

High Performance CMOS Transmitters for Wireless Communications

by

Jeffrey Arthur Weldon

B.S. (University of California, Berkeley) 1992

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering-Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Paul R. Gray, Chair

Professor Robert G. Meyer

Professor Paul K. Wright

Fall 2005

The dissertation of Jeffrey Arthur Weldon is approved:

Chair

Date

Date

Date

University of California, Berkeley

Fall 2005

High Performance CMOS Transmitters for Wireless Communications

Copyright © 2005

by

Jeffrey Arthur Weldon

To Mom and Dad

Abstract

High Performance CMOS Transmitters for Wireless Communications

by

Jeffrey Arthur Weldon

Doctor of Philosophy in Engineering-Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Paul R. Gray, Chair

The demand for wireless technology has dramatically increased in recent years. Wireless communications has not only allowed for increased portability for voice and data communications but has also facilitated the deployment of systems in which wires are either too difficult or costly to install. Two fundamental forces have largely fueled this demand: the desire for information and the advancement of technology. Despite the improvement of the underlying integrated circuit technology, high performance transmitters typically use a number of discrete components and several integrated circuits because they employ multiple device technologies.

This thesis describes advancements at both the circuit and architectural levels which allow the construction of a single-chip CMOS transmitter while enabling high performance and the ability to operate with multiple radio frequency standards. Single-chip integration without the need for off-chip filtering has been addressed at the circuit level with the design of a mixer that eases the filtering requirements by canceling the closest harmonics created in the mixing process. The mixer uses multiple phases of the

local oscillator signal to effectively multiply the input signal by a sampled-and-held version of a sine wave, as opposed to the more typical square wave.

By leveraging the abilities of the mixer to ease filtering requirements, a transmitter architecture was designed that simultaneously allowed for single chip integration in CMOS and high performance. This architecture's ability to accurately modulate a baseband signal is evident by its high image rejection. The architecture is based on I/Q modulation and therefore retains the potential for multi-standard operation.

To evaluate the mixer and the transmitter, a prototype transmitter was fabricated in a 0.35- μm five-metal double-poly CMOS process. The test chip included the entire signal path which was comprised of the digital-to-analog converters, baseband filters, IF mixers, RF mixers, frequency synthesizers and the power amplifier. The transmitter achieved 1.3 degrees of RMS phase error and met the close-in spectral mask requirements of the DCS1800 cellular communications standard. Furthermore, the rejection of the third and fifth IF harmonics was measured at -68 dBc and -69 dBc, respectively and the untuned image rejection was measured at -56 dB.

Contents

Figures	v
Tables	viii
Acknowledgments	ix
CHAPTER 1 INTRODUCTION	1
1.1 Communications Technology	1
1.2 Background	3
1.3 Research Goals	6
1.4 Societal Impact	8
1.5 Thesis Organization	9
CHAPTER 2 TRANSMITTER FUNDAMENTALS	11
2.1 Introduction	11
2.2 Wireless Communications System	12
2.3 Modulation	14
2.3.1 Constant Envelope vs. Non-constant Envelope	17
2.3.2 GMSK Modulation	22
2.4 Performance Metrics	24
2.4.1 Modulation Accuracy	25
2.4.2 Spectral Emissions	26
2.5 Non-Idealities	29
2.5.1 Quadrature Phase Mismatch	29
2.5.2 Gain Mismatch	33
2.5.3 DC Offsets and LO Feedthrough	35
2.5.4 Noise	38
2.5.5 LO Phase Noise	41
2.5.6 Filters	43
2.5.7 Nonlinear Distortion	49
CHAPTER 3 TRANSMITTER ARCHITECTURES	53
3.1 Introduction	53
3.2 PLL Based Architectures	55
3.3 Direct Conversion Architecture	59
3.4 Heterodyne Architecture	63
3.5 Transmitter Architecture Comments	65
CHAPTER 4 HARMONIC-REJECTION MIXER	67
4.1 Introduction	67

4.2	Switching Mixers	70
4.3	Harmonic-Rejection Background	74
4.4	Harmonic-Rejection Principles	85
4.5	Harmonic-Rejection Mixer	92
4.5.1	Applicability to Switching Mixers	93
4.5.2	Current Commutating Mixer Gain	95
4.5.3	HRM Schematic and Design Issues	97
4.5.4	Gain of the HRM	106
4.6	Matching in HRM	108
4.7	Harmonic Rejection in the General Case	114
4.8	LO Phase Generation	119
4.8.1	Generation of Eight Phases using a Frequency Divider	120
4.8.2	Generation of Eight Phases Using a Passive RC Filter	123
4.9	Summary	130
CHAPTER 5 HARMONIC REJECTION TRANSMITTER		131
5.1	Introduction	131
5.2	Harmonic Rejection Transmitter Architecture	134
5.2.1	Integration Advantages of the HRT	136
5.2.2	Performance	138
5.2.3	Multi-Standard Capability	140
5.3	Image Rejection	141
5.3.1	Amplitude mismatch tuning method	154
5.4	Synthesizer Integration	157
5.5	Summary	160
CHAPTER 6 TRANSMITTER PROTOTYPE		161
6.1	Introduction	161
6.2	Image Rejection and Carrier Feedthrough	163
6.3	Mixers	166
6.3.1	HRM Circuit	167
6.3.2	SNR in HRMs	173
6.3.3	I and Q HRMs	179
6.3.4	RF Mixers	182
6.4	Additional Circuits	185
6.4.1	Digital to Analog Converter	186
6.4.2	Baseband Filter	188
6.4.3	Frequency Synthesizers	190
6.4.4	Eight Phase Generation	191
6.4.5	Power Amplifier	192
6.5	On-chip Signal Coupling	194
6.6	Transmitter Test Chip	196
6.7	Summary	198

CHAPTER 7	MEASUREMENT RESULTS	199
7.1	Introduction	199
7.2	Measurement results	202
7.3	Summary	211
CHAPTER 8	CONCLUSIONS	212
8.1	Research Summary	212
8.2	Future Work	214
APPENDIX A	IMAGE REJECTION DERIVATION OF HRT	216
	REFERENCES	225

Figures

Figure 2.1	Simplified block diagram of wireless communications system.	13
Figure 2.2	Quadrature modulator.	17
Figure 2.3	Third-order distortion of AM signal.	20
Figure 2.4	GMSK constellation diagram.	24
Figure 2.5	Graphical representation of the error vector and phase error.	26
Figure 2.6	GSM in-band spectral mask requirement.	27
Figure 2.7	GSM spectral mask at low offset frequencies with GMSK modulated signal.	28
Figure 2.8	Quadrature modulator with phase mismatch.	30
Figure 2.9	GMSK constellation with quadrature phase error.	31
Figure 2.10	Quadrature phase error constellation.	32
Figure 2.11	Single sideband spectrum with phase error.	33
Figure 2.12	Quadrature modulator with gain mismatch.	34
Figure 2.13	GMSK constellation with gain mismatch.	34
Figure 2.14	Single sideband spectrum with gain mismatch.	35
Figure 2.15	GMSK constellation with a DC offset.	36
Figure 2.16	Single sideband spectrum with DC offset and quadrature mismatch.	37
Figure 2.17	GMSK signal with thermal noise.	39
Figure 2.18	GMSK constellation diagram with thermal noise.	40
Figure 2.19	LO spectra. (a) Ideal spectrum. (b) Spectrum with phase noise.	41
Figure 2.20	GMSK constellation diagram with phase noise.	42
Figure 2.21	GMSK spectrum with phase noise.	43
Figure 2.22	Typical quadrature modulator showing baseband circuits.	44
Figure 2.23	Magnitude responses of 3 rd order Butterworth, Bessel, and Chebyshev filters.	45
Figure 2.24	Passband magnitude responses of 3 rd order Butterworth, Bessel, and Chebyshev filters.	46
Figure 2.25	Group delay of 3 rd order Butterworth, Bessel, and Chebyshev filters.	47
Figure 2.26	GMSK constellation with 3 rd order Butterworth filter.	48
Figure 2.27	Harmonic distortion.	49
Figure 2.28	GMSK spectrum with third order distortion.	51
Figure 2.29	Intermodulation distortion.	52
Figure 3.1	Generic transmitter architecture.	54
Figure 3.2	Offset phase locked loop architecture.	56
Figure 3.3	Basic polar transmitter block diagram.	57
Figure 3.4	Direct-conversion transmitter block diagram.	59
Figure 3.5	Third order intermodulation by a non-linear PA in a direct-conversion transmitter.	61
Figure 3.6	LO Pulling in a direct-conversion transmitter.	62
Figure 3.7	Conventional two-step or heterodyne transmitter block diagram.	64

Figure 4.1	Graphical representation of convolution of two cosine functions in frequency domain.	69
Figure 4.2	Single-balanced current-commutating mixer.	71
Figure 4.3	Unity amplitude square wave.	71
Figure 4.4	Output spectrum of a single-balanced mixer with sinusoidal baseband input.	72
Figure 4.5	Double-balanced current-commutating mixer.	73
Figure 4.6	Square wave generated by applying a sample-and-hold operation to a sine wave.	75
Figure 4.7	Model for sample-and-hold operation.	76
Figure 4.9	Graphical representation of the hold function on sampled sinusoid.	80
Figure 4.10	Sample and hold operation on sinusoid in time and frequency domains for three sampling frequencies. (a) $f_s=2f_i$. (b) $f_s=4f_i$. (c) $f_s=8f_i$.	83
Figure 4.11	Spectrum from a sample-and-hold for with arbitrary sampling rate N.	84
Figure 4.12	(a) Input, LO and output signals of a conventional switching mixer. (b) Input, LO and output signals of a harmonic-rejection mixer.	86
Figure 4.13	Four square waves which can be summed to generate the SHS waveform.	88
Figure 4.14	SHS waveform resulting from shifting of sampling position and the square waves which compose the SHS waveform.	89
Figure 4.15	Simplified current-commutating switching mixer.	96
Figure 4.16	Simplified circuit diagram of the harmonic-rejection mixer.	99
Figure 4.17	Graphical representation of HR_3 and HR_5 .	109
Figure 4.18	Third harmonic rejection as a function of gain and phase error.	112
Figure 4.19	Fifth harmonic rejection as a function of gain and phase error.	113
Figure 4.20	Square wave and a more realistic current waveform generated by a sinusoidal LO signal.	115
Figure 4.21	Divide-by-four frequency divider.	120
Figure 4.22	Expanded view of divide-by-four and timing diagram.	121
Figure 4.23	(a) RC low-pass filter. (b) CR high-pass filter.	123
Figure 4.24	Polyphase filter used for quadrature generation.	124
Figure 4.25	RCR circuits giving 22.5 degree phase shift.	125
Figure 4.26	Eight-phase polyphase filter.	128
Figure 4.27	Three stage eight-phase polyphase filter.	129
Figure 5.1	Block diagram of a Harmonic Rejection Transmitter.	135
Figure 5.2	Gain and phase mismatch in a quadrature modulator.	142
Figure 5.3	Model used to evaluate image rejection in the HRT.	144
Figure 5.4	Image rejection for HRT and quadrature modulator as function of θ_2 for two values of α and with $\theta_1=2^\circ$.	146
Figure 5.5	Image rejection for HRT as function of θ_1 and θ_2 for two different values of α .	147
Figure 5.6	Constellation diagrams illustrating the effect of LO_2 phase error at various points in the HRT. (a) I_{IF} . (b) Q_{IF} . (c) I_{RF} . (d) Q_{RF} . (e) I_{RF} .	149
Figure 5.7	Constellation diagram of the HRT output in one quadrant with LO_1 and LO_2 phase errors.	150

Figure 5.8	Constellation diagrams illustrating the effect of gain mismatch at various points in the HRT. (a) I_{IF} . (b) Q_{IF} . (c) I_{RF} . (d) Q_{RF} . (e) I_{RF} .	152
Figure 5.9	Phasor diagram of HRT showing the effects of gain mismatch and LO_2 phase error.	153
Figure 5.10	Feedback options for quadrature modulator.	155
Figure 5.11	Amplifier converted from Gilbert cell mixer.	156
Figure 5.12	Low frequency feedback with amplifier in place of mixers.	157
Figure 6.1	HRT prototype block diagram.	163
Figure 6.2	Simulated RMS phase error as a function of image rejection from only gain mismatch in a GMSK signal.	165
Figure 6.3	Simulated RMS phase error as a function of carrier feedthrough for GMSK signal.	166
Figure 6.4	HRM circuit diagram.	168
Figure 6.5	Simulated RMS phase error as a function of $ HD_3 $ for a GMSK signal.	169
Figure 6.6	GMSK spectrum with HD_3 level at -40 dB.	170
Figure 6.7	Mixers used for comparison. (a) Single mixer with ideal switches. (b) Parallel mixers with ideal switches.	174
Figure 6.8	Mixers used for comparison. (a) Single mixer. (b) HRM.	176
Figure 6.9	Output of divide-by-four showing multiple phases of the LO signal.	180
Figure 6.10	HRM layout.	182
Figure 6.11	Inductively loaded Gilbert cell mixer used for up-conversion to RF.	183
Figure 6.12	Layout of RF mixers.	185
Figure 6.13	Dual resistor string DAC.	186
Figure 6.14	Complementary switches in dual resistor string DAC.	187
Figure 6.15	Plot of simulated RMS phase error of a GMSK modulated signal as a function of -3 dB frequency for four Butterworth filters of order 2, 3, 4 and 5.	189
Figure 6.16	Sallen and Key low-pass filter with feedback buffer.	190
Figure 6.17	Schematic of class-C PA.	194
Figure 6.18	Transmitter test chip micrograph.	197
Figure 7.1	Transmitter test setup.	200
Figure 7.2	Photo of test board with chip-on-board technology.	201
Figure 7.3	Single sideband output at testing buffer output.	203
Figure 7.4	Vector signal analyzer output.	205
Figure 7.5	Single sideband spectrum at PA output.	206
Figure 7.6	Modulated output spectrum with DCS1800 spectral mask.	207
Figure 7.7	Wideband plot showing the rejection of 3 rd and 5 th IF harmonics.	208
Figure 7.8	Plot showing wideband noise at 20 MHz offset.	209
Figure 7.9	Breakdown of current consumption in HRT.	211
Figure A.1	Model used to evaluate image rejection in the HRT.	217

Tables

Table 7.1 HRT performance summary.

210

Acknowledgments

The path to a Ph.D. is often a circuitous one. Fortunately, this is not a solitary endeavor as support and guidance come from many sources. In my experience the assistance has been technical, emotional, and medical and the people involved have been family, professors, fellow students, friends and physicians. To these people I am grateful.

I had the good fortune to work with Professor Paul Gray. His technical guidance and support throughout the years have been invaluable. Not only has he taught me a great deal technically but I have found his other lessons equally important. His commitment to excellence and integrity have shaped my perspective and I would like to express my most sincere gratitude for his support throughout the years.

I would also like to acknowledge the other professors that influenced my graduate career. I'd like to thank Professors Robert Meyer, for his technical assistance and for serving on my dissertation committee, and Professor Paul Wright for serving on my dissertation committee. I'm also grateful to Professors Bob Broderson and Jan Rabaey for providing an exceptional research environment at the BWRC. Professor Borivoje Nikolic has been generous with his advice on technical matters and career paths for which I am appreciative and I'd also like thank Professor Kris Pister.

One of most beneficial aspects of working for Professor Gray was his ability to attract incredibly talented students. I was lucky to be part of this and I benefited greatly from their knowledge and friendship. Due to the collaborative aspects of the research project I was involved in, I have many fellow students to thank. I ended up spending a great deal of time with Sekhar Narayanaswami and his insights and motivation for circuit

design and testing were invaluable. I am thankful to Sekhar for being a great teammate and even better friend.

There were several other students who were involved with the transceiver project that I would like to acknowledge. I want to thank Chris Rudell for his help with so many different matters. Li Lin, Martin Tsai, Luns Tee and Cheol-Woong Lee were integral to the success of the transmitter project and I am thankful for their assistance. I'd also like to thank Sebastian Dedieu and Masanori Otsuka, two visiting industrial fellows, that were also deeply involved with the transmitter project.

In addition to the students in my immediate research project, there are a number of other students in the research group that I would like to thank. These include Andy Abo, Carol Barrett, George Chien, Yun Chiu, Thomas Cho, Arnold Feldman, Keith Onodera, Jeff Ou and Todd Weigandt.

Early in my graduate school career there were a few more senior students to whom I feel indebted. Greg Uehara has been so generous with his time and energy. He has been a role model in many ways and I want to thank him for everything he has done. Tony Stratakos and Dave Lidsky have also played an important role early on and I am very appreciative.

I have also benefited from a number of students from other research groups. I'd like to thank Dennis Yee for his technical insights as well as the many intriguing conversations over the years. Ada Poon and I shared a cubicle and I ended up learning so much from her. I'd also like to thank Tom Burd and Henry Jen for their help. I'd like to acknowledge a number of other students including Sayf Alalusi, Chinh Doan, Dejan

Markovic, Al Molnar, Ian O'Donnell, Kostas Sarrigeorgidis, David Sobel and Johan Vanderhaeen.

Many of the staff members have also been very helpful to me throughout my time in graduate school. In particular I would like to thank Tom Boot, Ruth Gjerde and Carol Sitea for their assistance with numerous issues.

I'd also like to acknowledge a few physicians and they include Dr. Joel Piser, Dr. Eric Small, Dr. Mack Roach and Dr. Craig Nichols.

I have been truly fortunate to have such a loving and supportive family. To Ed and Laura, thank you for all of your encouragement and understanding throughout my life. You are more than just siblings, you are also two of my best friends.

To Sanie, thank you for putting up with me even in the most trying of times. You definitely made me a better and stronger person and I am so lucky to have found you. You and Connor have brought be so much joy.

To Mom and Dad, thank you for your patience and unwavering support. I feel so lucky to have you as my parents. Your unconditional love and willingness to do anything for your children is beyond remarkable. I am so thankful to both of you.

Chapter 1

Introduction

1.1 Communications Technology

Technology has dramatically improved the way in which human beings communicate and transfer information. Although a number of different technical developments have contributed to this improvement, it can be argued that the technology of wireless communications has had a greater impact on modern communications than any other single technology. Beginning in the early 20th century with broadcast radio and progressing to wireless Internet access and cellular telephony of today, communicating wirelessly has become indispensable in our society. Although the goal of transferring information without the use of wires hasn't changed, new methods have been developed which have greatly reduced the cost and size of the equipment while increasing the range

and the number of users in the system. One specific advancement that has had a dramatic effect on wireless communications is the development of integrated circuit technology.

The invention of the integrated circuit in the late 1950s by Jack Kilby and Robert Noyce would eventually change the way most electronic equipment is made. The first commercial integrated circuits (ICs) were available in 1961 and the rapid miniaturization that followed would radically alter the electronics industry. The continual shrinking of transistor sizes and the concomitant increase in the number of transistors in a given area has been termed Moore's law. (In 1965 Gordon Moore predicted that the number of transistors in a given area would double every twelve months.) Although the "law" was amended in the 1970s to change the period from twelve to eighteen months, it has been remarkably accurate for over three decades. The resulting exponential growth in the density of transistors has fueled the growth of the semiconductor industry by reducing the size and cost, and simultaneously increasing the functionality, of electronic devices.

Wireless devices have clearly benefited from the utilization of integrated circuits in many ways, most notably their size and thus portability. For example, in the early 1980s, a cellular phone had the volume of (and nearly the weight of) a common brick; modern cellular phones are smaller and lighter by at least an order of magnitude. Although the size reduction and the superior performance of modern wireless devices were facilitated by the underlying semi-conductor technology, advancements at both the circuit level and the system level have realized the potential provided by the technology. The research in this thesis focuses on integrated circuits designed for wireless communications.

1.2 Background

The demand for wireless technology has dramatically increased in recent years. Wireless communications have not only allowed for increased portability, as is the case with cellular telephones, but has made feasible the deployment of systems in which wires are either too difficult or costly to install, such as trans-oceanic communications. Two fundamental forces have largely fueled this demand: the desire for information and the advancement of technology. Although the need for information is not new, the Internet has significantly changed the dissemination of information by allowing for constant connectivity that was previously unavailable. This penchant for continuous connectivity, which has driven people to demand further improvements in their access to information, and it is wireless communications that provides this improvement.

