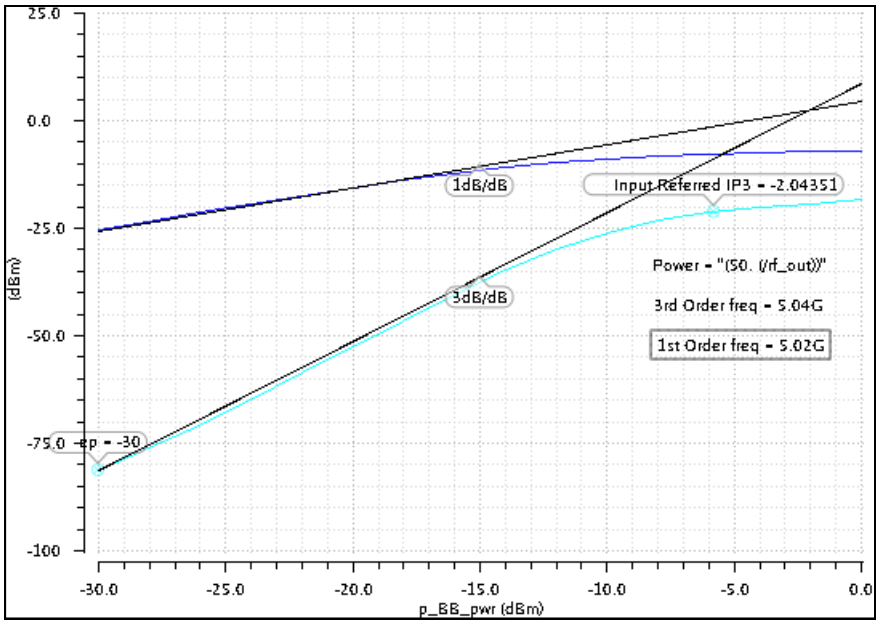

Figure 4.22 Right IMD component input referred third order intercept point

The plot shows that the input referred 1dB compression point is -2.3dBm in 50 Ohm. This power level translates into 485mVpp.

Note: The simulated figures are lower than the estimated ones due to non-linearity not accounted for in the simple calculations. A side by side comparison between calculated and simulated values is shown in the table below.

	Calculated	Simulated
P_1dB	-9.3 dBm	-11dBm
P_IIP3	1 dBm	-2.3 dBm

4.7 Chip Reliability Considerations

While achieving the best possible chip performance is the main focus of the designer, there are specific issues that need to be taken into account in order to make sure the production yield reasonably high. If not observed properly, there is a risk that the MOS devices may be damaged even before the silicon die is fully built. In the following section, some of the techniques applied for chip reliability improvement will be described briefly.

4.7.1 Antenna Rule

The antenna rule is part of the DRC required in the 0.130μm IBM process to ensure that the on die MOS transistors are not destroyed during the fabrication process. In such processes, the wafer containing the future IC dies is bombarded with ions in order to create the poly-silicon and metal layers. These ions need to find a path through the wafer and to eventually leak to the substrate and active layers at the bottom of the wafer. If there is a large area of poly or metal, and if it connects ONLY to a transistor gate, not to source or drain or any other conductive material, then these ions will travel through the transistors.

If the ratio of the poly or metal layers to the area of the transistors is too large, a random discharge of the floating node may result in the transistors being destroyed.

Solutions to reduce Antenna effects

Routing options

- The signal wires should be broken in smaller segments and routed to upper metal layers by inter layer vias insertion
- Etched metal areas should not be connected to MOS gates until the last metal layer is etched on the chip

Dummy transistors inclusion

- Addition of extra gates will reduce the capacitance ratio between the metal traces and the transistor gate
- However, this solution may lead to the Problem of Reverse Antenna Effect as described in reference [18].

Embedded Protection Diodes

- Connecting reverse-biased tie-down N+ diodes to the transistor gates susceptible to antenna effect creates a discharge path for the charge accumulation, thus preventing excessive voltage to damage the gate oxide
- In normal circuit operation, the diodes do not significantly affect functionality
- Simulations should be run after adding the tie-down diodes in order to confirm that the impact on functionality is minimal.

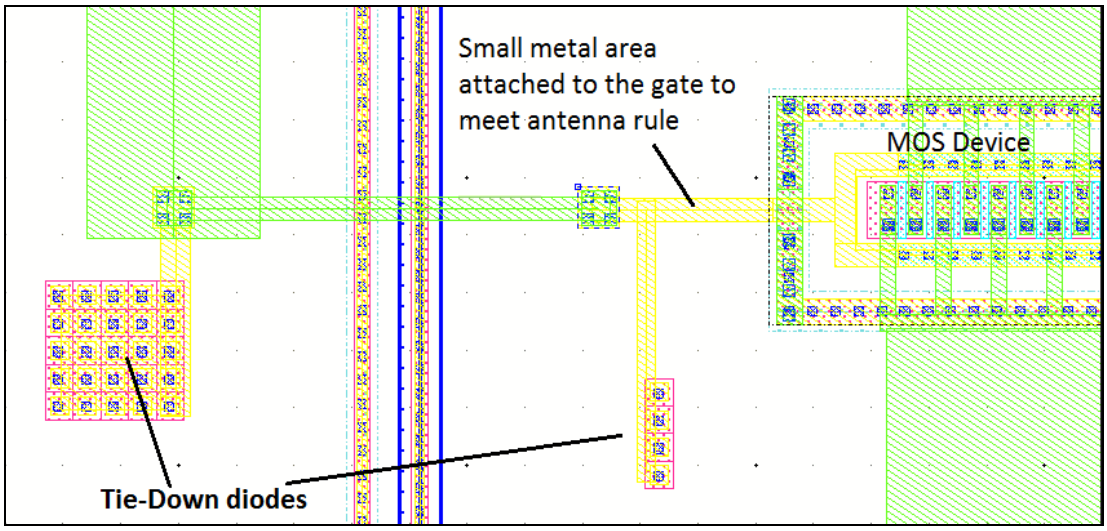

Figure 4.23 Tie-Down diodes and routing option that meets antenna rule

4.7.2 ESD Protection

Electrostatic Discharge (ESD) is a transient release of static charge that arises from either human handling or a machine contact. The continuous decrease of the physical dimensions and increase in doping in modern CMOS technology resulted in a significant decrease in gate-oxide thickness and pn-junction width. As a consequence, the small size devices require less energy and lower voltages to destroy MOS transistors.

Although ESD is the result of a static potential in a charged object, the energy dissipated and damages made are mainly due to the current flowing through the sensitive components during discharge. The majority of the ESD damages are thermally related in the form of device interconnect burn-out or oxide break-down. The basic phenomenon of ESD is that a large amount of heat is generated in a localized volume significantly faster than it can be removed. As a consequence, the dissipated energy leads to a temperature increase that is in excess of the materials' safe operating limits.

ESD testing and modeling is an important area of research. We briefly mention some of the modeling techniques used in the industry: Human Body Model, Machine Model, Charged Device Model and Transmission Line Pulse Model. ESD study is outside the scope of this work, but the reader can find more information by consulting the existing standards and publications in the field.