Although wireless communications have become common, the advancement of technology has been critical to this change. Improved electronics have allowed for very high quality, technically advanced communications at a relatively low cost to the end user. For example, not long ago cellular phones were bulky, heavy, expensive and only capable of voice information. Today, phones that fit in the palm of the hand are capable of sending pictures and browsing the internet. Furthermore, these phones have become so inexpensive that almost everyone owns one. Integrated circuit technology has played a key part in this advancement, as have improvements at both the circuit and the system level. The ability of the technology to deliver high performance, low-cost electronics combined with the ever increasing demand for constant connectivity has driven existing

wireless communications and have also motivated new wireless standards and new wireless applications.

Fundamental to wireless devices is the ability to transmit and receive signals without the use of wires. The hardware ultimately responsible for this function, the transceiver, is present in nearly every wireless device with the exception being devices that only receive, such as some GPS units, or devices that only transmit. A transceiver is comprised of two basic blocks: the transmitter and the receiver. The transmitter converts low frequency or baseband information to a high frequency or radio frequency (RF) signal. This high frequency signal is then radiated via the antenna. The receiver performs the complementary function converting the RF signal to a baseband signal. Transceiver design is critical to the wireless devices because it affects many aspects of the device including cost, performance, size, and power consumption.

A wide variety of signal processing is needed in a typical transceiver design. The necessary signal processing typically includes baseband digital, baseband analog, intermediate frequency (IF) analog, and RF analog. Commonly each of these types of processing will require an individual IC. In addition to these processing elements, the interface between the baseband analog and digital sections will typically employ a dedicated IC and a special high power analog RF chip is needed to drive the antenna. Consequently, the transceiver usually consists of many separate integrated circuits as well as large numbers of passive components. The different sections use distinct ICs because different integrated circuit technology may be best for each section. For example, the technology used for the digital portion is typically silicon CMOS but much of the analog and RF signal processing is performed by other, more expensive, technologies such as

silicon bipolar, gallium arsenide (GaAs) or silicon germanium (SiGe). While this multi-chip solution may be advantageous for the performance of the transceiver, it also increases both the cost and size.

It would obviously be economically attractive to integrate these multiple ICs and passive components into a single IC based on an inexpensive technology: the reduction in size and cost would be substantial. Because of its low cost and compatibility with digital circuits, the most attractive technology for single chip integration is CMOS.

Although many different technologies are used for integrated circuits, CMOS remains the least expensive and the most widely used. The vast majority of circuit implementations, including digital and low frequency analog ICs, are well suited for CMOS. Driven by the demand for digital ICs, the investment in infrastructure for their production has resulted in a lower cost as compared with other technologies. Furthermore, with shrinking feature sizes leading to higher integration, both the density and functionality have increased.

While CMOS may be best for digital circuits, it traditionally has not been the technology of choice for high frequency analog and RF integrated circuits [1]. CMOS is ideal for digital ICs because CMOS transistors behave very much like ideal switches. However, the transistors are not well suited for analog circuits because of their relatively low current driving capability. Consequently, for high performance analog and RF circuits, other more expensive technologies, such as silicon bipolar, GaAs and SiGe, have been utilized.

The challenge of building a fully integrated transceiver lies in the ability to implement these functions in CMOS and simultaneously remove the need for additional

passive components. To build the analog and RF sections in CMOS presents some significant engineering challenges that require research into new techniques at both the circuit and architectural level.

In addition to single chip integration, the ideal transceiver would also operate with multiple radio standards. Currently many different wireless standards exist in the United States and even more worldwide. For example, cellular telephony throughout the country uses a number of standards including GSM, PCS-1900, IS-95, and AMPS. In addition, multiple standards have also emerged for wireless data such as 802.11a, 802.11b, 802.11g, and Bluetooth. Multiple standards have evolved for many reasons and consequently offer a variety of options for the end user. However, the multiple standards can also limit the effectiveness of a particular wireless device. For example, if a person changes their cellular service, they will in all likelihood need a new phone because different carriers use different standards. Primarily to minimize cost, wireless devices are typically built to operate in for one particular standard and do not inter-operate with other standards. Each wireless device has hardware that is dedicated to operate in the particular standard of interest. From the user's perspective, the ideal wireless device would be compatible with numerous standards.

1.3 Research Goals

The ultimate goal of this research is to develop new techniques to facilitate the integration of high performance RF transceivers in CMOS. The specific topic addressed here is the design and implementation of a single-chip transmitter. Advances at both the

architectural and circuit level of transmitters will be necessary to realize the ultimate goal of a complete single chip, multi-standard transceiver. To realize the goal of a single-chip CMOS transmitter, three research topics were investigated.

First, switching mixers typically generate significant harmonics at their output. These harmonics then need to be filtered before the signal is transmitted. A mixer was designed with inherent rejection of the local oscillator harmonics. This mixer, termed the Harmonic-Rejection Mixer (HRM) was analyzed and the performance limits will be discussed. Furthermore, circuit design techniques were developed to enable the incorporation of this design in a single-chip CMOS transmitter.

Second, changes to current transmitter architectures were necessary to facilitate complete integration. To understand the changes that are necessary transmitter architectures were analyzed for their potential for single-chip integration in CMOS, high performance and multi-standard operation. This resulted in a dual-conversion architecture that needed minimal filtering while maintaining high performance. This architecture, the Harmonic Rejection Transmitter (HRT) was analyzed and shown to have other significant benefits with respect to performance.

Third, a single-chip CMOS transmitter IC was designed and implemented. Experimental results show that the combination of architectural changes and circuit advancements have resulted in a high performance, single-chip transmitter that has the potential for multi-standard operation and that was realized in CMOS technology. Measured results show the effectiveness of the HRT architecture and the HRM circuits. The prototype was designed to meet the requirements of DCS1800 [2], a cellular standard which is an up-banded version of GSM.

1.4 Societal Impact

Over the past 25 years the cellular telephone has evolved from a large and cumbersome device costing several thousand dollars and with limited capabilities to the multi-capable small phones of today costing under one hundred dollars. The research presented in this thesis is one important step towards furthering this progression by facilitating the realization of a high-performance single-chip CMOS transceiver. Should this progression continue, the phones of the year 2030 will be the size of stick of gum, costing only few dollars with even more capabilities. The future applications of such technology appear to be unlimited and are difficult to predict, but speculating about the more likely candidates is interesting.

One of the future uses of such a phone might be the advent of the disposable cell phone much like the disposable cameras of today. Instead of buying a phone calling card, one might buy a disposable phone. This would be particularly useful when traveling to other countries that often use different radio standards.

Another potential use could be in medical applications. Every senior citizen might carry a phone that would automatically dial emergency assistance should some debilitating medical condition occur.

Other applications might involve tracking of important objects. For example, if every wild animal were fitted with a low-cost, high performance transceiver, a great deal could be learned about their behavior and patterns. Every piece of luggage or mailed package could be include a transceiver eliminating lost items.

Although the research presented here was demonstrated for cellular telephony, which is a modern high-performance wireless communications system, as wireless technology evolves and spectra become even more valuable, the performance demands of all future wireless transceivers is likely to increase. In this case the ideas discussed in this thesis might be applicable to many types of future wireless communications systems and the societal impact of this research could extend well beyond the cellular telephony of today.

1.5 Thesis Organization

This thesis consists in eight chapters. Chapter 2 will discuss the basics of transmitter operation and their fundamental function. The role of transmitter in a complete radio will be presented as well as metrics with which to judge performance. In addition, the effect of non-idealities on the transmitted signal will be discussed. Chapter 3 will focus on transmitter architectures and current implementations. The choice of architecture is critical to performance, facilitating integration as well as multi-standard operation. This chapter will discuss various architectures including direct conversion, heterodyne, and PLL based transmitters. Chapter 4 will describe the circuit design of the Harmonic-Rejection Mixer (HRM). The mixer will be analyzed for non-idealities and how they will affect the overall transmitter performance. Chapter 5 will focus on a transmitter architecture termed the Harmonic-Rejection Transmitter (HRT). This architecture will be analyzed for non-idealities and the results will be compared with other architectures. Furthermore, the performance limits of this architecture will be discussed. Chapter 6

will describe a prototype IC that was built to demonstrate the HRT and Chapter 7 will present the measurement results from this transmitter. Finally, Chapter 8 will draw some conclusions about the research presented and suggest possible future research.

Chapter 2

Transmitter Fundamentals

2.1 Introduction

As was stated in the previous chapter, almost all wireless devices use a transceiver to receive and transmit information. To better understand the technical challenges that accompany the goal a single-chip, multiple standard, CMOS transmitter, some background information regarding transmitters is helpful. This chapter will discuss the role of the transmitters in wireless systems and the important aspects of transmitter design including performance metrics, transmitter non-idealities and the effect of different radio standards on transmitter design. The radio standard that is implemented can have a significant effect on the transmitter design due to the differences between

constant envelope (CE) and non-constant envelope (NCE) modulation. This is further complicated when the goal is a multiple standard capable transmitter in which varying channel bandwidths are used.

In addition to the effect of modulation, the transmitter performance requirements are discussed. Ultimately the quality of the transmitted signal is the most important requirement and it contains two fundamental metrics: modulation accuracy and spectral emissions. Modulation accuracy refers to how closely the in-band transmitted signal replicates the ideally transmitted signal. Spectral emissions requirements limit the levels of the transmitted signal both in an out of band. Non-idealities in the transmitter circuits cause the transmitted signal and spectrum to deviate from the ideal.

Although many non-idealities exist within a given transmitter, a few are critical to integrated transmitter design. These include thermal noise, phase noise, non-linearities and mismatch. The effects of each non-ideality on the transmitted signal will be discussed. Understanding these non-idealities is important when evaluating transmitter architectures for their performance, multi-standard capability and single-chip integration potential. The following chapter focuses on these issues in detail.

2.2 Wireless Communications System

The transmitter is one key building block of a wireless communication system. Shown in Figure 2.1 is a simplified block diagram of a complete modern digital wireless communications system. The input is typically voice or data depending on the application. The entire transmitter is composed of two major sections: the digital signal

processor (DSP) and the section which performs the analog and RF signal processing. The latter section makes up the analog portion of the transmitter and is the focus of this research.

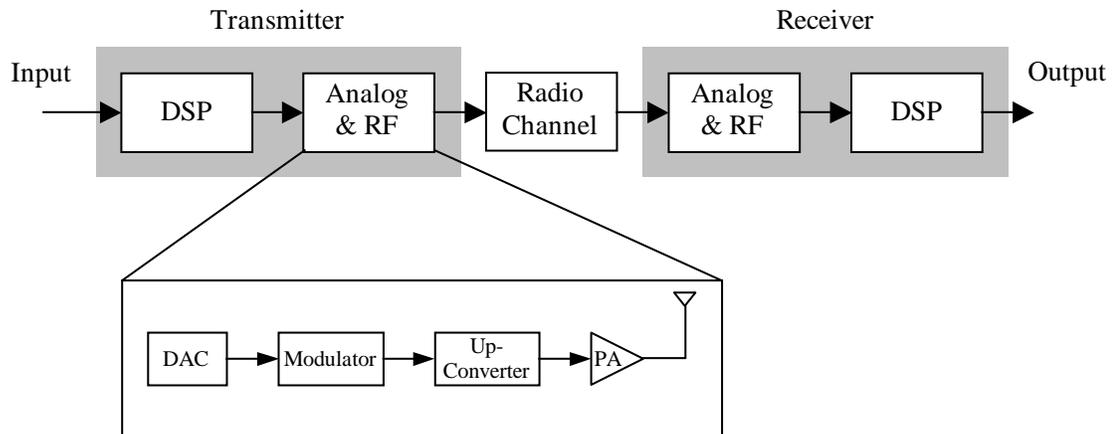


Figure 2.1 Simplified block diagram of wireless communications system.

The analog and RF section is expanded in Figure 2.1 to diagram the blocks within the analog and RF section. Although the particular breakdown of the signal processing varies between implementations this example illustrates a typical example. The input to the analog section comes from the DSP and this signal is then converted to an analog signal by the digital-to-analog converter (DAC). This baseband analog signal is then attached to a higher frequency signal. This process, termed modulation will be discussed in more detail in the following section. The signal is then up-converted to the RF spectrum. Modulation and up-conversion often occur in the same circuit but for the purposes of clarity the function are divided in this example. Finally, the signal is amplified and driven onto the antenna. In summary, the analog and RF section converts a baseband digital signal to a modulated, high-power, RF signal. For the purpose of

simplicity, in the remaining portion of this document the term transmitter will refer to the analog and RF section of the complete radio transmitter.

The focus of this research is to investigate architectures and circuit implementations that allow this analog and RF section to be integrated onto a single CMOS IC, potentially with the DSP, while implementing more than one radio standard. Although the power amplifier (PA) and the DAC will be referenced and discussed, their circuit implementation is not the focus. Instead, the circuit design will concentrate on the modulation and up-conversion process. As will be shown in Chapter 3, reducing the need for filtering in these sections will lead to increased integration. To better understand the issues involved with this goal, the process of modulation will be the focus of the next section.

2.3 Modulation

With the increased demand for wireless communications, radio spectrum has become extremely valuable. To get the most efficient use from a given section of spectrum, data is transmitted in such a way that it occupies as little spectrum as possible while still maintaining the desired quality. However, the complexity of the implementation is another important factor. Although high spectral efficiency is desired to maximize the use of the spectrum, the cost of implementation might offset these savings. Therefore, a trade-off typically exists between efficiency and complexity.

Wireless systems typically transmit and receive at frequencies that are considerably higher than the baseband signal. Modulation is the process in which a low-

frequency baseband signal is converted to a high-frequency bandpass signal. Modulation is performed in the transmitter and the receiver is responsible for demodulating the high frequency or radio frequency (RF) signal.

A modulated signal may be represented by

$$x(t) = a(t) \cos[2\pi f_c t + \phi(t)]. \quad (2.1)$$

The signal modulated signal $x(t)$ is essentially a sinusoid centered at f_c in which both the amplitude and phase may contain the desired information. If $a(t)$ is fixed and $\phi(t)$ is time varying, the signal is said to be angle modulated and if the reverse is true, the signal is amplitude modulated.

Although the modulating signal may be either digital or analog, digital modulation has become the predominant choice in applications such as cellular and cordless telephony, wireless LANs, and paging. The advantages of digital modulation include better noise immunity and greater resistance to channel variations. In addition, digital modulation is more spectrally efficient in a multi-user environment, such as cellular telephony, and it is easier to incorporate multiple data types like voice, video, or paging. However, these advantages come at the cost of hardware complexity.

Although digital modulation techniques often require the use of complex signal processing, the advancement of very large scale integration (VLSI) technology, and consequently the advancement of digital signal processing (DSP) capability, has mitigated the complexity disadvantage and has thus made digital modulation superior for many communications systems. While advances in the DSP allow for more complex

modulation schemes, implementation of the modulation function is still dependant on analog and RF signal processing.

To implement the modulation function of Equation (2.1) an alternative form is often useful. Expanding Equation (2.1) results in the following:

$$x(t) = a(t) \cos[\phi(t)] \cos[2\pi f_c t] - a(t) \sin[\phi(t)] \sin[2\pi f_c t]. \quad (2.2)$$

Although Equation (2.1) and Equation (2.2) are equivalent, implementation by direct interpretation of each equation would lead to very different results. Compared with Equation (2.1), the form shown in Equation (2.2) allows for a more separation between the RF signal processing and the baseband signal processing. Consequently, for most modulation schemes it is easier to base the implementation on Equation (2.2). Although the differences between these two implementations will be discussed in more detail in Chapter 3, it is important to introduce what is termed quadrature modulation in Equation (2.2).

Direct implementation of Equation (2.2) as shown in Figure 2.2, would lead to quadrature modulator. In quadrature modulation two baseband signals are multiplied by two RF quadrature carriers and then summed. The baseband signal $a(t)\cos(\phi(t))$ and $a(t)\sin(\phi(t))$ are typically referred to as the in-phase (I) and quadrature (Q) baseband signals respectively. Likewise, the signals $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ are typically referred to as the I and Q carrier signals. Notice that the two carriers are orthogonal and thus can carry independent information in the same spectrum. This is another advantage of quadrature modulation.

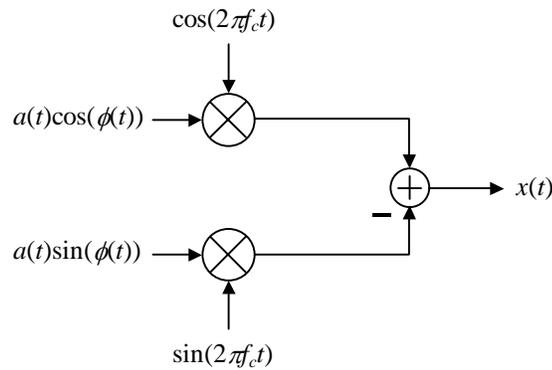


Figure 2.2 Quadrature modulator.

2.3.1 Constant Envelope vs. Non-constant Envelope

A modulated signal, as given in Equation (2.1), may contain information in both the amplitude, $a(t)$, and the phase, $\phi(t)$ of $x(t)$. If $a(t)$ is time-varying, the signal is termed non-constant envelope (NCE) because the envelope of the modulated signal is time varying. Angle modulation occurs if the information is contained in the $\phi(t)$ while $a(t)$ is constant. Angle modulation, either of the phase or frequency, may result in a constant magnitude signal in which the envelope of $x(t)$ does not vary with time. In this case the signal is termed constant envelope (CE). However, not all angle modulation techniques result in a CE modulated signal. For instance, a phase-modulated signal may contain abrupt phase transitions that will often lead to variations in the envelope and thus a non-constant envelope. Therefore, if $a(t)$ is time-varying or if abrupt phase transitions affect the amplitude of the carrier, the signal is NCE.

When designing a wireless communications system, the choice of CE versus NCE modulation has a significant effect on the design of the transmitter. Specifically the

linearity requirements change dramatically with the choice of modulation schemes. To better understand the linearity issue, $x(t)$ is applied to a memoryless third order nonlinearity and that results in the following:

$$y(t) = a_1x(t) + a_3x^3(t). \quad (2.3)$$

To understand the effect when $x(t)$ is NCE, a simplified case will be helpful. In this case, pure amplitude modulation (AM) is assumed and for simplicity $\phi(t)$ is set to zero. This results in

$$\begin{aligned} y(t) &= a_1a(t)\cos[\omega_c t] + a_3a^3(t)\cos^3[\omega_c t] \\ &= \left(a_1a(t) + \frac{3a_3a^3(t)}{4} \right) \cos[\omega_c t] + \frac{a_3a^3(t)}{4} \cos[3\omega_c t]. \end{aligned} \quad (2.4)$$

Notice that the signal now has spectral components at both ω_c and $3\omega_c$. Although the spectral component centered at $3\omega_c$ will need filtering, due to the relatively high frequency of the carrier signal compared with the baseband signal, it will have little effect on the desired signal centered at ω_c . However, the spectrum centered at ω_c has also been affected by the nonlinearity and to understand this effect a frequency domain representation is helpful.

Taking the Fourier transform of the term centered at ω_c results in

$$Y(f) = \frac{1}{2}a_1[A(f - f_c) + A^*(-f - f_c)] + \frac{3}{8}a_3[B(f - f_c) + B^*(-f - f_c)] \quad (2.5)$$

where

$$A(f) = \mathcal{F}[A(t)] \quad (2.6)$$

and

$$B(f) = A(f) * A(f) * A(f). \quad (2.7)$$

The results of Equation (2.5) are important and show that the signal centered at ω_c has been distorted by the non-linearity. To illustrate this effect Figure 2.3 shows an arbitrary NCE, AM modulated signal in both the time domain and the frequency domain. The signal is then applied to a third order non-linearity and the output signal is shown on the right side. Notice that in addition to the desired signal, the output now contains a third-order distortion term that is spread out in frequency. The spectrum centered at ω_c has been altered and this change is termed spectral regrowth. Spectral regrowth is a particular problem in NCE modulation schemes because nonlinearities in the transmitter cause can degrade the modulation accuracy and cause violations of the spectral mask requirements.

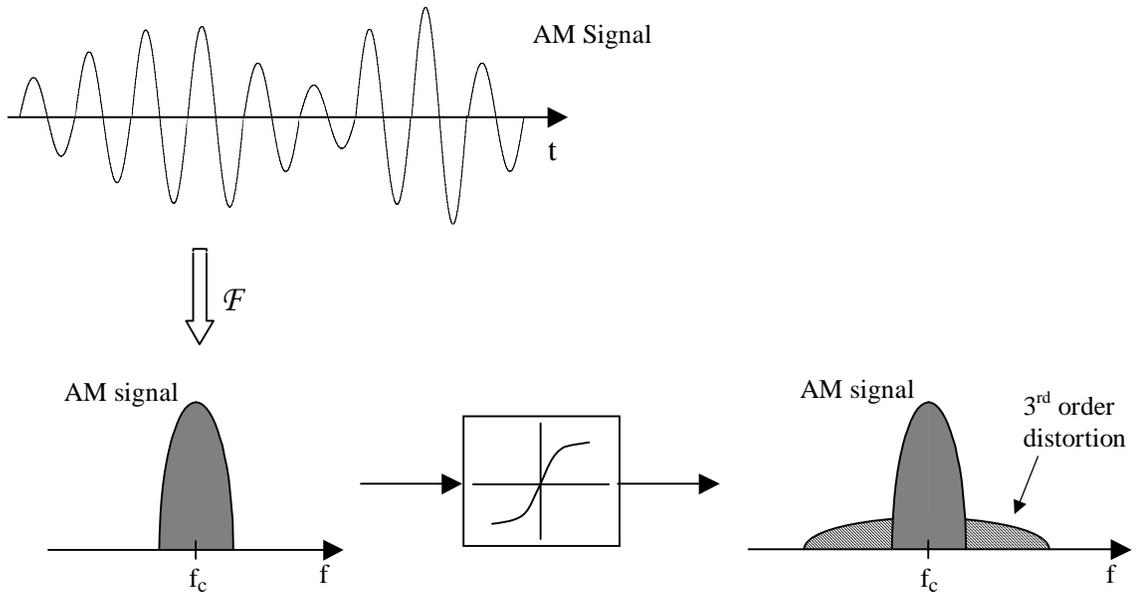


Figure 2.3 Third-order distortion of AM signal.

Although the example shown in Figure 2.3 examined a simplified case, the same effect is present when any envelope variations are present. As previously mentioned these envelope variations can be caused by AM modulation or angle modulation schemes that result in abrupt changes to the phase. Modulation schemes that require high linearity include QPSK, QAM, and some forms of FSK.

While NCE modulation is typically beneficial for spectral efficiency the linearity requirements lead to other problems. Specifically, in general a trade-off exists between power consumption and linearity. The higher the linearity requirement the more power is needed with all the other specifications fixed. This effect is of particular importance to the PA because often the majority of the power dissipated in a transmitter is done so in the PA particularly in cellular telephone transmitter. For example if a cellular telephone

PA transmits one watt of output power, with fifty percent efficiency, the PA dissipates two watts of power. While this is relaxed for smaller, lower power systems, the effect remains.

This trade-off between power consumption and linearity is only exacerbated when the goal is a single-chip CMOS implementation. Typically CMOS PAs are less efficient than PAs implemented in other technologies. Although research is being done to remedy this problem, currently, the best solution is to ease the linearity requirements of the PA. Consequently, integrating a linear PA in CMOS will negatively impact the power overall transmitter power consumption for two reasons: the inherent trade-off between linearity and power consumption, and the relative inefficiency of CMOS PAs compared with other technologies. Consequently, high power systems that employ NCE modulation are very difficult to implement in a single-chip CMOS solution.

To ease the linearity requirements a CE modulation scheme is often employed. When a CE signal is applied to a non-linearity the output spectrum near the carrier is minimally affected. Consequently, non-linear amplification may be used in the transmitter chain with little effect on the output signal. Therefore in CE systems highly efficient PA classes can be used such as class-C or class-E. In systems that require high output power, the PA will still dominate the overall power consumption but the effect is lessened.

2.3.2 GMSK Modulation

In digital modulation schemes, the binary data is mapped to the baseband signal. To recover this data, the signal is sampled by the receiver and a decision is made as to whether the transmitted bit was logical “1” or “0”. For example, a popular digital modulation scheme often used in cellular communications is Gaussian minimum shift keying (GMSK). A modulated GMSK signal may be represented by

$$s(t) = a_c \cos[2\pi f_c t + \phi_m(t)] \quad (2.8)$$

where f_c is the carrier frequency, a_c is the carrier amplitude, and the modulated phase, $\phi_m(t)$, is given by

$$\phi_m(t) = \sum_i \alpha_i \frac{\pi}{2} \int_{-\infty}^t g(\tau - iT) d\tau. \quad (2.9)$$

The symbol sequence $\alpha_i \in \{1, -1\}$ is the modulating index and $g(t)$ is defined as the convolution of the impulse response of a Gaussian filter with the *rect* function. This convolution can be written as

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma T} e^{-\frac{1}{2}\left(\frac{t}{\sigma T}\right)^2} * \text{rect}\left(\frac{t}{T}\right) \quad (2.10)$$

where T is the symbol period and σ is defined as

$$\sigma = \frac{\sqrt{\ln 2}}{2\pi B T} \quad (2.11)$$

where B is the 3 dB bandwidth of the Gaussian filter. The function *rect*(t) is defined as

$$rect(t) = \begin{cases} \frac{1}{T} & |t| < \frac{T}{2} \\ 0 & \text{otherwise} . \end{cases} \quad (2.12)$$

An important property of the function $g(t)$ is that

$$\int_{-\infty}^{\infty} g(t) = 1 . \quad (2.13)$$

Referring to Equation (2.9) and applying the property of Equation (2.13) it can be shown that for each data bit a phase change of $\pi/2$ radians occurs. It is also worth noting that the carrier amplitude is constant and only the phase is changing.

To graphically illustrate the impairments of a modulated signal a constellation diagram is often used. A constellation diagram plots the magnitude and phase of the complex envelope of the modulated signal in polar coordinates in the I-Q plane. The in-phase component is plotted on the x-axis while the quadrature component is plotted on the y-axis. A constellation diagram is useful in identifying non-idealities in the modulated signal and can help to identify transmitter impairments.

The constellation diagram of an ideal GMSK signal is shown in Figure 2.4. In the constellation diagram the distance from the origin represents the magnitude of the signal. The trajectory of the signal is represented by the dotted line and it is worth noting that the magnitude of the signal is constant for all time while the phase changes with each bit. Therefore, GMSK is a constant-envelope modulation scheme. Also due to inter-symbol interference (ISI) in the GMSK signal, each quadrant has three distinct points. This is in contrast to a modulation scheme such as QPSK that has one point in each quadrant.

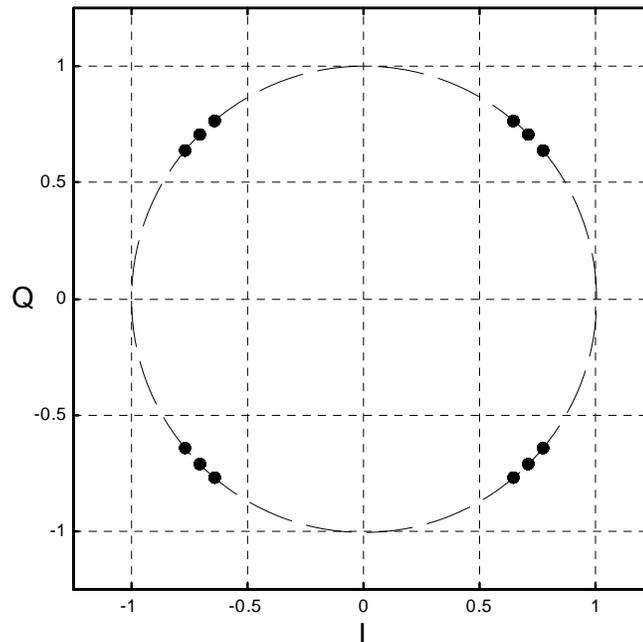


Figure 2.4 GMSK constellation diagram.

2.4 Performance Metrics

The growth of the wireless communications industry has been driven by the demand from users for constant connectivity. Systems that require the ability to accommodate a many users and simultaneously offer wide areas of coverage is a challenge to radio system designers. In such systems the wireless device must be able to transmit and receive data over long distances and with the chance that other users of the same system will be very nearby. These requirements put stringent performance constraints on both the transmitter and receiver. To quantify these constraints, radio standards set limits on certain aspects of the transmitted signal.

The limits set for the transmitter generally fall under two broad categories: modulation accuracy and spectral emissions. Modulation accuracy refers to how well the transmitted signal replicates the ideal signal. High modulation accuracy is important in a communications system to ensure that the transmitted signal has a high signal-to-noise ratio. This in turn allows for either a higher data rate or longer range.

Regarding spectral emissions, these are unwanted signals from the transmitter that may fall either in-band or out-of-band. These signals have the potential to interfere with users in the same system or even other systems. This may result in lower quality communications for other users or a lower number of total users. These limits will be discussed in more detail in the following sections.

2.4.1 Modulation Accuracy

The ideal RF transmitter would transmit the desired signal with out any deviation from the ideal signal. In reality, the transmitted signal is not an exact replica of the ideal signal and the difference between them falls under the general category of modulation accuracy.

The difference between the ideal signal and the actual signal is quantified depending on the type modulation. For digital modulation schemes, the difference between the ideal signal and the actual signal is shown more clearly using an I-Q constellation diagram. Shown in Figure 2.5 is a one quadrant of a constellation diagram with the ideal signal and a signal with errors. The difference between the two signals is a measure of the modulation accuracy.

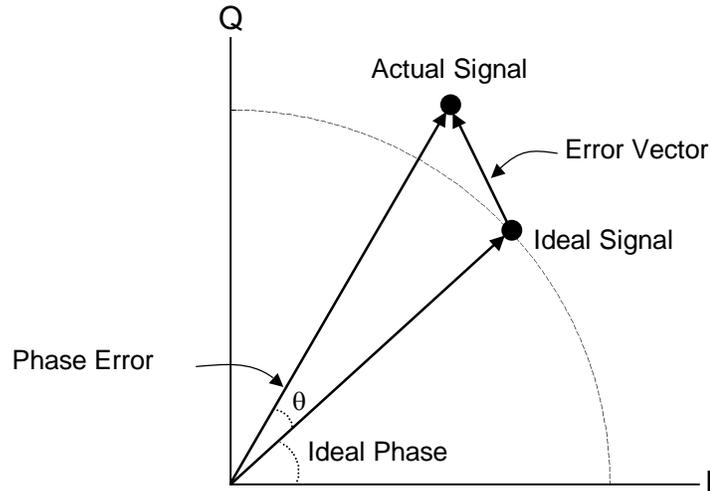


Figure 2.5 Graphical representation of the error vector and phase error.

For certain types of modulation the difference is measured by the error vector magnitude (EVM) while in other systems only the phase error is the critical metric. Generally, phase and frequency modulation schemes, which are constant envelope, use phase error as the metric and NCE modulation schemes use EVM. For example for GMSK modulation used in DCS1800 cellular phones, the standard requires an RMS phase error of less than 5 degrees.

2.4.2 Spectral Emissions

In addition to modulation accuracy, it is critical that a transmitter only emit a specified amount of radiation so as not to interfere with other devices both in the same system and in other systems. Ideally, the transmitter would only transmit a perfectly modulated signal with no other undesired spectral emissions. However, this is rarely the case and thus limits must be set for the levels of unwanted spectral emissions. The limit of the

spectral emission is set by each radio standard as a spectral mask requirement. While transmitting a signal, the emission levels must fall below the limits set by spectral mask.

The spectral mask requirements affect both in-band and out-of-band emissions. Unwanted emissions are caused by a number of factors including non-linearity in the system, noise resulting from interference with other circuits or spurious tones created by clocks or frequency synthesizers. Because these non-idealities affect the in-band signal they can also have an effect on the modulation accuracy. The in-band spectral mask requirement for DCS1800 is illustrated in Figure 2.6.

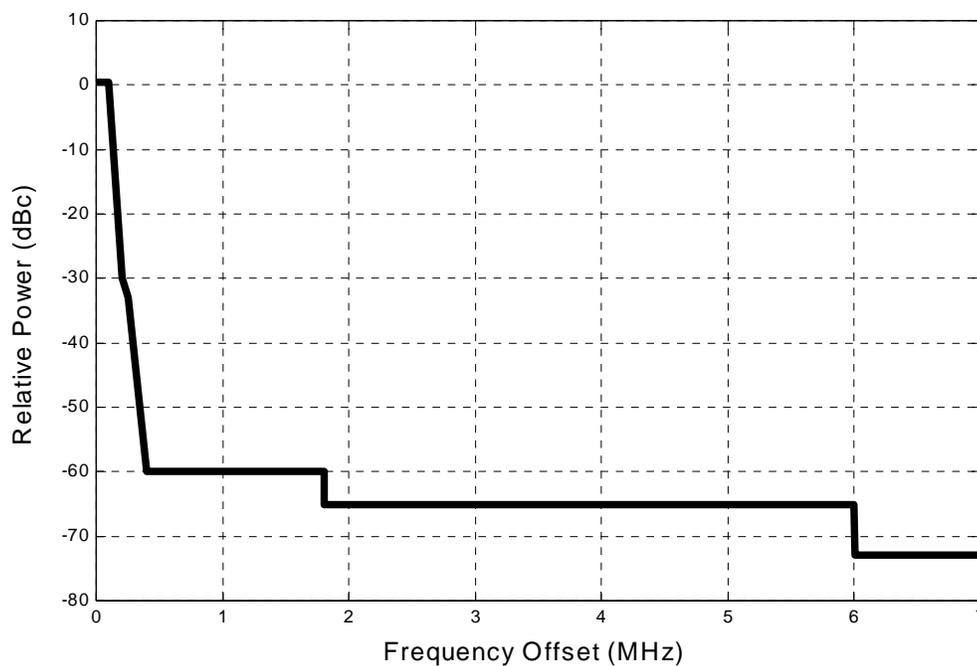


Figure 2.6 GSM in-band spectral mask requirement.

Typically, one of the most difficult portions of the spectral mask requirement is close to the carrier at. For example, Figure 2.7 shows the same GSM spectral mask at relatively low offset frequencies. In addition a GMSK modulated signal has also been illustrated in the same plot. Notice that the modulated signal is always below the spectral

mask. Non-idealities in the transmitter, which distort the signal, will bring the modulated signal higher and thus closer to violating the spectral mask. These non-idealities will be discussed in more detail in the following section.

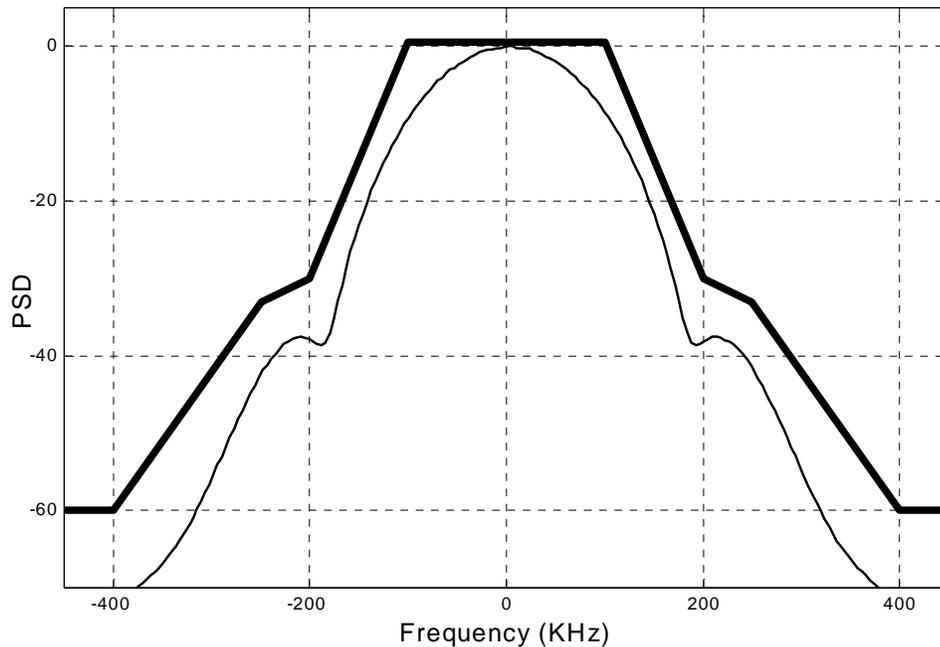


Figure 2.7 GSM spectral mask at low offset frequencies with GMSK modulated signal.

In addition to in-band requirements, out-of-band spectral emissions requirements also have an affect on transmitter design. While the out-of-band emissions don't affect the modulation accuracy or other transmitters in the same band, the limits are often lower than the in-band limits, making them more difficult to satisfy. For example, one of the most difficult requirements in the DCS1800 standard is the emissions requirement that falls in the DCS1800 receive band.

2.5 Non-Idealities

A wireless communications system must maintain a certain signal quality if it is to operate correctly. This implies that both the transmitter and the receiver must the receiver must perform the necessary functions while maintaining good signal quality. Furthermore the system depends on a channel that will not corrupt the signal beyond a certain point.

Within the transmitter, the exponential scaling of CMOS technology has vastly improved the capabilities of the DSP. This in turn has allowed for more complex digital modulation schemes, which results in a higher throughput. However, the impairments caused by the analog and RF sections of the transmitter have not benefited from scaling and are still very important to the overall performance of the system.

Any deviation from the ideal signal will cause degradation in the overall system performance. These deviations include the effects of mismatch, noise, and distortion. Although the effect on the overall system performance will vary depending on the modulation scheme and the architecture,

2.5.1 Quadrature Phase Mismatch

Quadrature phase mismatch is typically caused by errors in the quadrature generation of the LO signals. Although the baseband signals can also have quadrature phase errors, in practice these are typically negligible. Figure 2.8 shows a block diagram of a quadrature modulator with phase error, θ , in the LO signal. The output, $x(t)$, now contains errors which will degrade the system performance.

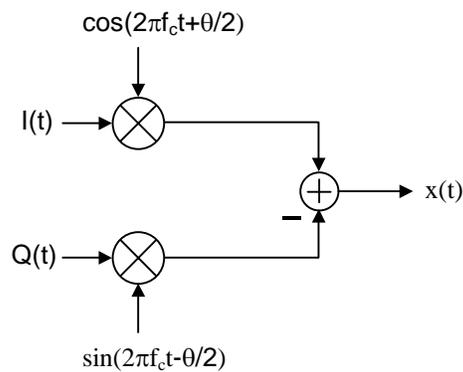


Figure 2.8 Quadrature modulator with phase mismatch.

Shown in Figure 2.9 is the constellation diagram of a GMSK signal with a quadrature phase mismatch. Ideally a GMSK signal has no envelope variations and thus the constellation points fall on a circle. However, quadrature phase mismatch causes envelope variations in the transmitted signal. This can be seen in the constellation diagram where half of the points fall inside the circle while the others are outside.

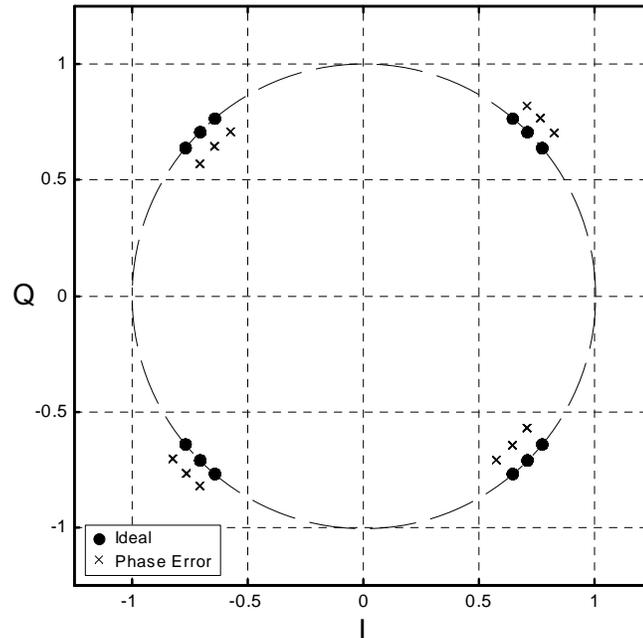


Figure 2.9 GSMK constellation with quadrature phase error.