The 0.13µm IBM CMOS process provides an ESD library containing a whole range of components designed for ESD protection of the fabricated chips.

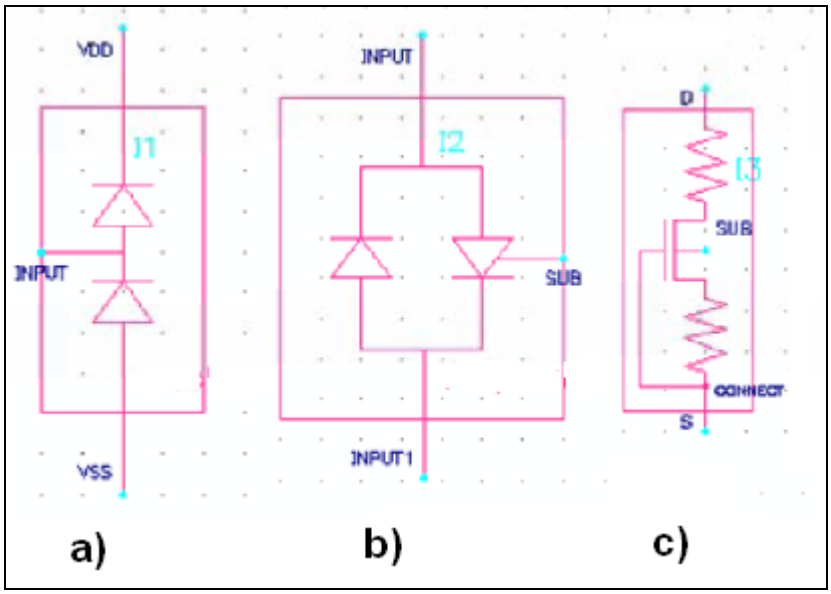

Figure 4.24 ESD Input Protection Components: a) double diode; b) antiparallel diodes; c) ESD-NFET

All the I/O pads of the design are protected with double diode as shown in Figure 4.24 a). The electric charge resulting from an ESD event that may happen at one of the inputs, is provided with low impedance paths towards the Vdd/Vss rails. The expectation is that the low impedance path prevents the damage that may occur at the transistor gate level.

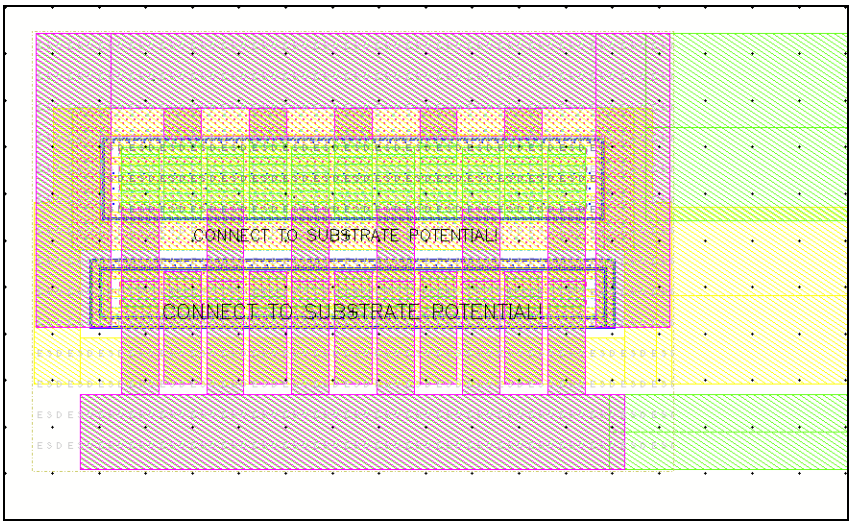

Figure 4.25 ESD double diode structure on the layout

4.7.3 Latch-up Protection

Latch-up is a failure mechanism found in CMOS integrated circuits which manifests itself by excessive drain current further associated with functional failure, performance degradation and, in severe cases, device destruction. Latch-up may be a temporary condition that ends upon removal of the triggering stimulus, a catastrophic condition that requires the shutdown of the system power supply or a fatal condition that requires replacement of damaged parts. Due to its potential to reduce component reliability and MTBF of the system, the components must be designed to avoid latch-up. Fortunately, this is possible in most of the designs.

The opportunity for latch-up exists in all junction-isolated or bulk CMOS processes. It is caused by the inherent formation of parasitic PNPN paths which act as an SCR (also known as thyristor).

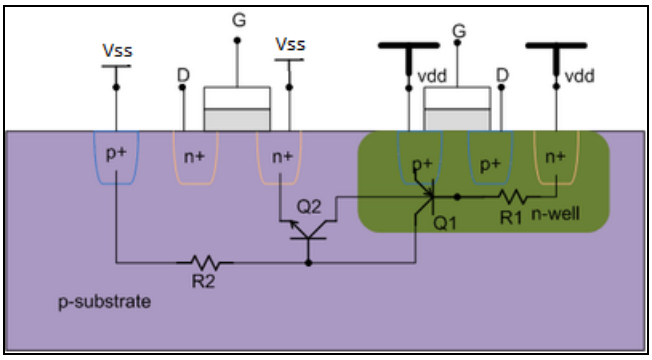

Figure 4.26 Cross-section of the CMOS process, emphasizing the formation of the parasitic bipolar transistors in an SCR type of configuration

Figure 4.26 shows a typical CMOS process cross section manufactured on N-substrate silicon. The parasitic NPN and PNP bipolar transistors in an SCR configuration are also shown that usually trigger the latch-up phenomena. The P+ sources and drains of the PMOS devices act as the emitters of the lateral PNP device. On the other hand, the N-well creates the base of this bipolar device and the collector of a vertical NPN device.

The P-substrate acts as the collector of the PNP and the base of the NPN. Finally, the N+ sources and drains of the N-channel MOS devices serve as the emitter of the NPN. The substrate is normally connected to Vss, the most negative voltage in the circuit, using an P+ diffusion tap while the N-well is attached to the Vdd node, the most positive circuit voltage, through a N+ diffusion. These power supply connections involve parasitic resistance to all points of the substrate and N-well, as depicted in Figure 4.26.

A range of successful counter measures are currently applied in the IC industry in order to prevent the destructive effects of the latch-up. Guard banding, device placement, the installation of pseudo-collectors between the P-channel devices and the N-well, and the use of a low resistivity substrate under an epitaxial layer are a few of the IC design tactics now being practiced to reduce the current gain or to control the action of the lateral PNPN structures in state-of-the-art CMOS devices.

By using this approach, structures which would normally be part of the latch-up loops are decoupled and are less of an issue. Even from the design phase, the CAD packages used

for implementation use elaborated Design Rule Check (DRC) scripts that are able to identify potential latch-up issues. Latch-up prone areas are protected by similar geometries where any risk is significant [19], [20], [21].

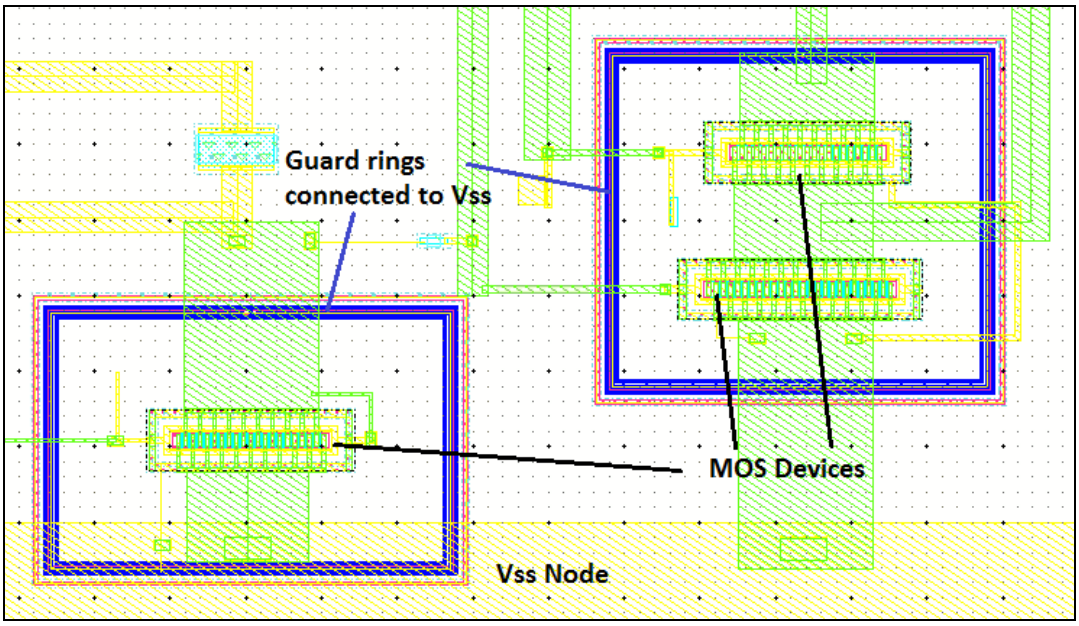

Figure 4.27 Guard rings placement around MOS transistors to prevent latch-up

Figure 4.27 above shows an example area in the current design where a guard ring is in place in order to protect the device from latch-up events.

4.8 **Layout of the complete design**

The layout of the wideband up-converting mixer is carried out in a 0.13 μm CMOS eight layer process and implemented with the assistance of the Canadian Microelectronics Corporation (CMC). The modular design strategy during the schematic and layout generation allowed for fast turnaround validation simulations and modifications. The design is divided into smaller functional cells and then integrated together to form the complete mixer. Layout of the mixer is depicted in Figure 4.28.

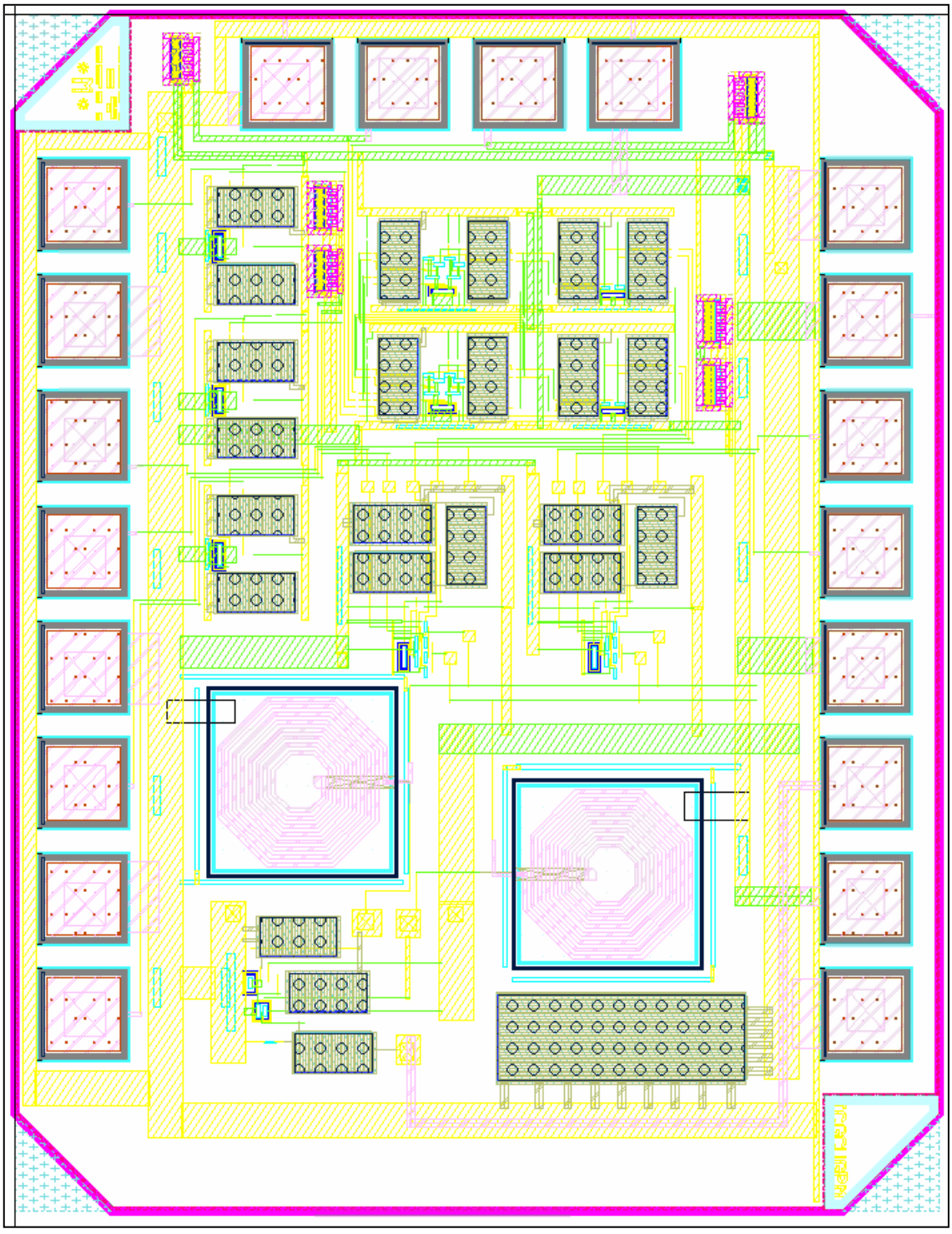

Figure 4.28 Layout of the complete upconverting mixer.

Guard rings are visible around the whole circuit and around matching inductors to protect them from noise coupling and provide effective isolation. DC bypass capacitors are scattered around the chip in order to improve the isolation between modules. DRC was run to validate the design rules of the process and LVS to match the schematic of the mixer with layout before parasitic RC extraction. The simulation results presented in section 4.6 are based on the extracted schematic therefore it includes the parasitic elements.