To illustrate the effects of quadrature phase mismatch on the modulation accuracy, Figure 2.10 shows one quadrant of a constellation diagram with one ideal constellation point at phase $\pi/4$ and the points caused by a by a negative and positive quadrature phase mismatch of θ . In this example, because the phase of the ideal signal is $\pi/4$, the quadrature phase mismatch only causes a magnitude error, not a phase error. However, this is the not always the case and quadrature phase mismatch will often lead to phase errors as well.

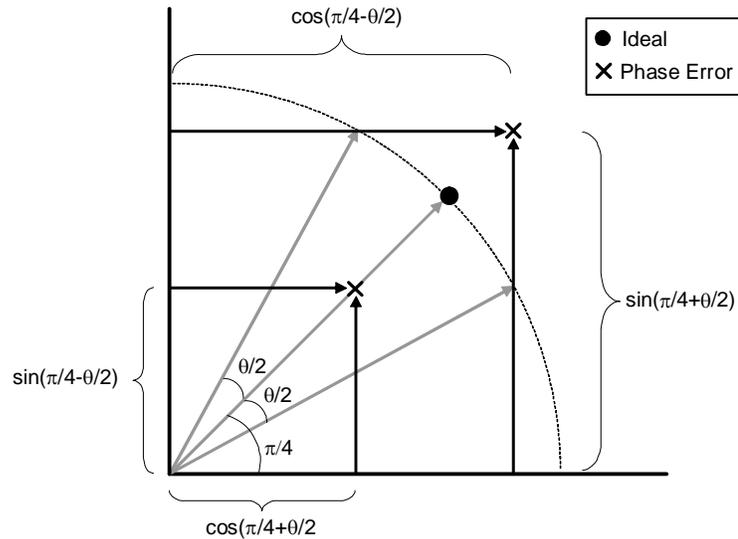


Figure 2.10 Quadrature phase error constellation.

Another common and useful graphical representation of accuracy of a transmitted signal involves altering the baseband input. Typically modulated data is applied to the baseband but in this test, the baseband inputs are low frequency sinusoids given by

$$I(t) = \cos(2\pi f_{bb}t) \quad (2.14)$$

and

$$Q(t) = \sin(2\pi f_{bb}t). \quad (2.15)$$

It might be expected that the up-converted spectrum would contain two frequencies, but in the ideal case either the negative or positive sideband is rejected, depending on the sign of $I(t)$ and $Q(t)$. When the modulator contains errors, the unwanted sideband is present and can be used as a measure of the accuracy of the modulator. For example the SSB test output of a modulator with two degrees of phase error is illustrated in Figure 2.11. The upper, desired sideband is, is located at the carrier frequency plus the baseband frequency

and the lower, undesired sideband is located at the carrier frequency minus the baseband frequency. In this case, two degrees of phase error leads to approximately 35 dB of sideband suppression or image rejection.

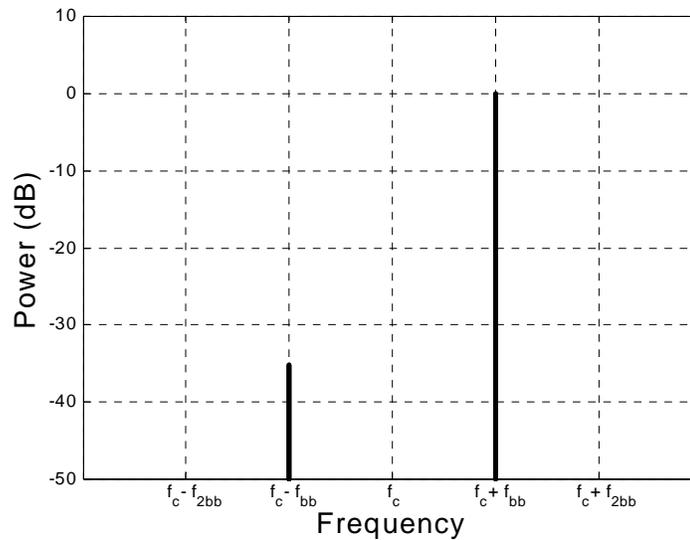


Figure 2.11 Single sideband spectrum with phase error.

2.5.2 Gain Mismatch

A quadrature modulator can also suffer from gain mismatches between the I and Q paths. Due to the nature of the mixers which are used the gain mismatch is typically caused by mismatch in the baseband signal, not the LO signal. Therefore to model a quadrature modulator with gain mismatch, as illustrated in Figure 2.12, the gain is inserted in the baseband signal path.

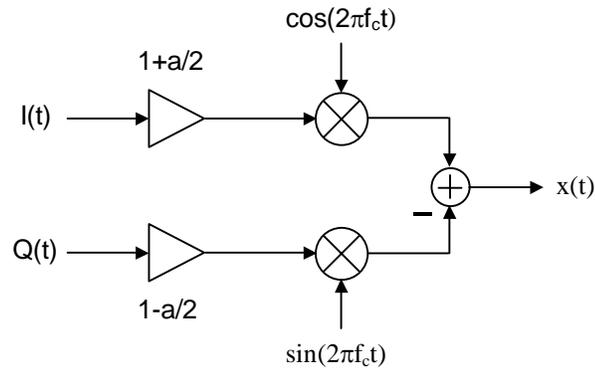


Figure 2.12 Quadrature modulator with gain mismatch.

Gain mismatch leads to a constellation diagram as shown in Figure 2.13. The in-phase component of each point is larger while the quadrature component is smaller which leads to a distorted constellation. Therefore envelope variations are introduced and the modulation accuracy is affected.

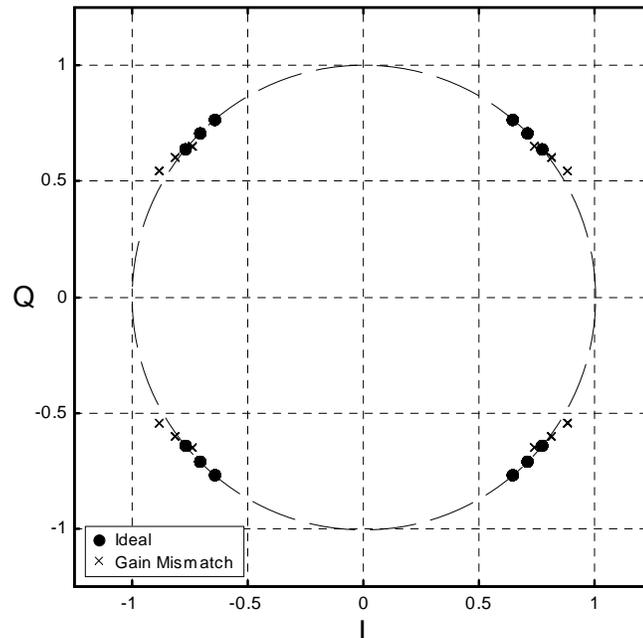


Figure 2.13 GMSK constellation with gain mismatch.

Much like the quadrature phase mismatch, gain mismatch will also be very evident in the SSB test. Shown in Figure 2.14 is a SSB test output of a quadrature modulator with a two percent gain mismatch. It is interesting to note that while a gain mismatch will affect both the constellation diagram and the SSB output, it has almost no affect on the modulated output spectrum. Consequently, the modulation accuracy will be affected by a gain error but the spectral mask will not.

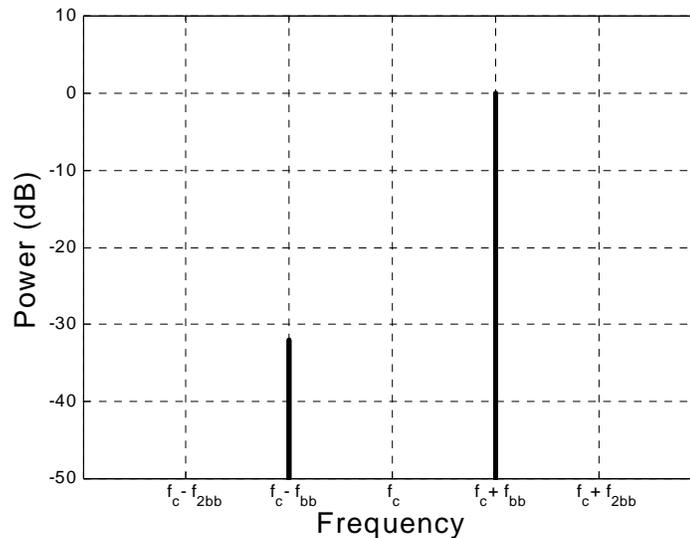


Figure 2.14 Single sideband spectrum with gain mismatch.

2.5.3 DC Offsets and LO Feedthrough

Circuit mismatches in the baseband section can also lead to DC offsets at the output of a quadrature modulator. The effect of a DC offset on a GMSK constellation is shown in Figure 2.15. The constellation is shifted in both the I and Q direction by an amount equal to the DC offset in each side of a quadrature modulator. The points in the ideal constellation fall on a circle and with the DC offset this is still the case. However, the

origin of this circle is shifted by an amount equal to the vector sum of the I and Q offset. The DC offset causes EVM errors and phase errors but the shape of the transmitted spectrum is relatively unaffected.

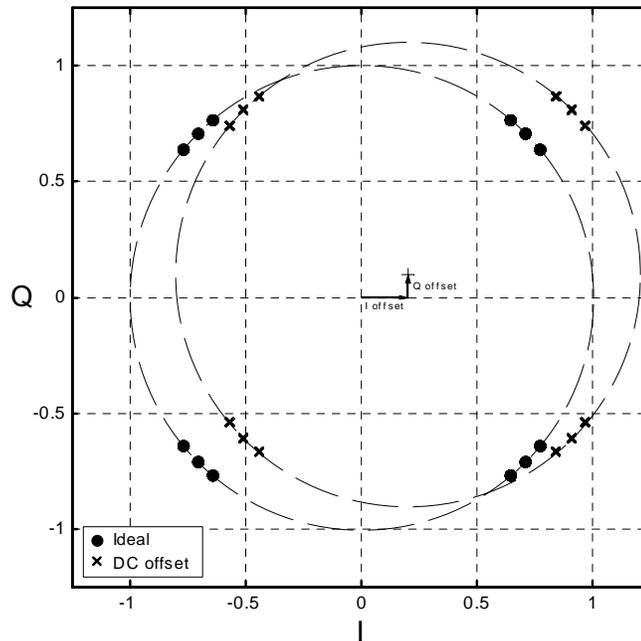


Figure 2.15 GMSK constellation with a DC offset.

DC offsets also have a very clear affect on the SSB output spectrum. Shown in Figure is a SSB spectrum with both quadrature mismatch and a DC offset. Much like Figure 2.16 the desired signal is located at a frequency of f_c+f_{bb} and the signal due to the quadrature mismatch is located at f_c-f_{bb} . However, the DC offset has generated a new frequency component at the carrier frequency. Although DC offsets are one cause of a carrier frequency component in the SSB test, other limitations, such as LO feedthrough, can also cause this effect.

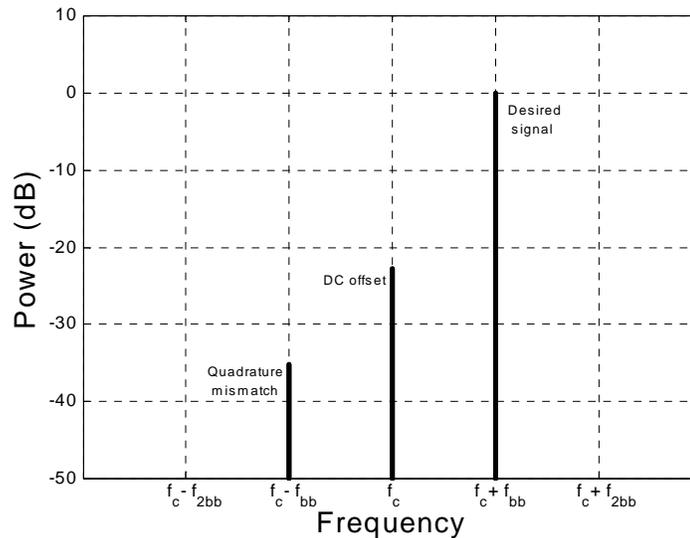


Figure 2.16 Single sideband spectrum with DC offset and quadrature mismatch.

The frequency synthesizer creates a tone that is mixed with the baseband signals. Coupling between the frequency synthesizer signal and the output of the mixers can lead to a tone in the modulated output located at the center frequency. In a single-sideband test, this LO feedthrough would be indistinguishable from a DC offset in the baseband and would lead to modulation error. This effect is a function of the coupling path and the frequency of the LO signal. Higher LO frequencies lead to larger LO feedthrough and thus increase the modulation error. The primary concern with LO feedthrough is corruption of the modulated signal. Therefore, it is typically only a problem when baseband signals are up-converted. When an intermediate frequency (IF) signal is mixed with an LO signal the feedthrough is typically well below the output signal and is much less of a problem.

2.5.4 Noise

Noise is present in all active devices and thus a common problem in transmitter design. The presence of noise lowers the SNR and that degrades the overall system performance. In CMOS transmitters the dominant noise types are typically thermal noise and flicker noise. Thermal noise is present in both passive resistors and active transistors. The mean-square voltage of a resistance R is given by

$$\overline{v^2} = 4kTR\Delta f \quad (2.16)$$

and defined by Boltzmann's constant, k , the temperature, T , and the bandwidth, Δf . Thermal noise is also present in CMOS transistors and in this case the mean-square noise drain current is given by

$$\overline{i_d^2} = 4kT\gamma g_{d0}\Delta f \quad (2.17)$$

where g_{d0} is the drain-source conductance when the drain-source voltage is zero and γ is a technology dependant parameter. The power spectral density (PSD) is constant over frequency for both of the previously mentioned thermal noise processes. The fact that the noise processes are white is particularly important for transmitter design.

White noise can lead to problems with the spectral mask requirement, especially in situations that require very low spectral emissions. This issue is exacerbated when switching mixers are used because the white noise in the baseband signal is copied to the harmonics of the LO signal. This effect, termed noise folding, raises the noise level both close to the carrier and far away. The close-in noise can degrade the SNR while the wideband noise can lead to spectral mask violations.

Shown in Figure 2.17 is a DCS1800 spectral mask and the power spectrum of two GMSK signals: one with added thermal noise and one without. The two spectra are shown together to illustrate the effects of the noise. At low offset frequencies the two spectra are indistinguishable and overlap one another. However, for frequency offsets above approximately 300 kHz and below -300 kHz, the noise is clearly evident leading to a spectral mask violations.

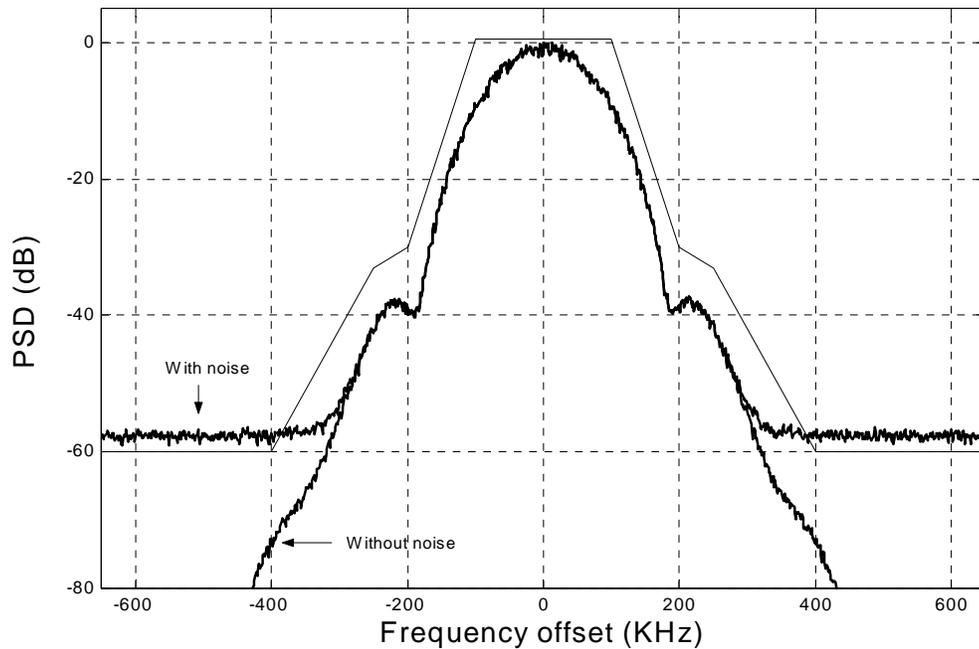


Figure 2.17 GMSK signal with thermal noise.

The effects of thermal noise on the constellation of a GMSK signal are shown in Figure 2.18. Thermal noise will clearly cause error in the magnitude and phase of the transmitted signal.

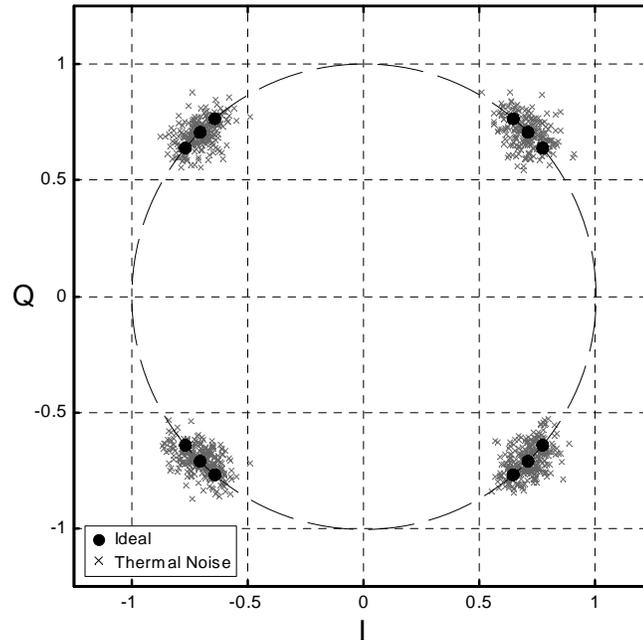


Figure 2.18 GMSK constellation diagram with thermal noise.

In addition to thermal noise, flicker noise can also negatively impact a transmitted signal. Flicker noise is present in CMOS transistors and the mean-square drain current is given by

$$\overline{i_d^2} = K \frac{I_D^a}{f} \Delta f \quad (2.18)$$

where K is a constant for a given device, I_d is the drain bias current, and a is a technology dependent constant. Due to the inverse relationship between the spectral density and the frequency, flicker noise is a more significant problem at low offset frequencies. Therefore flicker noise can impact the modulation accuracy of the transmitted signal but generally will have little impact on unwanted spectral emissions.

2.5.5 LO Phase Noise

The primary impairment that originates from the local oscillators is phase noise. While the LO is supposed to deliver a pure tone to the modulator, in practice this is not the case. Shown in Figure 2.19 are the spectra of an ideal LO output and an LO spectrum with phase noise. The phase noise is manifested as random deviations from the ideal LO frequency. This has the effect of altering the phase of a transmitted signal.



Figure 2.19 LO spectra. (a) Ideal spectrum. (b) Spectrum with phase noise.

Phase noise causes errors in both the modulation accuracy by affecting the phase of the transmitted signal. Shown in Figure 2.20 is a GMSK constellation diagram with added phase noise. The phase noise shifts the points on the constellation around the unit circle resulting in phase errors. However, the points all continue to lie on the unit circle and thus contain no envelope variations.