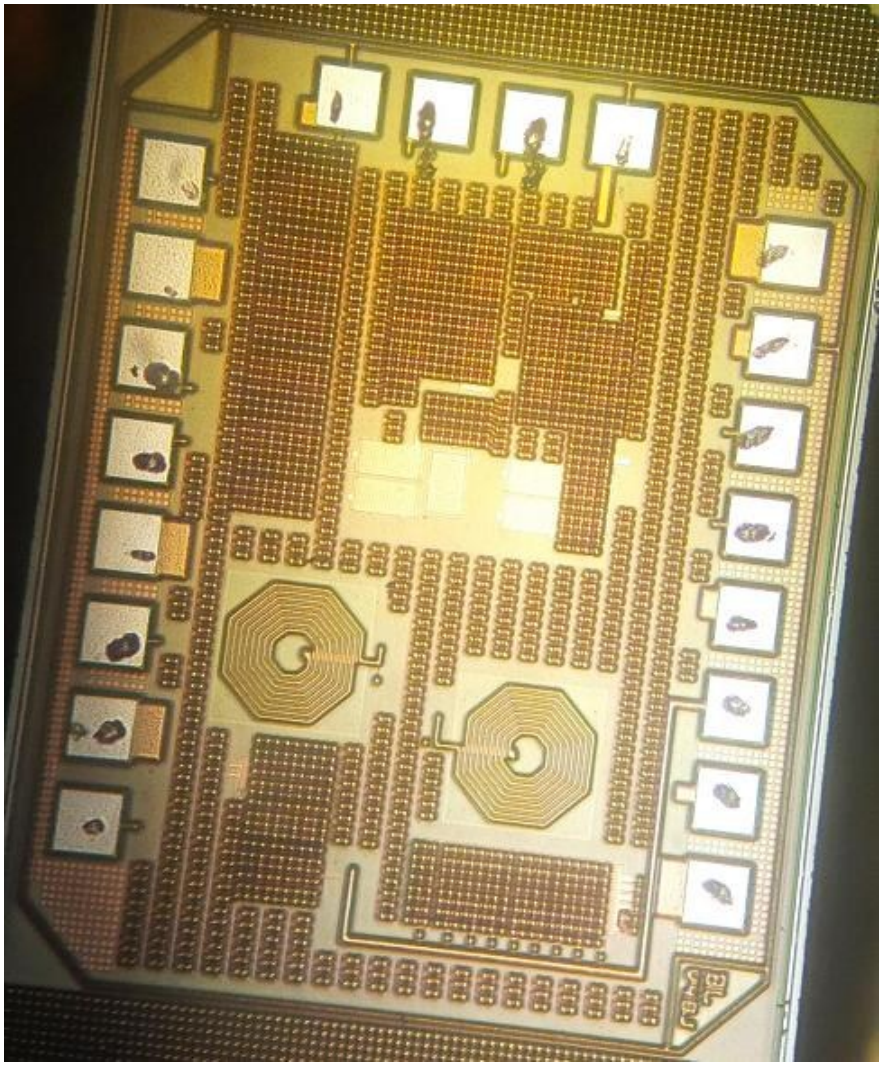

Figure 4.29 Photograph of the measured chip. Scratches from the test probes are visible on the connecting pads.

5 Lab Measurement Results

The fabricated silicon chips were measured in the lab in order to evaluate their performance. The block diagram of the test setup is shown in Figure 5.1 below.

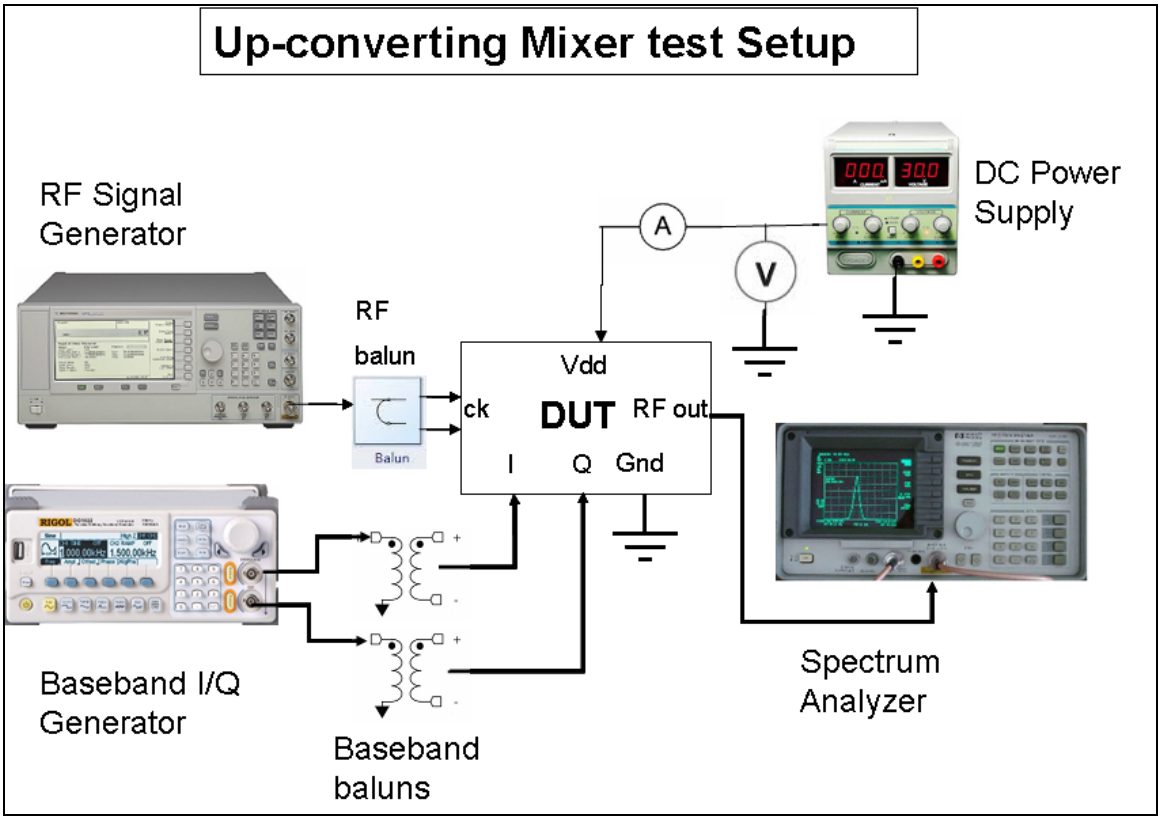

Figure 5.1 Up-Converting Mixer Test Setup

The following test equipment units were used in the test setup:

- GW Labs power supply, Model GPS-1830
- Agilent Signal Generator, Model E8241A PSG-L Series
- Rigol Dual Arbitrary Waveform Generator, Model DG1022A
- HP Spectrum Analyzer, Model 8595A
- Digital Multimeters: MY-68, Fluke

5.1 Power Consumption

The measured power consumption is shown in Table 4 below.

Table 4 Power consumption of the measured chip

Vdd [V]	1.8	2	2.2
Id [mA]	29.9	35	40
P [mW]	53.8	70	88

Correlation between measurements and simulations

Simulated current consumption was 32mA at nominal power supply voltage Vdd=2V. This correlates well with the measured 35mA nominal.

5.2 Clock Sensitivity vs. Input Frequency

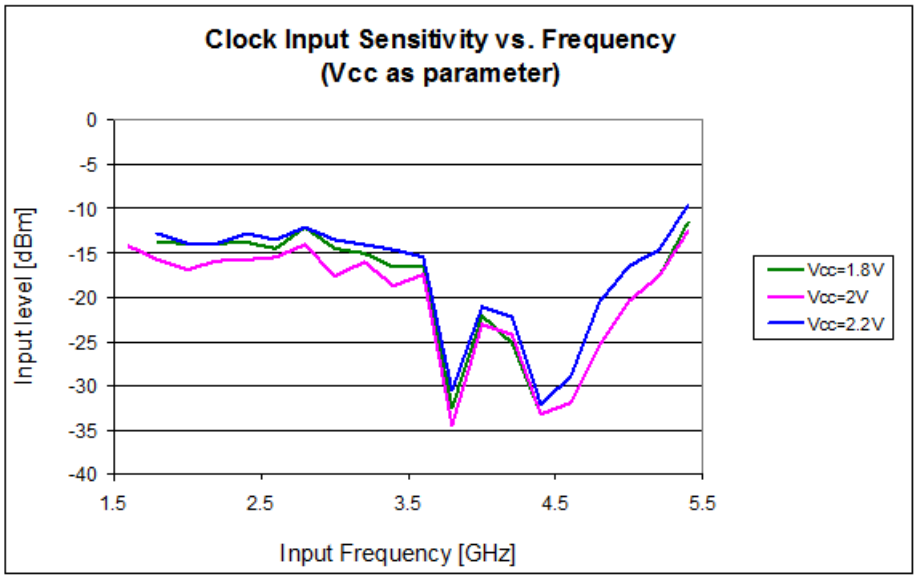

Figure 5.2 Clock Input Sensitivity vs. Frequency

Discussion

Attempts to increase the clock frequency beyond 5.4GHz were not successful. This may be the result of the limited frequency range of the balun available for the clock input. Balun characterization on the network analyzer revealed the S21 shown in Figure 5.3 below. Inherent mismatches may have further reduced the balun performance above 5GHz.

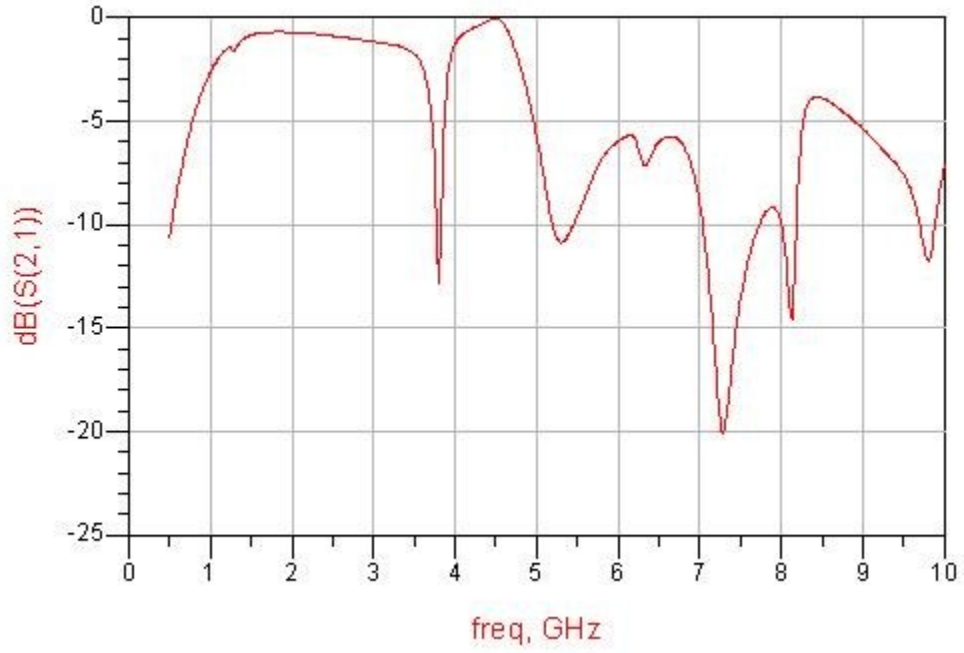

Figure 5.3 Transfer (S21) characteristic of the balun used for the clock input.

Correlation between measurements and simulations

Input sensitivity correlation shows opportunities for improvement. See the previous paragraph for discussion.

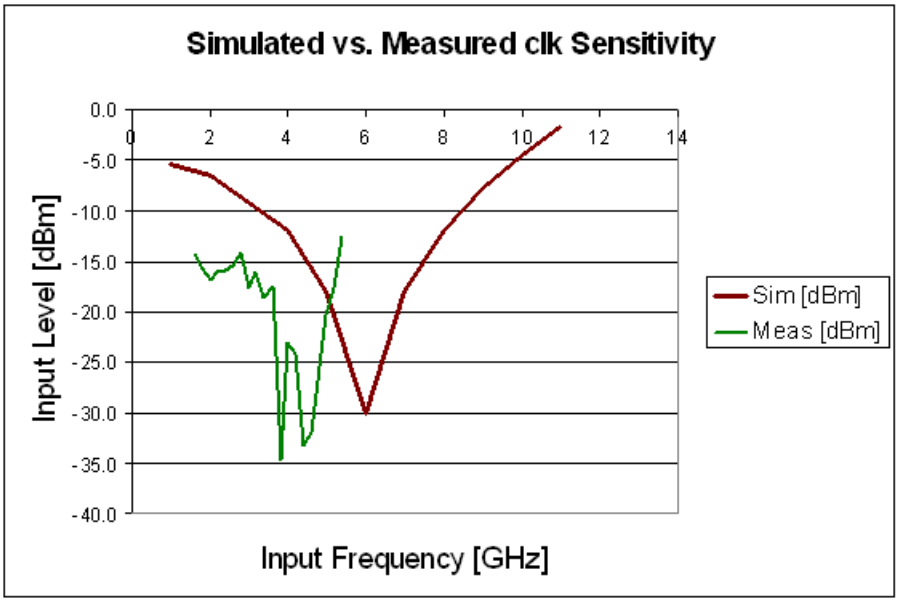

Figure 5.4 Simulated vs. Measured RF clock sensitivity

5.3 **Output 1dB compression point – P1dB**

1 dB compression point measurements and comparisons to calculations and simulations are shown in Table 5 and Table 6 and plotted in Figure 5.5.