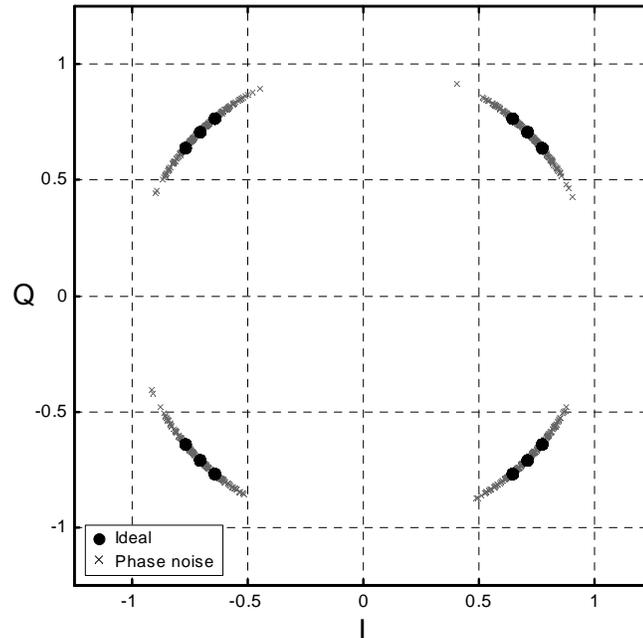


Figure 2.20 GSMK constellation diagram with phase noise.

In addition to affecting the modulation accuracy, phase noise also changes the modulated spectrum as illustrated in Figure 2.21. Much like the case with thermal noise, phase noise can raise cause spectral mask violations. In this example the phase noise is large enough to cause spectral mask violations close to the carrier. However, unlike thermal noise, phase noise decreases as the frequency of interest moves away from the LO frequency. Even with this property, phase noise is still a major concern at large offset frequencies because the spectral mask often decreases faster than the phase noise. As a result phase noise at large offset frequencies is often problematic in transmitter design.

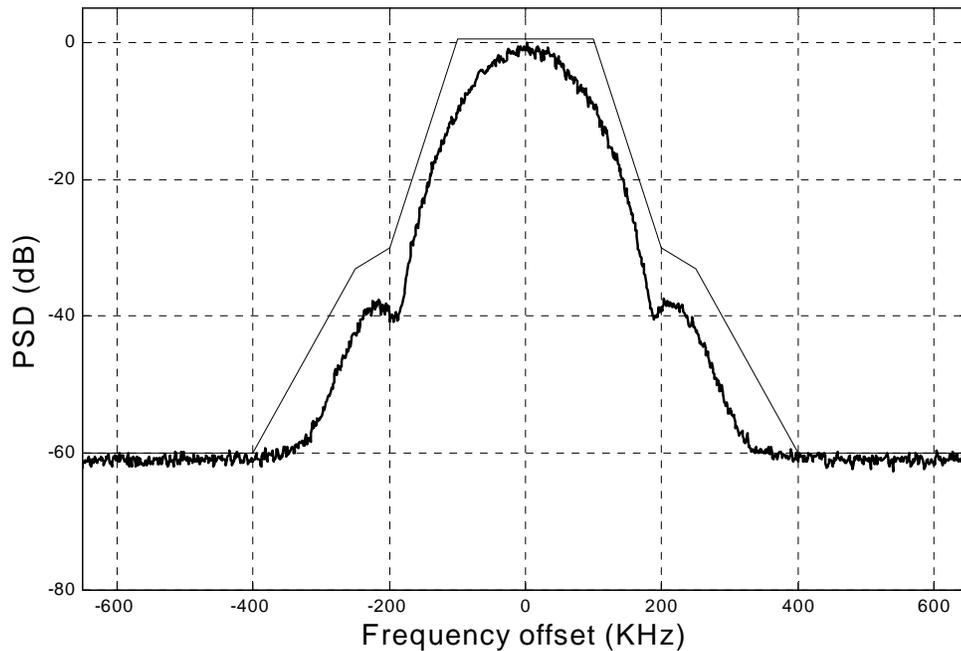


Figure 2.21 GMSK spectrum with phase noise.

2.5.6 Filters

Quadrature modulators typically employ a digital-to-analog converter (DAC) to interface between the digital signals of the DSP and the analog mixers. However, due to the nature of digital signals, a reconstruction filter is often used after the DAC is used to attenuate unwanted spectral content caused by sampling. A typical quadrature modulator is illustrated in Figure 2.22 with the DACs, filters and mixers. The design of this filter involves trade-offs between stop-band rejection and pass band distortion.

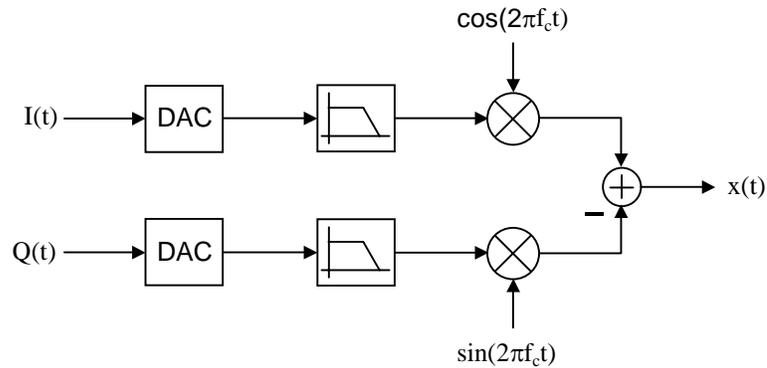


Figure 2.22 Typical quadrature modulator showing baseband circuits.

While the ideal low-pass reconstruction filter, with infinite stop-band rejection and zero in-band distortion, is not practical, many different filter responses are used in practice. To illustrate the trade-off involved with filter selection, three common low-pass filter responses, Butterworth, Bessel and Chebyshev [3], are compared. Shown in Figure 2.23 are the Magnitude responses for each of these three filters each designed with three poles and normalized in frequency. The stop-band rejection is best for the Chebyshev filter followed by the Butterworth and the Bessel. However, the stop-band rejection is only one factor in evaluating the effectiveness of a filter.

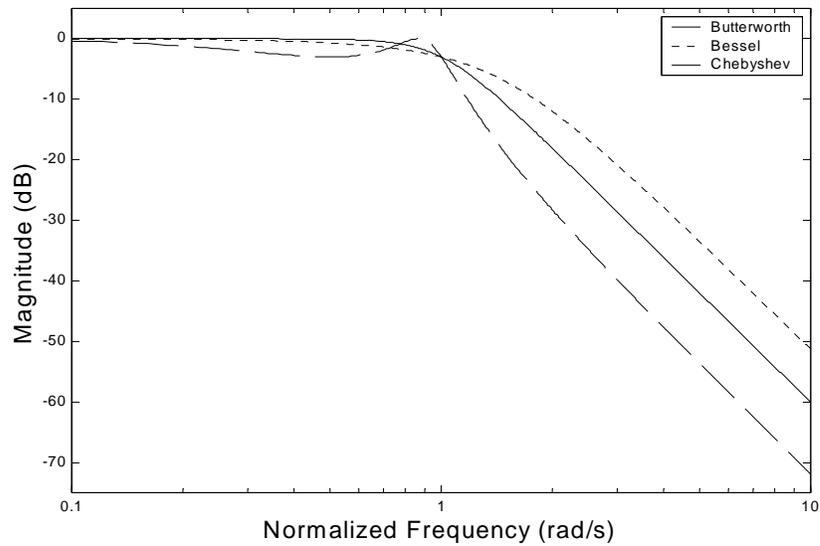


Figure 2.23 Magnitude responses of 3rd order Butterworth, Bessel, and Chebyshev filters.

The in-band frequency response is also critical in filter design. Shown in Figure 2.24 is the in-band magnitude response for the three filters. While the Chebyshev filter has the best stop-band rejection, it has the worst in-band amplitude distortion. The Butterworth response has a very flat amplitude response and moderate stop-band rejection. It is superior to the Bessel in both respects and it would seem a Bessel response would never be the correct choice. However, another factor that affects the choice of filters is the pass-band phase distortion.

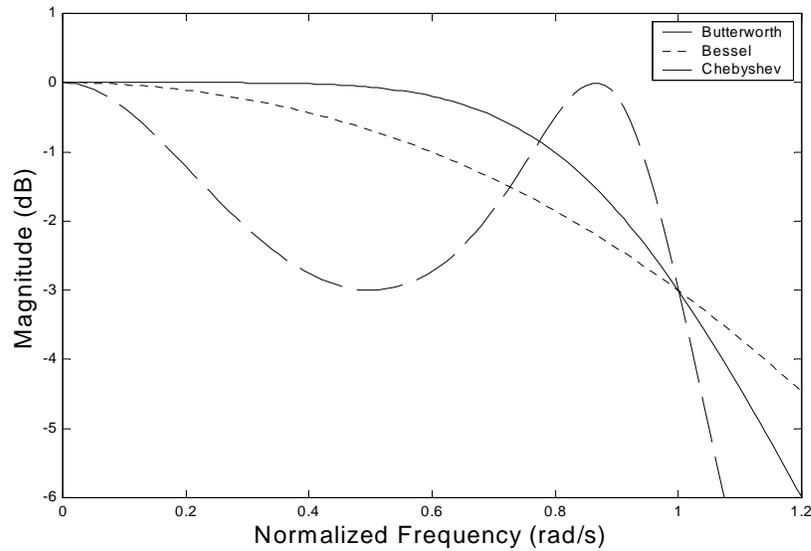


Figure 2.24 Passband magnitude responses of 3rd order Butterworth, Bessel, and Chebyshev filters.

Ideally, the pass-band phase response would be linear implying that the time delay for all frequencies would be equal. If the time delay were not equal, then different frequencies would experience varying delay, leading to signal distortion. This delay, typically termed group delay, is given by

$$D(\omega) = -\frac{d\theta(\omega)}{d\omega} \quad (2.19)$$

where $\theta(\omega)$ is the phase of the filter transfer function. The absolute value of the group delay is not critical but maintaining a constant value is important. Illustrated in Figure 2.25 is the pass-band group delay response of the three filters. Bessel filters are designed to minimize pass-band group delay but the Chebyshev filter clearly exhibits significant group delay variation. The Butterworth response exhibits fairly constant group delay that may be acceptable depending on the system requirements.

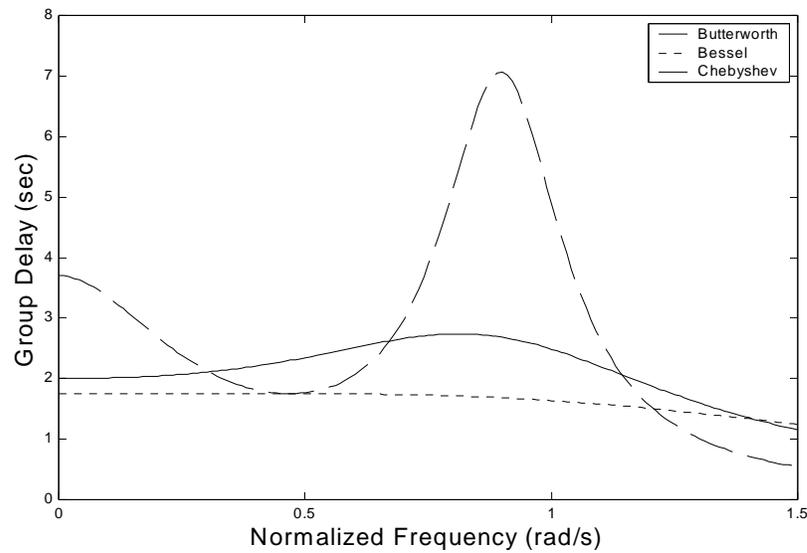


Figure 2.25 Group delay of 3rd order Butterworth, Bessel, and Chebyshev filters.

To illustrate the effects of filtering on a transmitted signal, the effects of a third-order Butterworth filter on a GMSK constellation are shown in Figure 2.26. The filter introduces errors in the modulation accuracy but has little effect on the spectral mask requirements. These errors are a function of the pass-band distortion and the position of the filter corner frequency relative to the bandwidth of the signal.

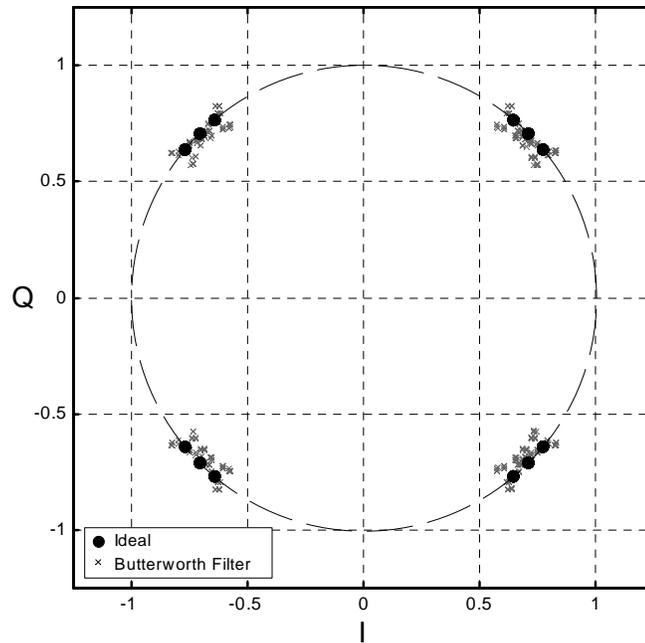


Figure 2.26 GSMK constellation with 3rd order Butterworth filter.

The choice of filters when for use in a transmitter depends largely on the overall system requirements. Trade-offs exist between pass-band distortion and stop-band attenuation. If the requirements place more emphasis on stop-band rejection then a filter with greater stop-band attenuation and consequently larger pass-band distortion, such as a Chebyshev filter, would be a suitable choice. Conversely, if pass-band distortion is critical, a Bessel filter might be ideal. The Butterworth filter represents a compromise between the two aforementioned filters and is often used in wireless transmitters. Other filter responses are also a possibility including elliptic and inverse Chebyshev filters.

2.5.7 Nonlinear Distortion

As was shown in the previous section, filters are one potential source of distortion in a transmitter. Other potential nonlinear elements include mixers, dividers, power amplifiers and DACs. The detrimental effect of these nonlinear elements depends on the properties of the input signal. If the signal $x(t)$ is applied to a nonlinear system, the output, $y(t)$ can be described by the following power series:

$$y(t) = a_1x(t) + a_2x(t)^2 + a_3x(t)^3 + \dots \quad (2.20)$$

If the input $x(t)$ is a pure sinusoid, the resulting output consists of a sum of sinusoids whose frequencies are integer multiples of the input frequency. The effect, termed harmonic distortion is illustrated in Figure 2.27. Harmonic distortion is relevant to the baseband and IF section of a transmitter. Harmonics created in the baseband can affect both modulation accuracy and the output spectrum and harmonics created at IF can also affect the output spectrum.

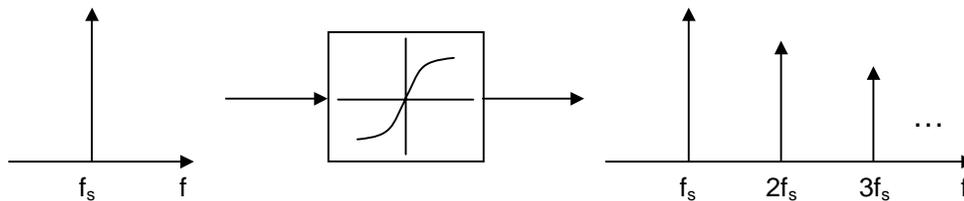


Figure 2.27 Harmonic distortion.

To quantify the effects of harmonic distortion, assume a single sinusoid given by

$$x(t) = X_i \cos \omega_i t \quad (2.21)$$

is applied to a non-linear system. The output can be determined by substituting Equation (2.21) into Equation (2.20) resulting in

$$\begin{aligned} y(t) &= a_1 X_i \cos \omega_i t + a_2 X_i^2 \cos^2 \omega_i t + a_3 X_i^3 \cos^3 \omega_i t + \dots \\ &= a_1 X_i \cos \omega_i t + a_2 X_i^2 \frac{1}{2} (1 + \cos 2\omega_i t) + a_3 X_i^3 \frac{1}{4} (\cos \omega_i t + \cos 3\omega_i t) + \dots \end{aligned} \quad (2.22)$$

where the generation of the harmonic distortion is evident. The fractional third harmonic distortion or HD_3 , is determined by the ratio of the amplitude of the third harmonic to the amplitude of the fundamental and is given by

$$HD_3 = \frac{a_3}{4a_1} X_i^2. \quad (2.23)$$

This is a common metric for quantifying the effects of distortion in a particular circuit or system.

Distortion of the baseband signals in a quadrature modulator can affect both the modulation accuracy and the spectrum. Shown in Figure 2.28 is the spectrum of a GMSK modulated signal with third-order baseband distortion along with an ideal spectrum and the GSM spectral mask. The distortion causes the modulated signal to violate the spectral mask requirements. Therefore the linearity of the baseband circuits is an important aspect of transmitter design.

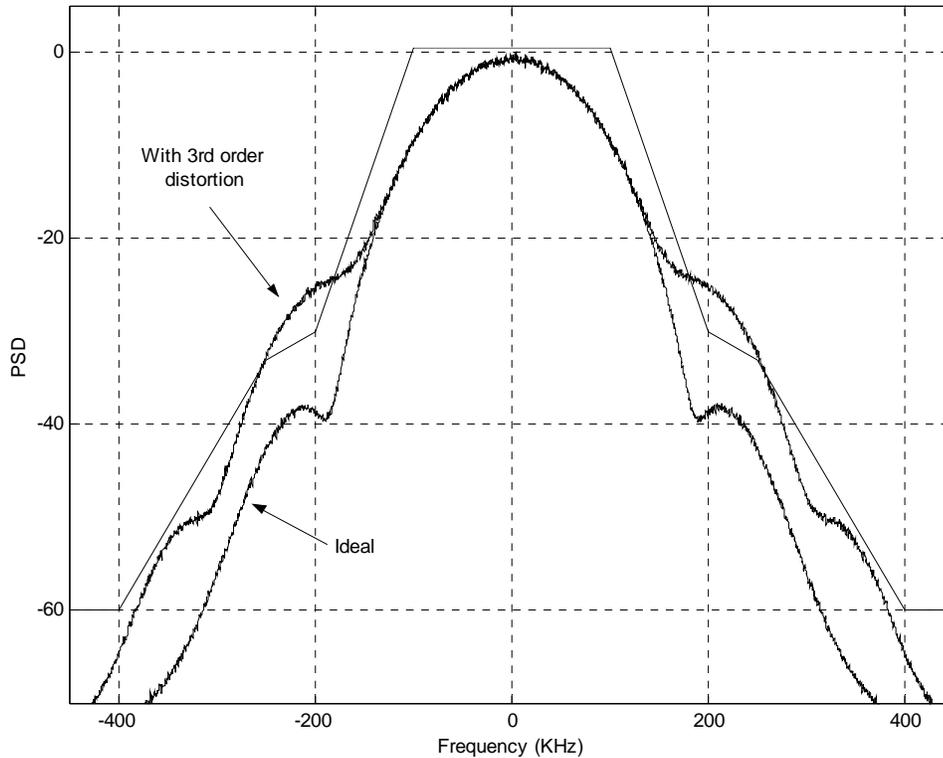


Figure 2.28 GSMK spectrum with third order distortion.

A second type of distortion is the result of multiple input signals. Intermodulation distortion occurs when the input to a nonlinear element consists of two signals. As shown in Figure 2.29, the input signal consists of two sinusoids at frequencies f_1 and f_2 . The resulting output signal has new components at $2f_1-f_2$ and $2f_2-f_1$. Intermodulation distortion usually occurs at IF or RF after harmonics have been created. These harmonics then intermodulate causing errors in the transmitted signal.