Table 5 Output 1 dB Compression Measurements

BB Input	Measured Output Levels [dBm]				Calculated Output for Linear Gain [dBm]		Gain Compression [dB]	
Vbb [mVpp]	10M_1 G	10M_2.4 G	20M_1 G	20M_2.4 G	lin_10M_1 G	lin_20M_2.4 G	GC_10M_1 G	GC_20M_2.4 G
10.00	-48.90	-52.79	-49.03	-54.20	-49.08	-54.07	-0.18	-0.13
20.00	-43.30	-47.59	-42.83	-47.70	-43.07	-47.84	0.23	0.14
40.00	-37.10	-41.59	-36.93	-41.50	-37.06	-41.61	0.04	0.11
80.00	-31.00	-35.59	-30.73	-35.50	-31.05	-35.38	-0.05	-0.12
120.00	-27.50	-32.19	-27.23	-32.10	-27.54	-31.74	-0.04	-0.36
200.00	-22.90	-28.09	-23.03	-27.80	-23.11	-27.14	-0.21	**-0.66**
300.00	-20.30	-24.89	-19.93	-24.80	-19.60	-23.50	**0.70**	**-1.30**
400.00	-18.40	-23.09	-18.13	-22.90	-17.11	-20.91	1.29	-1.99
500.00	-17.40	-21.89	-17.03	-21.80	-15.17	-18.91	2.23	-2.89

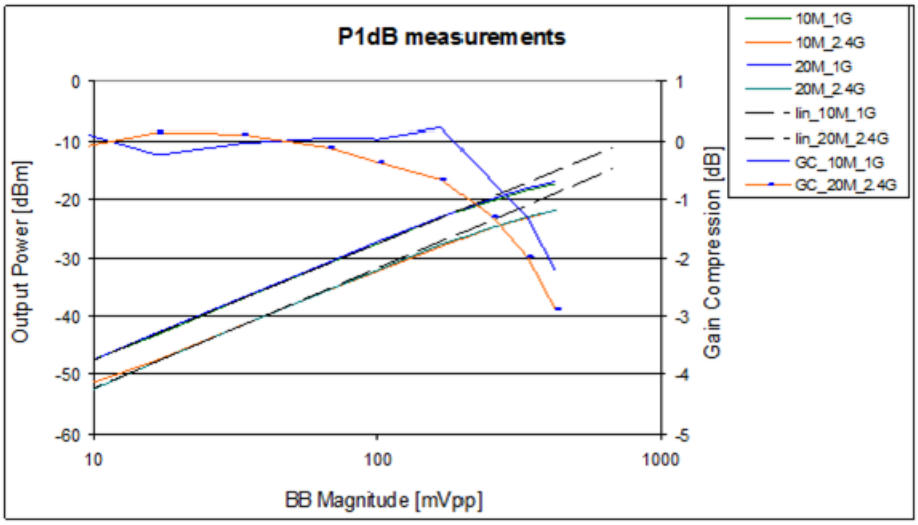

Figure 5.5 Output 1dB Compression Point

Correlation between measurements and simulations

Table 6 below shows a reasonably good correlation between calculation, simulation and lab measurement.

Table 6 Calculated, Simulated and Measured 1dB Compression point

	Calculated	Simulated	Measured
1dB Input Voltage Compression Point	217mVpp	178mVpp	250mVpp
Output Power 1dB Compression Point			-20dBm @ 1GHz -24 dBm @ 2.4GHz

5.4 **Output IP3**

IP3 measured at f_{out} = 1GHz is shown in Table 7 and plotted in Figure 5.6.

Table 7 Output 3rd order IMD measurements used for IP3 calculations at f_{out}=1GHz

Vin [mVpp]	P1_1G [dBm]	P3_1G [dBm]	OIP3_1G [dBm]
100	-27.22	-62.90	-9.38
110	-26.23	-59.42	-9.63
120	-26.40	-57.34	-10.92
130	-26.27	-54.95	-11.94

140	-25.14	-54.67	-10.38
150	-24.63	-52.41	-10.75
		AVG	-10.50

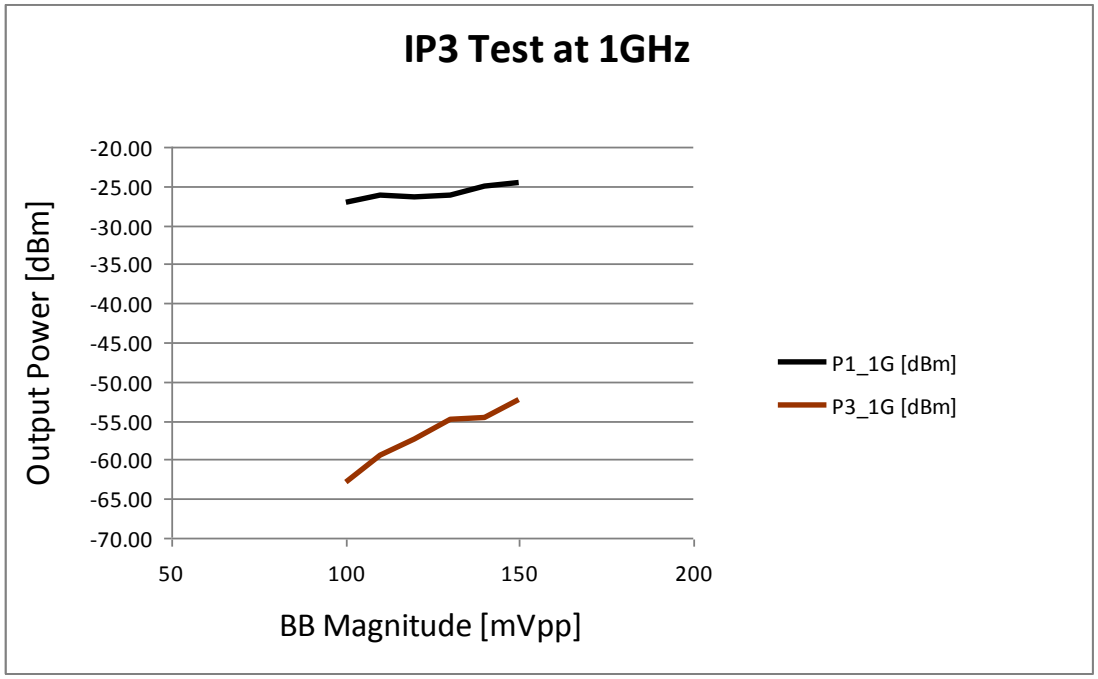

Figure 5.6 IP3 at f_out=1GHz. BB1=10MHz, BB2 = 15MHz, IMD = 20MHz

IP3 measured at f_{out} = 2.4GHz is shown in Table 8 and plotted in Figure 5.7.

Table 8 Output 3rd order IMD measurements used for IP3 calculations at f_{out}=2.4GHz

Vin [mVpp]	P1_2.4G [dBm]	P3_2.4G [dBm]	OIP3_2.4G [dBm]
100	-31.29	-68.84	-12.52
110	-27.74	-55.61	-13.80
120	-26.55	-52.92	-13.36
130	-25.80	-50.39	-13.51
140	-25.89	-49.75	-13.95
150	-25.46	-48.56	-13.92
		AVG	-13.51

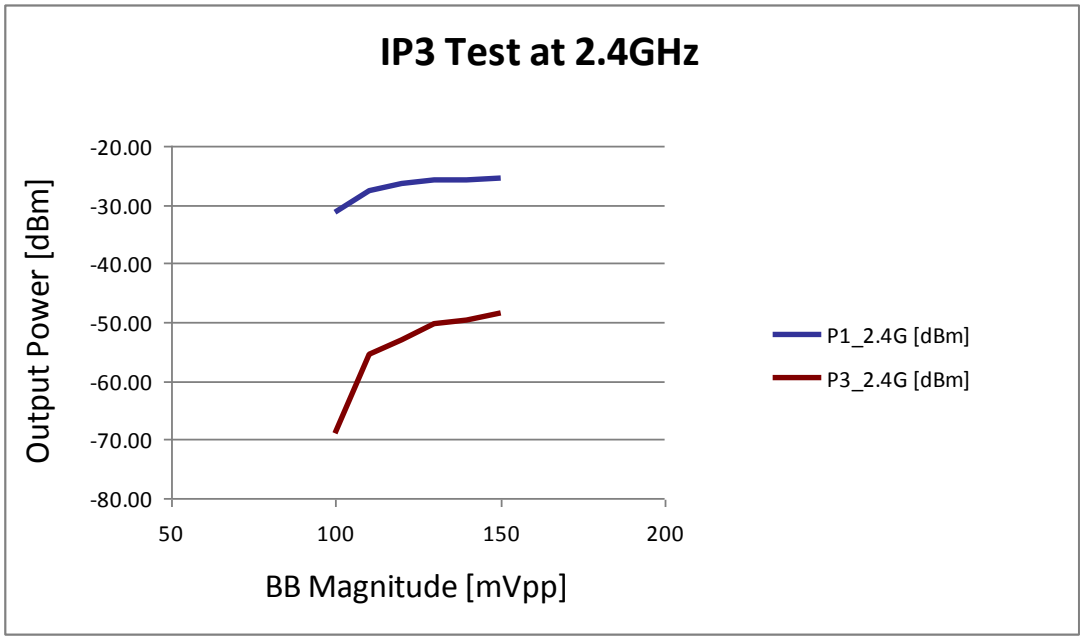

Figure 5.7 IP3 at f_out = 2.4GHz. BB1=10MHz, BB2 = 15MHz, IMD = 20MHz

Discussion

IP3 was calculated based on measurements and by applying (48). Due to the relatively low level of the IMD, multiple measurements were taken and averaged in order to get a figure close to the actual performance of the circuit.

The following notations were used:

- P1 – the measured output power on the fundamental component frequencies
- P3 – the measured intermodulation product
- OIP3 – the calculated third order intercept point at the output.

Correlation between measurements and simulations

IP3 results correlation is a bit more challenging to evaluate because the measurements were performed at the output, while the calculation and simulation were referred to the input. Also, the input impedance is higher than 50 Ohm, therefore the power gain is different from the voltage gain. It should be noted, however, that the output IP3 is approximately 10dB higher than the output 1dB compression point as predicted by theory. See Table 5 and Figure 5.5 for details.

5.5 **Sideband Suppression**

The sideband suppression measurements are shown in Figure 5.8 and Figure 5.9 below.

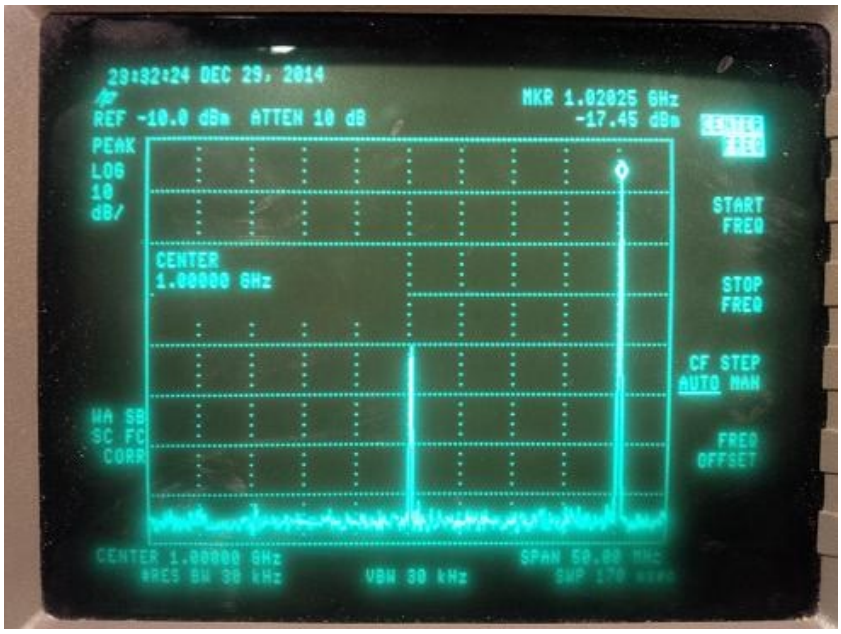

Figure 5.8 Sideband Suppression at fc = 1GHz, f_{BB} = 20MHz

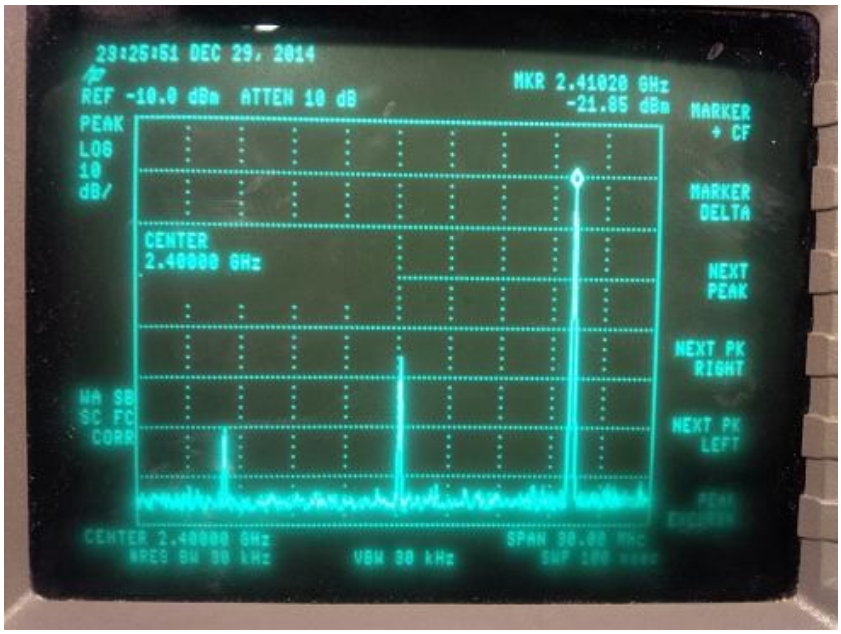

Figure 5.9 Sideband Suppression at fc = 2.4GHz, f_{BB} = 10MHz

Discussion:

In both 1GHz and 2.4GHz cases, the undesired sideband suppression is 50dB or better. The level of suppression shown was achieved after fine tuning the magnitude and phase of the baseband signal. In an actual transmitter, the baseband is generated by a DSP that has full control over the baseband magnitude and phase.

The carrier feed-through is fairly high compared to the undesired sideband level. The differential I and Q signals were generated using a simple transformer based balun. This approach did not allow provisions for baseband DC adjustment.

Correlation between measurements and simulations

Simulation results showed -43dB of unwanted sideband attenuation. The -50dB measured figure shows that the RF phase shifter is working properly and a good SBS can be achieved by properly adjusting the baseband I/Q signals.