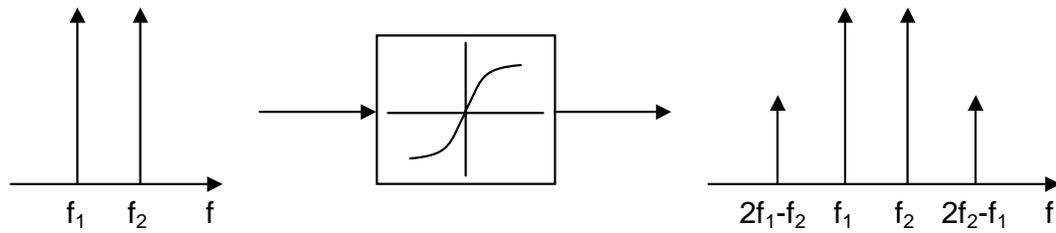


Figure 2.29 Intermodulation distortion.

The effect of nonlinear elements in a transmitter is particularly important in the PA. This problem is further exacerbated when NCE modulation is employed as was discussed in Section 2.3.1. Because the power consumption of the PA can dominate the overall power consumption, the transmitter should be designed to relax the linearity requirements on the PA. The transmitter architecture can significantly impact the linearity thus the potential for relaxed linearity requirements. Transmitter architectures will be discussed in more detail in the following chapter.

Chapter 3

Transmitter Architectures

3.1 Introduction

While the previous chapter focused on the fundamentals of transmitters including performance metrics and common sources of error, this chapter will focus on the architecture used to implement a transmitter. Transmitter architectures will be discussed with respect to three important criteria: the potential to fully integrate the device on a single CMOS substrate, the ability to achieve the high performance that is necessary in modern wireless communications and the ability to operate with multiple radio standards. The radio standards may employ different types of modulation and varying bandwidth which, require a flexible architecture. High performance is required if the multiple

standards are to include cellular communications and high data rate WLANs. Another factor that will influence the choice of architectures is power consumption. For high output power, the PA typically dominates the power consumption of the entire transmitter. However, the choice of architecture influences the power consumed by the PA.

This chapter will discuss the relative merits of mixer based transmitters versus phase locked loop (PLL) based transmitters. Furthermore a number of mixer based transmitters will be evaluated based on the previously mentioned criteria.

The fundamental role of a transmitter is to convert a baseband signal to a high-power RF signal. Shown in Figure 3.1 is a block diagram of a generic transmitter. The input is a baseband signal containing the desired information. This signal is modulated and is then typically filtered before up-conversion to RF. The RF signal is then amplified by a PA and again filtered before going to the antenna. The details regarding the sequence and structure of these blocks are dependant on the overall transmitter architecture and will be the focus of this chapter. In many cases, both filters, the PA and parts of the modulator and up-converter are discrete components.

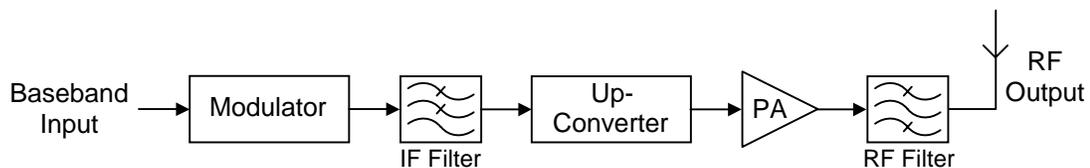


Figure 3.1 Generic transmitter architecture.

3.2 PLL Based Architectures

Transmitters generally fall into two broad categories: VCO based and mixer based transmitters. Although mixer based transmitters employ VCOs to generate LO signals, they do not directly modulate a VCO. Alternatively, in a VCO based transmitter, either the baseband signal or a modulated signal at IF, directly modulate a VCO. To accurately control the transmitted frequency, feedback is typically employed around the VCO in the form of a PLL. The PLL maintains a constant center frequency and can shape the unwanted noise from different circuits.

Shown in Figure 3.2 is a block diagram of a common PLL-based transmitter termed the Offset Phase Locked Loop (OPLL) [4]. In the OPLL, the baseband signals are up-converted to IF using a quadrature modulator and the IF signal is then filtered to attenuate unwanted harmonics created in the mixing process. This filtered IF signal is then applied to a PLL where it modulates the RF VCO. The RF feedback signal is down-converted using a channel-select RF frequency synthesizer, LO_1 . Thus the frequency of the transmitted signal is offset from LO_1 by the frequency of LO_2 . The down conversion is used to perform channel selection and mitigate the need for an RF frequency synthesizer with finer channel spacing.

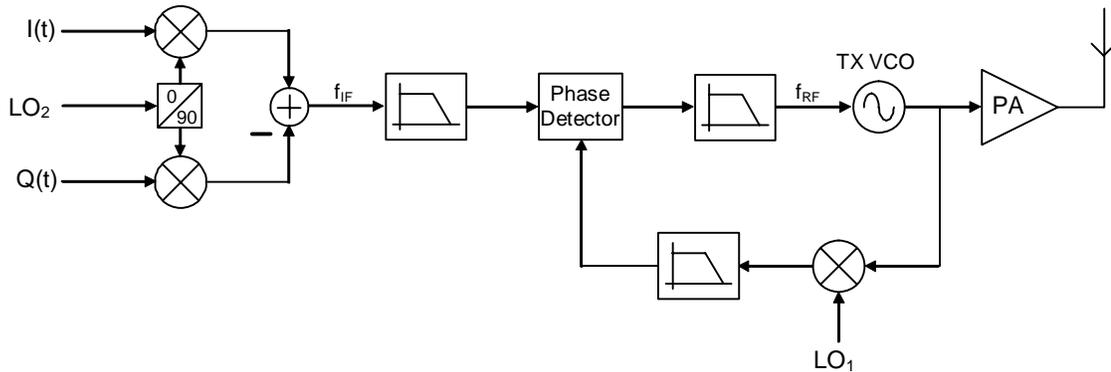


Figure 3.2 Offset phase locked loop architecture.

The OPLL behaves like an up-converting bandpass filter for the modulated IF signal. The out-of-band noise generated by the quadrature modulator is reduced by this filtering effect and this in turn can eliminate the need for an RF filter after the PA. With other transmitter architectures an RF filter is typically used after the PA for GSM systems because of the stringent noise requirements in the receive band. However, the use of an RF filter after the PA increases the overall power consumption because any losses after the PA require more power delivery from the PA. Since the PA can often dominate the overall power consumption, losses after the PA can have a significant impact on the total power consumption. However, the benefit of eliminating the RF filter, which is enabled by the OPLL, comes with a cost.

The output of the OPLL comes directly from a VCO and consequently, the amplitude of this signal is fixed. This makes the OPLL useful only for constant envelope modulation schemes. Therefore, the OPLL is not well suited for use as a multi-standard transmitter because of its inability to operate with non-constant envelope modulation. To address this problem a modification has been made.

To accommodate NCE modulation schemes, PLL-based transmitters have been modified with an additional amplitude control. These transmitters are typically termed polar transmitters because the rectangular coordinates, I and Q, are converted to polar coordinates and the signal is then represented by phase and amplitude. Separate signal paths for the phase and amplitude are used which are then combined at the output [5]. The basic idea behind current implementations of polar transmitters involves applying an OPLL for the phase modulation and a second path can be added to apply the amplitude modulation. Shown in Figure 3.3 is a block diagram of a generic polar transmitter. The data is converted to phase, ϕ , and amplitude A with baseband signal processing. The phase signal is modulated with an OPLL and the amplitude signal typically modulates the PA. In this way the transmitter can operate with NCE modulation and maintain the benefit of eliminating the need for an RF filter before the antenna. While the phase modulation is typically performed in a closed loop, the amplitude control can be applied either open loop [6][7][8], without feedback, or closed loop [9].

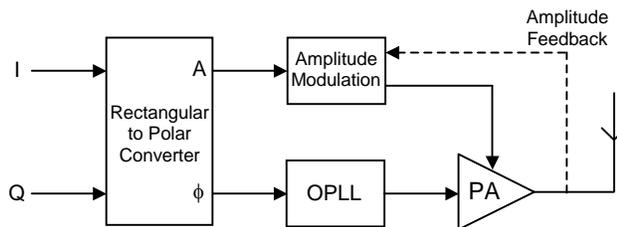


Figure 3.3 Basic polar transmitter block diagram.

A critical issue for polar transmitters is timing matching in the phase and amplitude paths. When the separate paths are combined to generate the RF modulated signal, mismatch in timing can lead to errors in the modulation accuracy or violations of

the spectral mask requirement. The matching requirements become even more difficult as the bandwidth of the transmitted signal is increased. Consequently, polar transmitters are not well suited for multi-standard operation where the signal bandwidths may vary from hundreds of kilohertz to tens of megahertz. This problem is exacerbated due to the nature of polar signals. The polar representation of a signal occupies significantly more bandwidth than its rectangular counterpart [7][8]. These larger bandwidths make the matching even more difficult.

Another potential problem for polar transmitters arises when open loop amplitude modulation is used. Typically an OPLL will use a high-efficiency, non-linear PA. This is possible because the envelope of the RF signal is constant. In the case of polar transmitters, a non-linear PA is also desirable. However, to maintain the benefits of using a highly non-linear, and thus more efficient, PA, the behavior of the PA needs to be well understood. Applying amplitude modulation (AM) to a non-linear PA will distort the amplitude output signal. Since this AM-AM distortion is not corrected with a feedback loop, pre-distortion is often required to achieve the necessary linearity. This use of pre-distortion requires calibration of the PA. Characterizing an integrated PA is not an option making this architecture ill suited for single chip integration.

Alternatively, closed loop amplitude modulation may be used in a polar transmitter. By employing feedback on both the phase and amplitude paths the need for characterization of the PA can be eliminated. However, with closed loop amplitude feedback the timing mismatch and bandwidth issues are even more severe.

Although PLL-based transmitters work very well for certain applications, they are limited with respect to single-chip integration and multi-standard operation. Mixer based

transmitters will be discussed in the following sections with respect to these two critical criteria.

3.3 Direct Conversion Architecture

The direct-conversion architecture, shown in Figure 3.4, is attractive because of the simplicity of the signal path. The digital baseband I and Q signals are converted to analog signals with two DACs and then filtered to attenuate the aliasing from the DAC. The signals are then modulated and up-converted to RF in a single step by the quadrature modulator. The RF signal then passes through a discrete bandpass filter before amplification by the PA. This pre-PA filter is not always necessary and is used to reduce intermodulation in the PA and to attenuate wide-band noise. Finally, after the PA, another discrete bandpass filter is typically needed to meet the spectral mask requirements. A single RF frequency synthesizer performs channel selection.

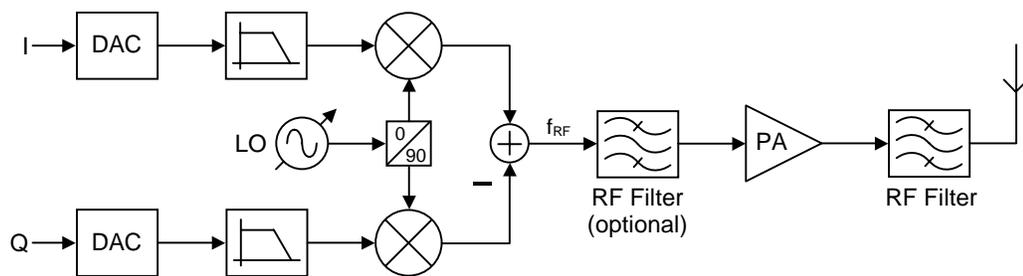


Figure 3.4 Direct-conversion transmitter block diagram.

The performance of a direct-conversion transmitter is typically limited by two factors. First, like all quadrature modulators, this architecture requires the generation of quadrature LO signals. Quadrature generation at RF is in less accurate than quadrature

generation at lower frequencies. Consequently, the modulation accuracy of the transmitter is reduced by this effect.

Another limitation to the performance of the transmitter involves intermodulation in the PA. When the baseband signals are mixed with the LO signals, harmonics of the LO are created. These harmonics, which will be discussed in more detail in the following chapter, are inherent to switching mixers. The harmonics create copies of the baseband signal which then intermodulate in the PA and reduce the modulation accuracy. To understand this effect, a quadrature modulator is shown in Figure 3.5 with tones applied to the baseband. The spectrum before the non-linear PA includes the desired signal, located at the sum of the LO and baseband frequencies, and an undesired signal located at one baseband frequency below the third harmonic of the LO. Third order intermodulation of two tones results in new tones located at $2f_1-f_2$ and $2f_2-f_1$. In this case the frequency of the distortion component in the output is given by

$$(3f_{LO} - f_{BB}) - 2(f_{LO} + f_{BB}) = f_{LO} - 3f_{BB}. \quad (3.1)$$

The new component, located three baseband frequency steps below the LO frequency, can lead to modulation error and spectral mask violations. As a result, the filter before the PA is often needed to attenuate the signals before intermodulation occurs in the PA. However, the need for this filtering before the PA often requires a discrete component and is thus not amenable to single chip integration.

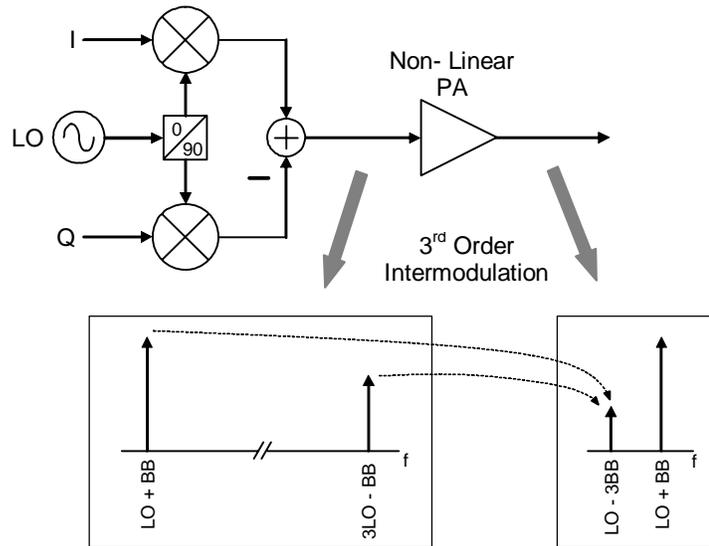


Figure 3.5 Third order intermodulation by a non-linear PA in a direct-conversion transmitter.

Another potential problem for this architecture involves interaction between different circuits. If the goal is an entirely integrated transmitter, large signals created by the PA will be present. These large signals can interfere with other sensitive analog circuits by coupling through the substrate. A particular problem occurs between the frequency synthesizer and the PA. The large modulated signal at the output is very close in frequency to the tone created by the frequency synthesizer. As shown in Figure 3.6 this effect, is termed LO pulling or injection pulling [10] because the large signal from the PA can pull and change the frequency of the LO. It is a particular problem for direct conversion transmitters because RF output and the LO are so close in frequency. Methods have been developed to reduce the effects of LO pulling but they introduce new problems.

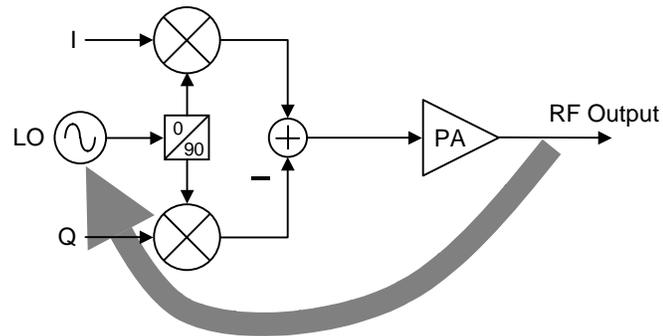


Figure 3.6 LO Pulling in a direct-conversion transmitter.

To maintain the direct-conversion architecture and reduce LO pulling a few methods have been used. One method involves using two frequency synthesizers and mixing them to create the LO signal [11][12]. In doing so, the frequency of the RF LO is moved away from the frequency of the PA output thus reducing the LO pulling. However, additional harmonics are created with this technique that either need to be filtered or the performance may suffer. A second technique is used in which a single frequency synthesizer output is divided and then mixed to create the LO signal [13]. Again, the frequency of the RF LO is changed to reduce the LO pulling. Yet another method employs multiple phases of a lower frequency LO to perform up-conversion [14]. This method also avoids an LO operating near the transmitted frequency. Although all of these solutions lessen the impact of LO pulling, they all add to the overall complexity and thus diminish the attractive simplicity of the direct-conversion transmitter. Furthermore, none of them address the other potential performance problems associated with direct conversion transmitters.

Other potential issues with direct conversion transmitters include the difficulty of power control and LO feedthrough. Because there is no IF stage, all the power control

must be done at baseband or RF. Baseband power control can place very strict linearity requirements on the baseband circuits. Coupling between the LO signal to the output is another potential problem due to the high frequency of the LO signal used in the quadrature modulator.

Although direct a conversion transmitter has a very simple signal path and offers potential for multi-standard operation, a number of limitations associated with the architecture exist. Some of these problems can be alleviated by using two steps to perform the up-conversion. This so-called heterodyne architecture will be the focus of the next section.

3.4 Heterodyne Architecture

The heterodyne architecture, which uses two steps to perform up-conversion, is a widely used topology. The heterodyne architecture, illustrated in Figure 3.7 uses a quadrature modulator to frequency translate the baseband signal to IF. At this point, an IF filter is needed to suppress harmonics created in the mixing process. The IF signal is then up-converted to RF where an RF filter is needed to attenuate the image created by the RF mixer. Finally, the RF signal is applied to the PA and typically filtered again to meet the spectral mask requirements.

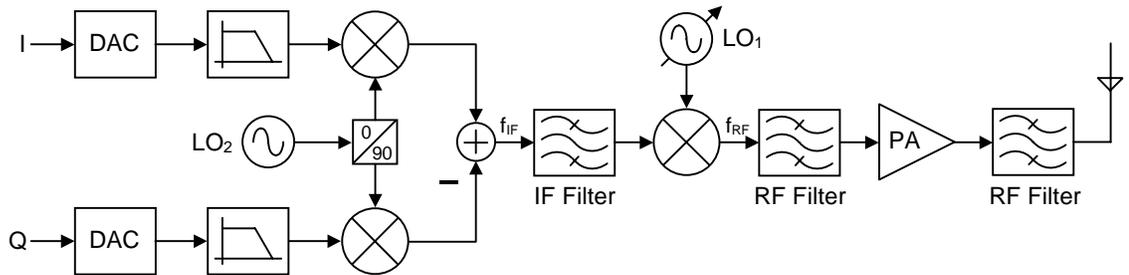


Figure 3.7 Conventional two-step or heterodyne transmitter block diagram.

The heterodyne architecture has a number of performance advantages over the direct-conversion approach. This architecture avoids the LO pulling problem because up-conversion is performed in two steps and therefore neither LO is operating at the transmitted frequency. Modulation accuracy in the quadrature modulator is also improved because the lower frequency LO allows for more accurate quadrature generation. Furthermore, the lower frequency of the LO leads to less LO feedthrough at the output. The issue of third order intermodulation is also alleviated because the signal located near $3(\text{LO}_1 + \text{LO}_2)$ is attenuated by the IF filter. Although the heterodyne architecture shows promise for higher performance, the architecture does have some drawbacks with respect to single chip integration.

To achieve the performance benefits previously mentioned two filters are needed before the PA. The IF filter is needed to attenuate harmonics created in the quadrature modulator and the RF filter before the PA is needed to filter the image created by the final up-conversion. Both of these filters are typically discrete components that are not amenable to integration on a single CMOS substrate.

Although single chip-integration with the heterodyne architecture may be difficult to achieve while maintaining high performance, the architecture is well suited for multi-standard operation because heterodyne transmitters are capable of both CE and NCE modulation schemes. Therefore, if the limitations to integration can be overcome, this architecture is a good choice when the goal is multi-standard operation on a single CMOS integrated circuit. The next chapter will address some of the critical issues involved with integrating a two-step transmitter.

3.5 Transmitter Architecture Comments

Although many different transmitter architectures have been developed, if the goal is a multi-standard, high performance transmitter, integrated onto a single CMOS substrate, the existing architectures are limited in some respects. The most promising architecture is the heterodyne or two-step architecture but it is limited with respect to complete integration. One of the key issues with integration is the need for filtering, which is difficult to perform on chip. Architectures in which the filtering requirements are reduced are therefore more amenable to complete integration.