One possible reason why the measured SBS result is better than the simulation might be the fine tuning of the magnitude and phase of the BB generator while seeking a minimum in the sideband magnitude. The procedure might have compensated the impairments of the phase shifter.

6 Conclusions

6.1 Achievements

The thesis work has resulted in the design, simulation and lab validation of a wideband up-converting mixer implemented on a 1.2x1.55mm silicon die that can be regarded as initial proof of concept for future developments and improvements. The whole design was divided in sub-blocks that were designed, analyzed and optimized separately. The following sub-blocks were implemented on chip:

A wideband digital phase shifter was implemented in CML configuration and optimized for wideband operation. The output of the digital phase shifter generates I and Q replicas of the LO that is further used in the complex mixing. A couple of design issues were encountered in the process and solutions were implemented. Due to the wide band and high operating frequency, the initial simulation of the extracted model had poor performance. It was determined that the resistance of the connecting traces was too high and it acted as a low pass filter that prevented the operation at high frequency. Trace widening resulted in successful operation. The CML latches could not drive capacitive loads higher than 16fF. Buffer stages were implemented in order to drive the Gilbert cells having an input capacitance around 50fF.

Two separate Gilbert cells were implemented for complex multiplication and I/Q modulation. The Gilbert cells were optimized for high frequency operation by transistor scaling and choice of operating current. Both input and output circuits were implemented in a differential configuration for improved LO feed-through rejection. Wideband operation was achieved by resistive output current summation.

An on chip active balun was implemented for conversion of the internal differential signal to the 50Ohm single ended output. Wideband operation was possible by appropriately scaling and biasing the balun circuit. The balun improves the LO rejection while converting the differential modulated signal to single ended. It also provides 50Ohm matching at the output of the mixer. The output is AC coupled; no external DC blocking capacitor is required

The three bias cells implemented in the design are identical; however, they allow for current adjustment in order to optimize each biased circuit. All three bias cells are provided with external connections for additional bias optimization for the main sub-blocks of the design.

In general, the lab measurements have shown reasonably good correlation with the calculations and simulations. Comparative tables and comments are provided in the Lab measurements section, chapter 5. For reference, the measurement results are provided below.

Nominal Supply Voltage [V]	2
Nominal Current Consumption [mA]	35
Nominal Power Consumption [mW]	70
Maximum Clock Sensitivity [dBm]	-30
Output P1dB @ 1GHz [dBm]	-19.9
Output P1dB @ 2.4GHz [dBm]	-24.8
Output IP3 @ 1GHz [dBm]	-10.5
Output IP3 @ 2.4GHz [dBm]	-13.51
Unwanted Sideband Suppression [dB]	>50

6.2 **Future Work**

Although the initial intended target for the maximum input clock frequency was 13GHz, the simulation results over PVT variations showed that this target can be reached only under best case scenarios. Further more, actual lab measurements resulted in a significant maximum frequency limitation compared to the intended target; this requires investigations, including the acquisition of test equipment with extended operating range.

Future work may focus on refining the wideband phase shifter such that the initial target of 13GHz clock input is achievable under worst PVT conditions. Main areas of focus are:

- Increase component density by reducing the spacing between them. Smaller distances results in lower interconnect parasitic capacitance and series resistance.
- A new attempt at transistor scaling in order to achieve the optimum current density for f_T optimization
- Attempt to use inductive peaking while watching closely the wideband performance of the system.

The chip area may be reduced by lowering the number of decoupling capacitors and grouping sub-blocks together that can use a common decoupling strategy. A thorough analysis on the supply line noise of each functional block is required prior to making a decision related to the minimum amount of decoupling capacitance. It is important to make sure the overall performance of the system is not compromised.

The output signal level is quite low; therefore an attempt to add a wideband amplifier at the output is another good candidate for further development.

7 References

[1] J.W.M Rogers and C. Plett, *Radio Frequency Integrated Circuit Design 2nd Ed.*, Artech House Publishers, 2010

[2] S. Voinigescu, *High-Frequency Integrated Circuits*", Cambridge University Press, 2013

[3] A. Georgiadis, "Gain, phase imbalance, and phase noise effects on error vector magnitude." *IEEE Transactions on Vehicular Technology*, pp. 443–449, 2004.

[4] B. Razavi, *RF Microelectronics*, Prentice Hall, 1998

[5] Agilent Application Note 1298, *Digital Modulation in Communications Systems — An Introduction*, 2001

[6] M. J. Gingell, "Single-sideband modulation using sequence asymmetric poly-phase networks", *Electric. Commun*. vol. 48, no. 1–2, pp. 21–25, 1973.

[7] R. C. V. Macario and I. D. Mejallie, "The phasing method for sideband selection in broadcast receivers", *EBU Rev. (Technical Part)*, no. 181, pp. 119–125, 1980.

[8] Mini-Circuits, *"*Understanding mixers*"*, in *RF/IF Designer's Handbook,* 1997

[9] Mini-Circuits, *"*Modern mixer terms defined*"*, in *RF/IF Designer's Handbook*, 1997

[10] B. H. Leung, .*VLSI for wireless communication*, Prentice Hall, 2002

[11] G. Gonzalez, .*Microelectronic Transistor Amplifiers, Analysis and Design, 2nd ed.*., Prentice Hall, 1997.

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

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

[14] J. Lee, *mm-Wave Silicon Technology: 60GHz and Beyond* chapter 5, Editors: A. M. Niknejad, H. Hashemi, Springer Science, 2008

[15] P. Heydari, "An Analysis of High Frequency Noise in RF Active CMOS Mixers", *Analog Integrated Circuits and Signal Processing*, vol.48, pp.199-209, 2006.

[16] H. Darabi and A. Abidi, "Noise in RF-CMOS Mixers: A Simple Physical Model", *IEEE Transactions on Solid State Circuits*, vol.35, no.1, pp.15-25, January 2000.

[17] M. Jouanneau-Douard, B. Herve, S. Bionaz, and D. Levy, "A 1 to 18 GHz out of phase combiner", *IEEE Microwave and Millimeter-Wave Monolithic Circuits Symposium digest*, 1992

[18] J.-F. Chen, C. Gelatos, P. Tobin, R. Shimer, and C. Hu, "Reverse Antenna Effect Due to Process-Induced Quasi-Breakdown of Gate Oxide", *IEEE International Integrated Reliability Workshop Final Report* , Lake Tahoe, CA, pp. 94-97, October 20-23, 1996.

[19] Texas Instruments, *"*Application Report SLYA014A*"*

[20] Fairchild Semiconductors, *"*Application Note AN-600*"*

[21] Microchip technology Inc., *"*Application Note AN763*"*

[22] P. Hurst, P. Gray, S. Lewis, *Analysis and Design of Analog Integrated Circuits*, John Wiley & Sons, 2009

[23] P. E. Allen, *Analog Integrated Circuits – Course Notes*

[24] A. J. Wilkinson, *Signals and Systems - Course Notes*
[25] J. P. Silver - http://www.odyseus.nildram.co.uk RF,RFIC & Microwave Theory, Design
[26] C. Plett, *Radio Frequency Integrated Circuit Design - Course Notes*