The filters in the heterodyne architecture filter harmonics and image signals created in the mixing process. These harmonics are fundamental to switching mixers and attenuating them is critical to full integration. Chapter 5 will introduce a two-step transmitter that does not require the use of a discrete IF filter because the harmonics have been attenuated. Fundamental to this architecture are mixers that can attenuate the harmonics and thus relax the filtering requirements leading to removal of the IF filter.

These mixers will be discussed in the following chapter allowing for a better understanding before the discussion of the complete architecture.

Chapter 4

Harmonic-Rejection Mixer

4.1 Introduction

A critical component in both RF transmitters and receivers is the mixer. Mixers are used to frequency translate the signal either from baseband to RF, as is the case in a transmitter, or RF to baseband in a receiver. Ideally this frequency translation would occur without the generation of other, unwanted signals. This would happen if the frequency translation was implemented with an ideal multiplier. However, in practice, multipliers are very difficult to implement and mixers are used instead. Although mixers perform the multiplication operation, they are not ideal and other unwanted signals are created in the process.

The fundamental role of a mixer is to multiply two signals in the time domain. To understand why this is important for frequency translation consider two signals $b(t)$ and $c(t)$ defined as

$$b(t) = \cos(\omega_b t) \quad (4.1)$$

and

$$c(t) = \cos(\omega_c t). \quad (4.2)$$

Based on a trigonometric identity, the multiplication of these two sinusoids results in the following output:

$$\begin{aligned} b(t)c(t) &= (\cos \omega_b t)(\cos \omega_c t) \\ &= \frac{1}{2} \cos(\omega_c - \omega_b)t + \frac{1}{2} \cos(\omega_c + \omega_b)t. \end{aligned} \quad (4.3)$$

Although the input consists of two tones at frequencies ω_b and ω_c , after multiplication, the output consists of tones at the sum and difference of frequencies ω_b and ω_c . Therefore frequency translation, both up and down, has occurred when the two signals were multiplied.

Multiplication in the time domain corresponds to convolution in the frequency domain. The results of Equation (4.3) can be rewritten in the frequency domain to provide some further insight. The frequency domain representation of equation (4.3) is given by

$$\begin{aligned}
 B(\omega) * C(\omega) &= \pi[\delta(\omega - \omega_1) + \delta(\omega + \omega_1)] * \pi[\delta(\omega - \omega_2) + \delta(\omega + \omega_2)] \\
 &= \frac{\pi}{2}[\delta(\omega - (\omega_1 - \omega_2)) + \delta(\omega + (\omega_1 - \omega_2))] \\
 &\quad + \frac{\pi}{2}[\delta(\omega - (\omega_1 + \omega_2)) + \delta(\omega + (\omega_1 + \omega_2))].
 \end{aligned}
 \tag{4.4}$$

A graphical representation of Equation (4.4) is illustrated in Figure 4.1. The multiplication has created signals at the sum and difference of the two input frequencies. As will be shown in subsequent section, this frequency domain representation is very useful when examining the non-ideal behavior of mixers.

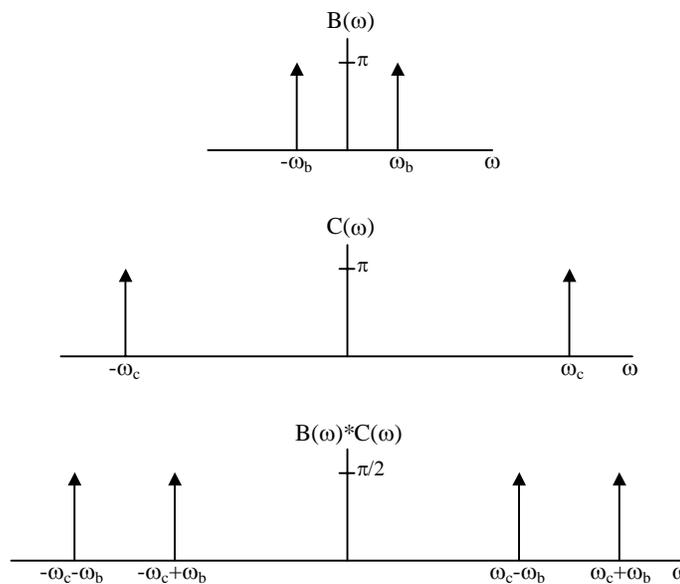


Figure 4.1 Graphical representation of convolution of two cosine functions in frequency domain.

In practice, mixers do not implement a perfect multiplication of two signals. One of the mixing products is the product of the two signals but many other signals exist due to the nature of switching mixers. The next section will illustrate why these unwanted signals are created in virtually every mixer that is used.

4.2 Switching Mixers

The vast majority of mixers used today, fall under the category of switching mixers. The name stems from the fact that transistors acting as switches are used to either commutate a current or a voltage to provide the multiplication. These mixers can be either passive or active mixers but for transmitter applications, active mixers are typically used because of their higher conversion gain. Within the broad category of switching mixers, current commutating, active mixers are frequently used.

In a typical current-commutating mixer, the input voltage is typically converted to a current and then the path of the current is controlled with switches. A simplified current-commutating active mixer is illustrated in Figure 4.2. The LO signal is applied to the gates of the upper transistors and a baseband signal controls the lower transistor. In addition to the baseband signal, a bias voltage is used to set a bias current. The voltage swing of the LO is typically large compared to the turn-on voltage of the transistors and thus the transistors behave like switches. The baseband voltage signal is converted to a current signal by the lower device and this current is then switched between the positive and negative output by the LO controlled switches. This effectively multiplies the baseband current by a square wave version of the LO at the output.

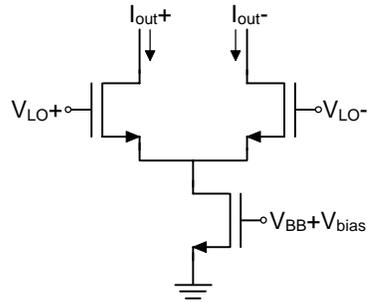


Figure 4.2 Single-balanced current-commutating mixer.

Since the output of the single-balanced mixer is essentially the product of the baseband signal and a square wave created by the LO signal, the output spectrum is the convolution of the baseband spectrum with the spectrum of a square wave. The Fourier series representation of the square wave illustrated in Figure 4.3 is given by

$$s(t) = \frac{4}{\pi} \left(\cos \omega t - \frac{1}{3} \cos 3\omega t + \frac{1}{5} \cos 5\omega t - \dots \right) \quad (4.5)$$

where $\omega = 2\pi/T$.

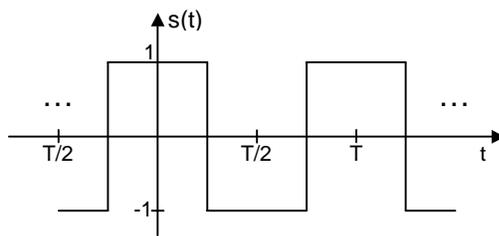


Figure 4.3 Unity amplitude square wave.

The square wave is composed of a fundamental cosine term and an infinite number of odd harmonics. Consequently, the output of a single-balanced mixer, when the baseband is driven with a sinusoid, is illustrated in Figure 4.4. The baseband signal has been mixed both up and down in frequency from the LO frequency. In addition,

harmonics of the LO, resulting from the square wave nature of the mixer, create copies of the baseband signal centered at the odd harmonics of the LO.

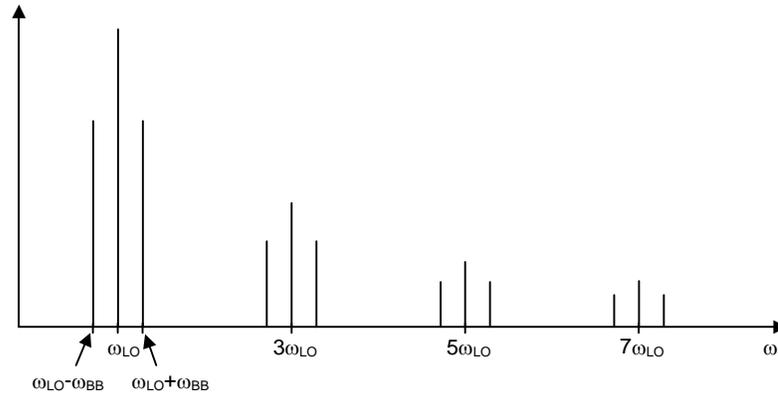


Figure 4.4 Output spectrum of a single-balanced mixer with sinusoidal baseband input.

In this case, due to the nature of a single-balanced mixer, a component also exists at the LO frequency and each odd harmonic of the LO. The LO component is caused by the bias current in the single-balanced mixer. The bias current is commutated along with the baseband current resulting a signal at the LO frequency and each odd harmonic of the LO. To avoid this problem, a double-balanced mixer is used to cancel the undesired LO signal at the output.

The vast majority of double-balanced mixers fall under the category of Gilbert cells [15]. The Gilbert cell, illustrated in Figure 4.5, is a current-commutating active mixer. Although originally designed with bipolar technology, the CMOS version retains the name. Like the single-balanced mixer, the LO signal is applied to transistors that behave like switches. The baseband signal now requires both a positive and a negative input which are applied to transistors acting as transconductors. These transconductors

convert the baseband voltage signal to a current signal which is then commutated by the switching transistors.

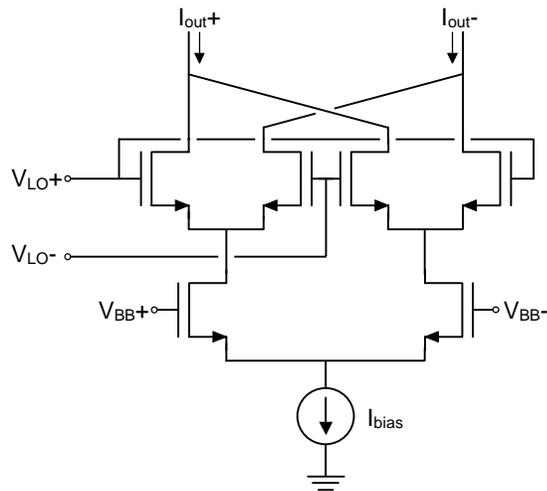


Figure 4.5 Double-balanced current-commutating mixer.

The Gilbert cell avoids the problem of LO feedthrough at the output because the bias current behaves like a common mode signal. To understand this, assume that the differential baseband voltage is zero but the LO voltage is still activating the upper transistors. Assuming the upper transistors are behaving like perfect switches, at any given moment, two transistors are on and two are off. The two transistors that are on will each carry half of the bias current, one from the positive output and one from the negative output. When the transistors switch, the other two transistors will then draw half of the bias current from either the positive or negative output. Consequently, the differential output current is zero as the upper transistors continue switching. As a result, the bias current is behaving like a common mode signal at the differential output. Although the transistors do not behave like ideal switches, this configuration still yields very high LO to RF isolation.

Although the double balanced mixer alleviates LO feedthrough, the switches still effectively multiply the baseband signal by a square-wave version of the LO signal. The harmonics created in this process present a considerable filtering problem. In reality, the switching transistors don't behave like perfect switches and have a finite transition time. As a result, the output of the mixer does not exactly replicate the product of the baseband signal and a square wave. However to maximizing the conversion gain of the mixer requires the switching transistors to act as close to ideal switches as is possible. Consequently, modeling the behavior of the switches as a square-wave is applicable in most circumstances.

Typically the only method to remove these unwanted harmonics from the output is by using a filter. The following section will present the basic principles behind a mixer, which can significantly attenuate the harmonics created in the mixing process, and thus eliminate the need for filtering. In this mixer the harmonics are still created but they are then cancelled at the output.

4.3 Harmonic-Rejection Background

The discussion in the previous section focused on the fundamentals of switching mixers. The simplified effect of the switching mixer was to multiply the baseband signal by a square-wave version of the LO signal. This section will discuss in more detail the nature of the harmonics that result from converting a sine wave to a square wave. Furthermore, the basis for a method to attenuate these unwanted harmonics is introduced.

In a mixer, the switching transistors generate the square wave but this provides little insight into the frequency content of a square wave. Alternatively, applying a sample-and-hold operation to a sine wave can also generate a square wave. Exploring this operation in more detail provides valuable insight into the harmonics that are present in a square wave

Applying a sample-and-hold (S/H) operation to a sine wave can generate a square wave. Figure 4.6 shows such a square wave as well as the underlying sinusoid from which it was sampled. Since a square wave can be generated in this way, understanding the effects of a sample-and-hold operation on a sine wave can provide insight into the spectrum of the square wave. For the remainder of this text, the waveform generated by applying a sample-and-hold operation to a sine wave will be referred to as a sample-and-hold sine (SHS) waveform.

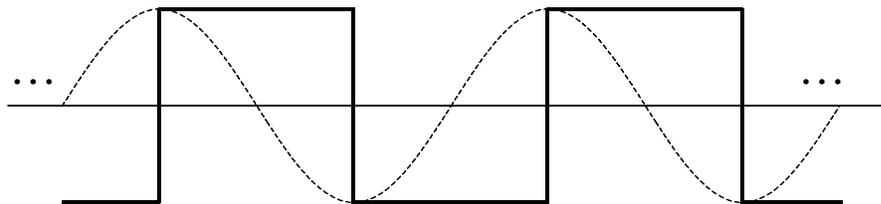


Figure 4.6 Square wave generated by applying a sample-and-hold operation to a sine wave.

The sample-and-hold operation, illustrated in Figure 4.7, can be modeled by a sampling function followed a filter with a rectangular impulse response [16]. The sampling function multiplies the input signal by an impulse train in the time domain. The hold operation then holds the signal at the sampled value until the next sample is taken. This hold operation can be modeled as a LTI filter with a rectangular impulse response.

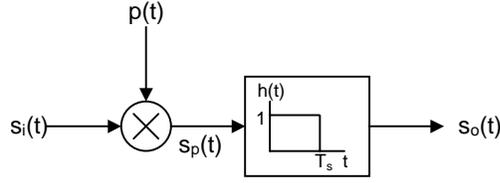


Figure 4.7 Model for sample-and-hold operation.

A frequency domain representation of the sample-and-hold operation is critical to understanding the harmonics is a square wave. The operation illustrated in Figure 4.7 leads to the following relationships between the input and output:

$$s_o(t) = (s_i(t)p(t)) * h(t). \quad (4.6)$$

The frequency domain representation of Equation (4.6) is given by

$$S_o(f) = (S_i(f) * P(f))H(f). \quad (4.7)$$

Now that the basic operation of the sample-and-hold is presented, consider the case with a sinusoidal input when $s_i(t) = \cos(2\pi f_i t)$ where f_i is the frequency of the input signal. The Fourier transform of this signal is given by

$$S_i(f) = \frac{1}{2} [\delta(f - f_i) + \delta(f + f_i)]. \quad (4.8)$$

In a sampling operation $p(t)$ is an impulse train in time defined as

$$p(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (4.9)$$

where T_s is the sampling period. The Fourier transform of $p(t)$ is given by

$$P(f) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T_s}) \quad (4.10)$$

and the sampling frequency $f_s=1/T_s$. In the frequency domain, the sampling operation corresponds to another impulse train where the separation between the impulses is equal to the frequency of the sampling operation. The output from the sampling operation, $s_p(t)$ is equal to the product of $s_i(t)$ and $p(t)$ which equals

$$s_p(t) = s_i(t)p(t) = \cos(2\pi f_i t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (4.11)$$

and this corresponds to the following function in the frequency domain:

$$\begin{aligned} S_p(f) &= S_i(f) * P(f) \\ &= \frac{1}{2} [\delta(f - f_i) + \delta(f + f_i)] * \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T_s}) \\ &= \frac{1}{2T_s} \sum_{n=-\infty}^{\infty} (\delta(f - nf_s - f_i) + \delta(f - nf_s + f_i)) \end{aligned} \quad (4.12)$$

where $f_s=1/T_s$. This shows that sampling operation convolves the spectrum of the input with an impulse train resulting in a copy of the input spectrum centered on each delta function in the impulse train.

A graphical representation of the Equations (4.11) and (4.12) is illustrated Figure 4.8 showing the generation of $s_p(t)$ and $S_p(f)$ in the case where the sampling frequency is equal to twice the frequency of the sinusoidal input making $f_s=2f_i$. Notice that a sampling rate of twice the input frequency results in harmonics at odd multiples of the input frequency in $S_p(f)$. These are created because through the convolution process, copies of $S_i(f)$ are generated and centered at each impulse in $P(f)$. Understanding the generation of $S_p(f)$ will be critical when the sampling rate is increased in subsequent sections.

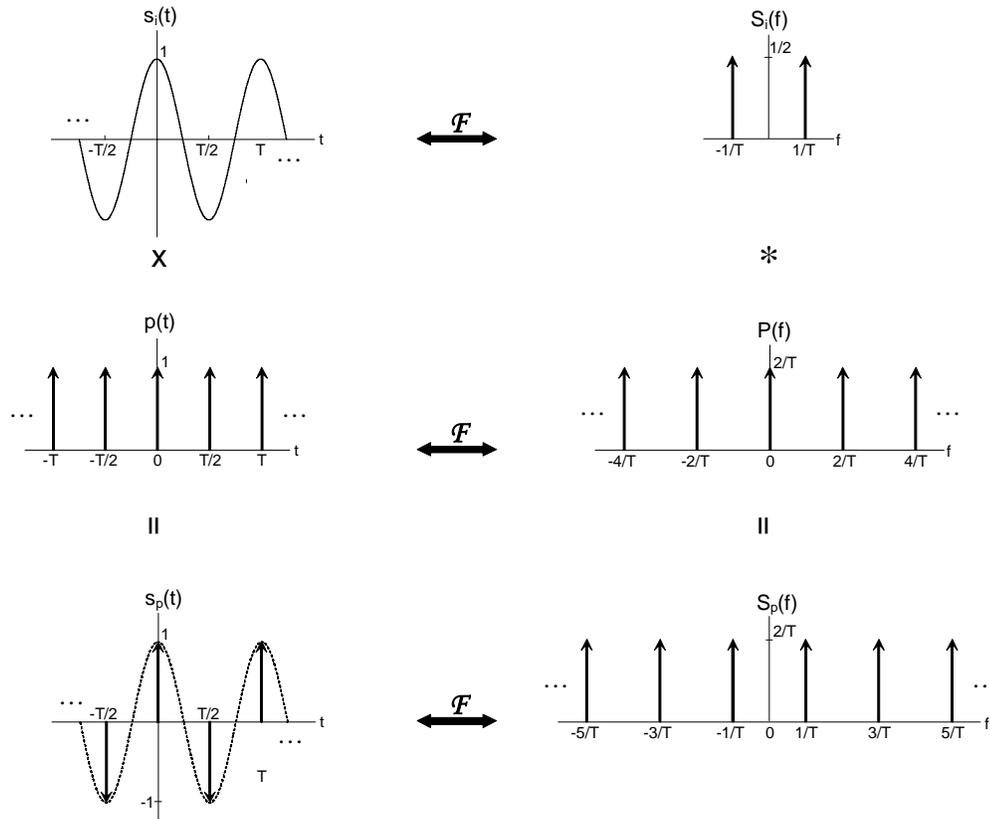


Figure 4.8 Sampling of a sinusoid in both the time and frequency domains.

After the sampling operation comes the hold operation. This operation is modeled by a filter with a rectangular impulse response that is defined as

$$h(t) = \begin{cases} 1, & 0 < t < T_s \\ 0, & \text{otherwise} \end{cases} \quad (4.13)$$

where the hold time, T_s , is equal to the sampling time. The Fourier transform of $h(t)$ is given by

$$\begin{aligned} H(f) &= \frac{\sin(\pi f T)}{\pi f} e^{-j\pi f T} \\ &= T \operatorname{sinc}(\pi f T) e^{-j\pi f T} . \end{aligned} \quad (4.14)$$

Consequently, the output of the sample-and-hold operation in the frequency domain is equal to

$$S_o(f) = \frac{1}{2} \text{sinc}(\pi f T) e^{-j\pi f T} \sum_{n=-\infty}^{\infty} (\delta(f - n f_s - f_i) + \delta(f - n f_s + f_i)). \quad (4.15)$$

The results of Equation (4.15) allow for a complete picture of the sample-and-hold operation. To complete the sample-and-hold operation, Figure 4.9 illustrates the effects of the hold operation on the sampled sinusoid from Figure 4.8. The hold operation convolves the sampled sinusoid with a square pulse resulting in a square wave. The Fourier transform of the sampled sinusoid is a uniform impulse train and the Fourier transform of the square pulse is a SINC function defined by Equation (4.14). The product of these two frequency-domain functions results in the output, $S_o(f)$ which is a series of impulses whose magnitude is shaped by the SINC function. This output agrees with the expected result in Equation (4.5).

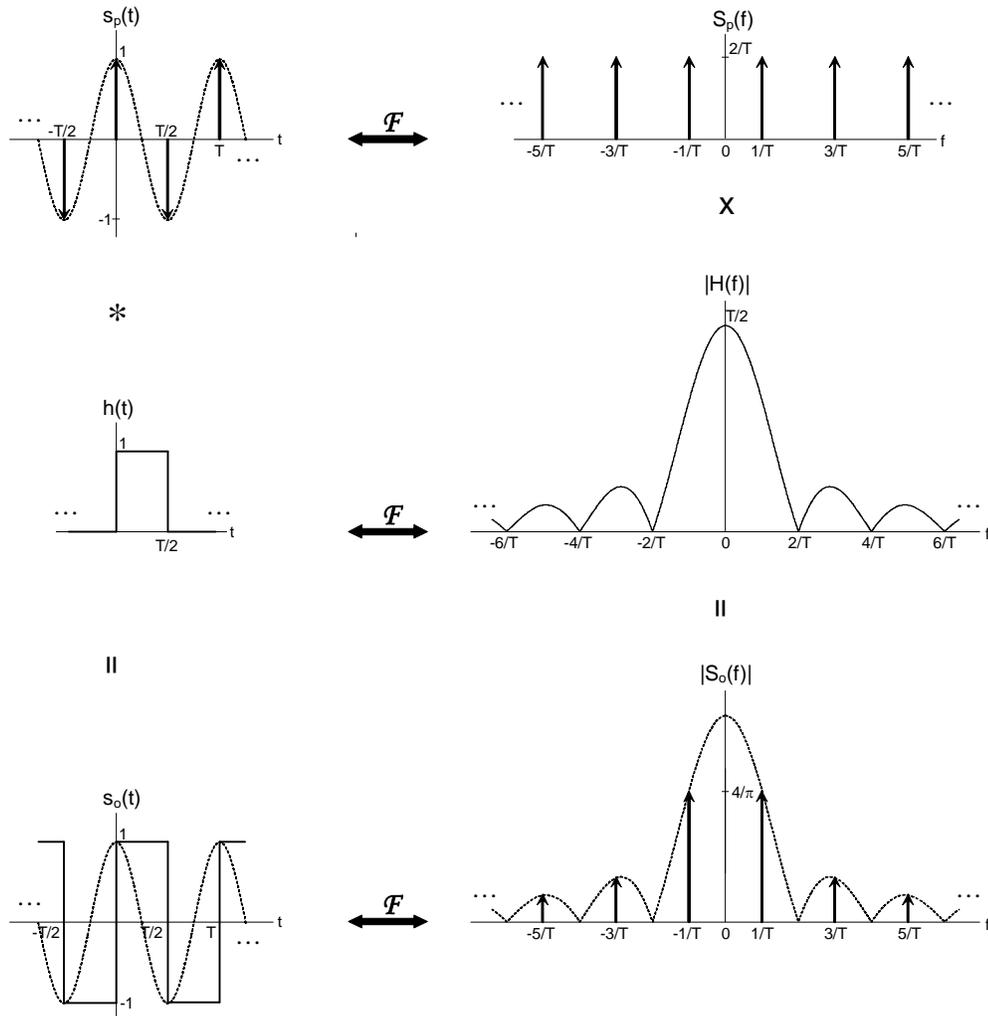


Figure 4.9 Graphical representation of the hold function on sampled sinusoid.

As was discussed in previous section, in a transmitter the harmonics of the square wave require filters that may be difficult to integrate on a CMOS substrate. Figures 4.8 and 4.9 show how a square wave is generated by applying a sample-and-hold operation to a sinusoid. It became clear that, in the frequency domain representation of a square wave, the location of the harmonics is a function of the sampling operation and the magnitude of these harmonics is then shaped by the hold operation. Consequently, to

significantly alter the filtering that is required, changing the sampling rate is promising option.

Although a switching mixer and a sample-and-hold operation both result in a square wave, the method by which this occurs is significantly different. Therefore, an in depth investigation of the sample-and-hold operation on a sinusoid may seem irrelevant to the problems of switching mixers. However, it will be shown that increasing the sampling rate of the sample-and-hold operation on a sinusoid results in a SHS waveform that contains significantly fewer harmonics than a square wave and furthermore it can be implemented with a modified switching mixer. It is for these reasons that understanding the effect of a sample-and-hold is important even though it doesn't necessarily provide an accurate model of a switching mixer.

If the sampling frequency is increased, the waveform resulting from the sample-and-hold operation on a sinusoid is changed. As the sampling frequency is increased, the impulse train given by Equation (4.10) will spread out in frequency. Specifically, the frequency difference between impulses will increase because this difference in frequency between impulses is equal to the sampling frequency. As a result, the harmonics at the output of the sample-and-hold will also spread out in frequency. These harmonics are shaped by the hold operation defined by Equation (4.14). Increasing the sampling rate dictates a decreased hold time resulting in the SINC function also spreading out in frequency. Figure 4.10 shows the output of the sample-and-hold operation, with the same sinusoidal input, in both the time and frequency domains for three different sampling frequencies: $f_s=2f_i$, $f_s=4f_i$, and $f_s=8f_i$. Figure 4.10a shows the output when the sampling rate is double the input frequency and this is taken directly from Figure 4.9.

Figure 4.10b corresponds to $f_s=4f_i$ and the unwanted odd harmonics are still present. It is interesting to note that the spectra are very similar in these two cases but in the time domain the signals are quite different. Figure 4.10c shows the case where $f_s=8f_i$ and now the spectrum has changed dramatically. The third and fifth harmonics are eliminated leaving the seventh harmonic as the closest harmonic to the fundamental frequency. This rejection of unwanted harmonics close to the fundamental illustrates the basic principles of harmonic-rejection.

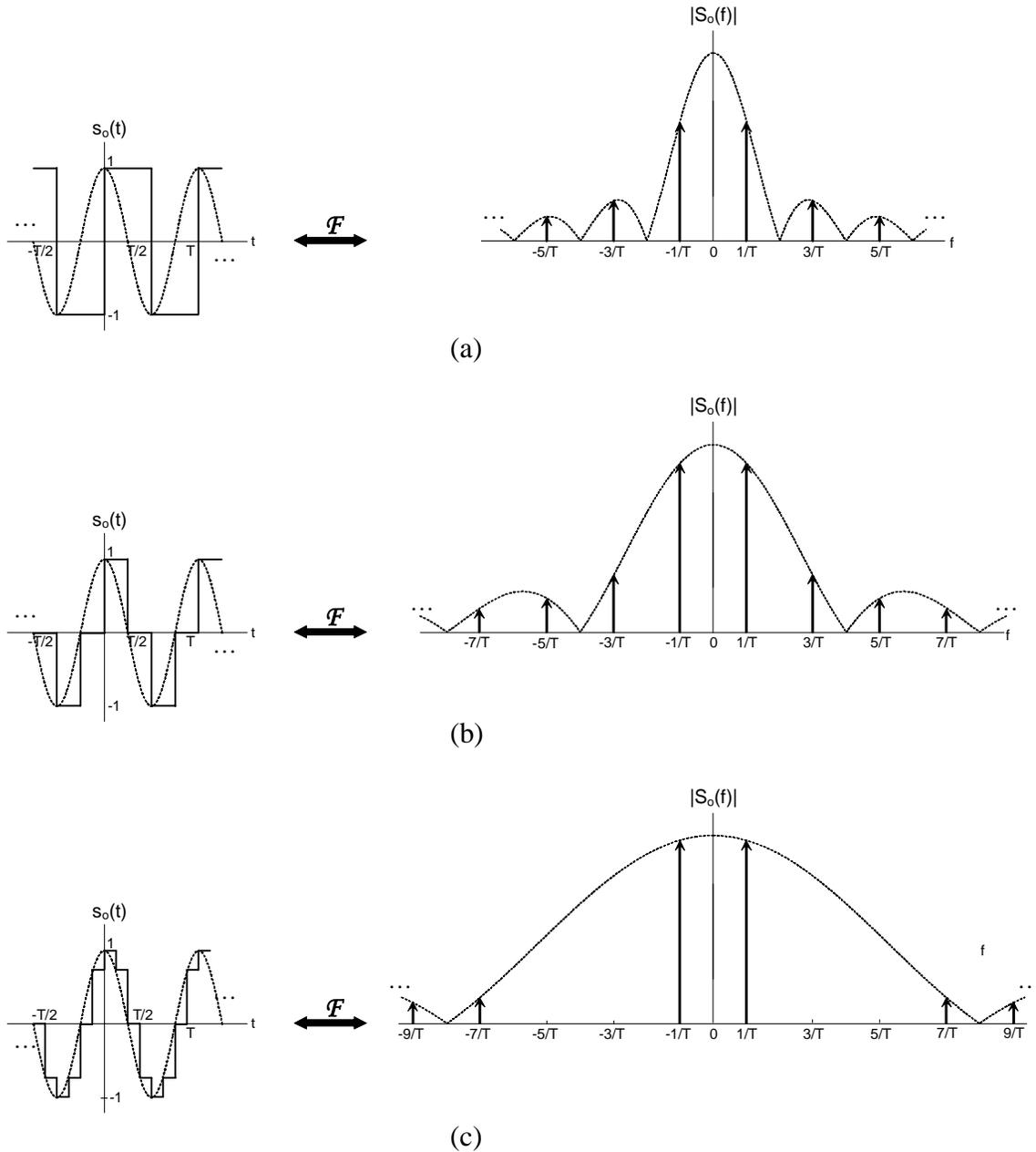


Figure 4.10 Sample and hold operation on sinusoid in time and frequency domains for three sampling frequencies. (a) $f_s = 2f_i$. (b) $f_s = 4f_i$. (c) $f_s = 8f_i$.

As the sampling rate is pushed even higher, the resulting SHS waveform will have even fewer harmonics. Intuitively this makes sense by inspecting the time-domain signal

in Figure 4.10c. A higher sampling rate in the sample-and-hold results in a waveform that has looks more sinusoidal, implying fewer harmonics.

Figure 4.11 shows the spectrum for an arbitrary sampling rate in which $f_s = Nf_i$. In the case where $N=16$ the fifteenth harmonic would be the first undesired signal. In the limit as the sampling rate is increased, the resulting waveform will be a sinusoid with no harmonics. Clearly increasing the sampling rate, N , increases the frequency separation between the fundamental tone and the closest unwanted harmonic.

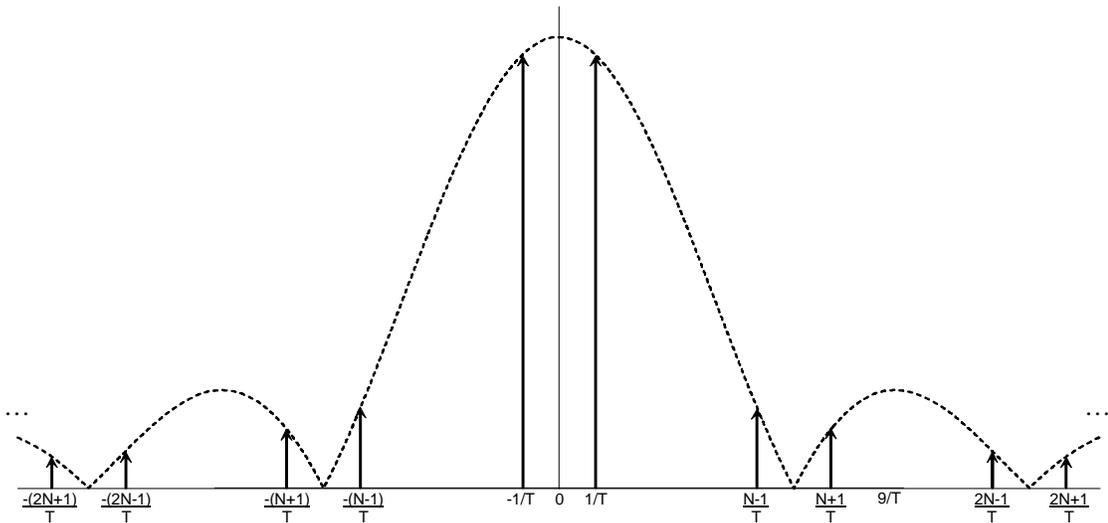


Figure 4.11 Spectrum form a sample-and-hold for with arbitrary sampling rate N .

The elimination of harmonics illustrated in Figure 4.10c shows the basic principle behind the harmonic-rejection mixer presented in the next section. If the SHS waveform from Figure 4.10c was multiplied with a baseband signal the filtering requirements would be relaxed to the point that a discrete filter might be unnecessary. In the implementation a trade-off exists between complexity and the number of harmonics that are rejected. To ease the filtering requirements, the waveform resulting from a high

sampling rate is desired. However, as can be seen by the sampled sinusoid waveform of Figure 4.10c, generating this sampled sinusoid waveform becomes increasingly complex as the sampling rate is increased. Therefore, understanding how much rejection is needed for a particular system is critical when choosing the sampling rate. Furthermore, although increasing the sampling frequency has produced a waveform that is more sinusoidal, it remains to be seen how this waveform might be implemented with a mixer. For the remainder of this text an SHS waveform with eight samples per period will become the standard SHS waveform. Although generating SHS waveforms with higher sampling rates is certainly possible, the added complexity adds little instructive value to the reader.

4.4 Harmonic-Rejection Principles

The basic principle behind the harmonic-rejection mixers (HRM) lies in the generation of a SHS waveform. Multiplying this waveform with baseband signal eases the filtering requirements when compared to the traditional switching mixer that essentially multiplies the baseband signal with a square wave. This concept is illustrated in Figure 4.12a in which a baseband signal is multiplied with a square wave LO signal, shown in both time and frequency, by a conventional switching mixer. The LO signal is effectively converted to square wave even if the LO signal is actually sinusoidal, which is a result of the current-commutating switches. The output spectrum contains fairly large unwanted signals centered on the odd harmonics of the LO signal and these unwanted signals will typically need filtering to meet the spectral emissions requirements.

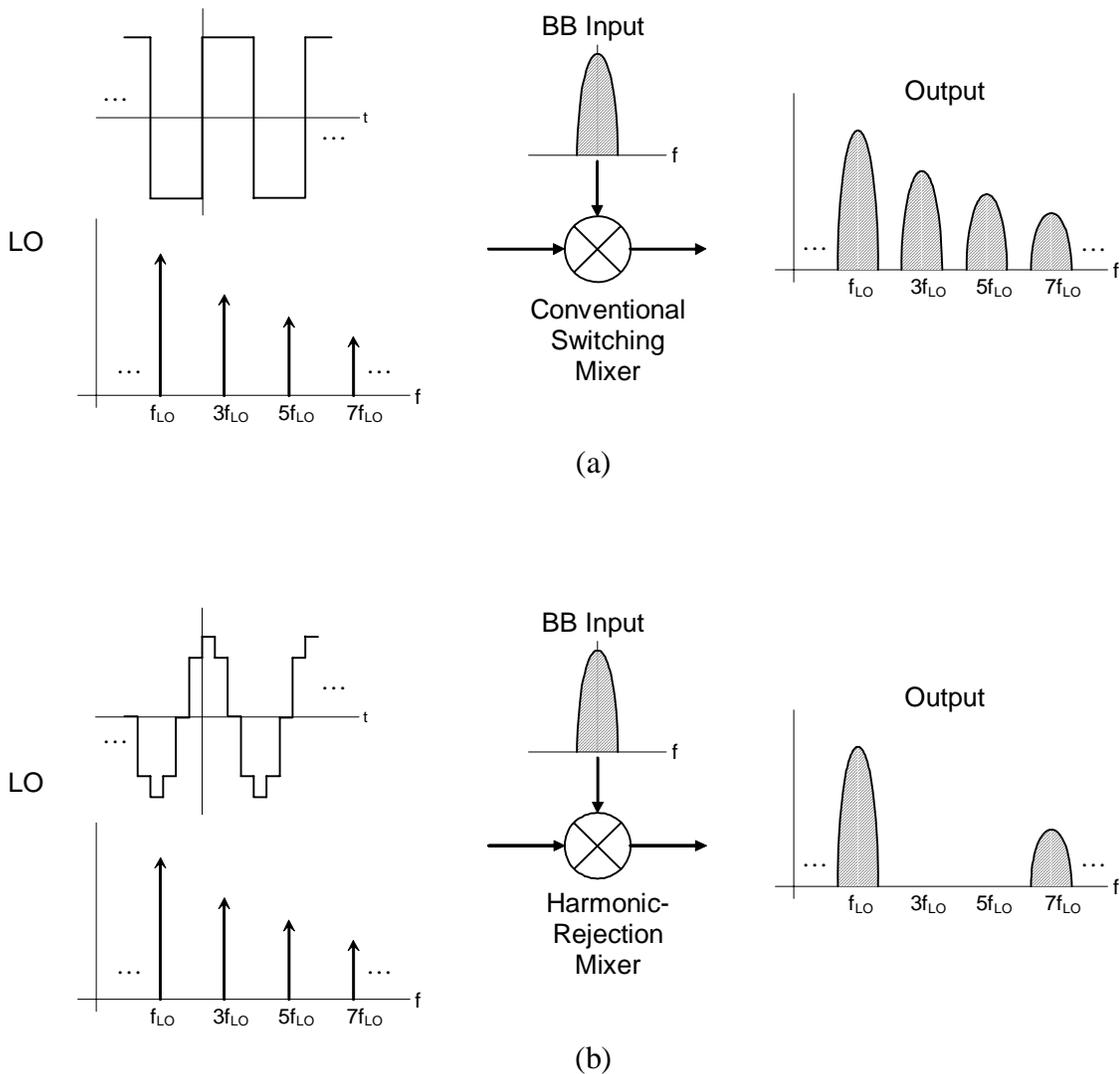


Figure 4.12 (a) Input, LO and output signals of a conventional switching mixer. (b) Input, LO and output signals of a harmonic-rejection mixer.

In contrast to the conventional mixer, Figure 4.12b shows the output when a SHS waveform is used for the LO signal instead of a square wave and a harmonic-rejection mixer is used to multiply the two signals. In this illustration the sampling rate for the SHS waveform is eight times the LO frequency. The closest unwanted signal in the

output spectrum is now centered on the seventh harmonic, which significantly eases the filtering requirements.

Although Figure 4.12 illustrates the ideal behavior of the HRM, it remains to be shown how this is actually accomplished. A critical component is the generation of the SHS waveform and it was shown that applying a sample-and-hold operation to a sinusoid could generate the desired SHS waveform. However, this waveform is not necessary applicable to switching mixers and thus a method to generate the SHS waveform is needed that does not require the use of a sample-and-hold operation and is applicable to switching mixers. After this method has been introduced, the HRM will be discussed in more detail. Specifically, the architecture of the HRM will be presented and in subsequent sections some of the limitations of the HRM will be analyzed.

As was mentioned, the SHS waveform is a key aspect of the HRM and the SHS waveform from Figure 4.10c is repeated in Figure 4.13, along with the underlying sinusoid. Applying a sample-and-hold operation to a sinusoid can generate this waveform but another method, more suited for a mixer application is also shown. A visual inspection of the waveform reveals that summing four phase shifted square waves can generate the same SHS waveform as a the sample-and-hold operation. This result is very interesting because square waves are amenable to implementation with switching mixers and therefore it might be possible to frequency translate a signal while rejecting the nearest unwanted harmonics. The four phase shifted square waves, also shown in Figure 4.13, all have the same frequency but two different amplitudes are required among the four square waves.

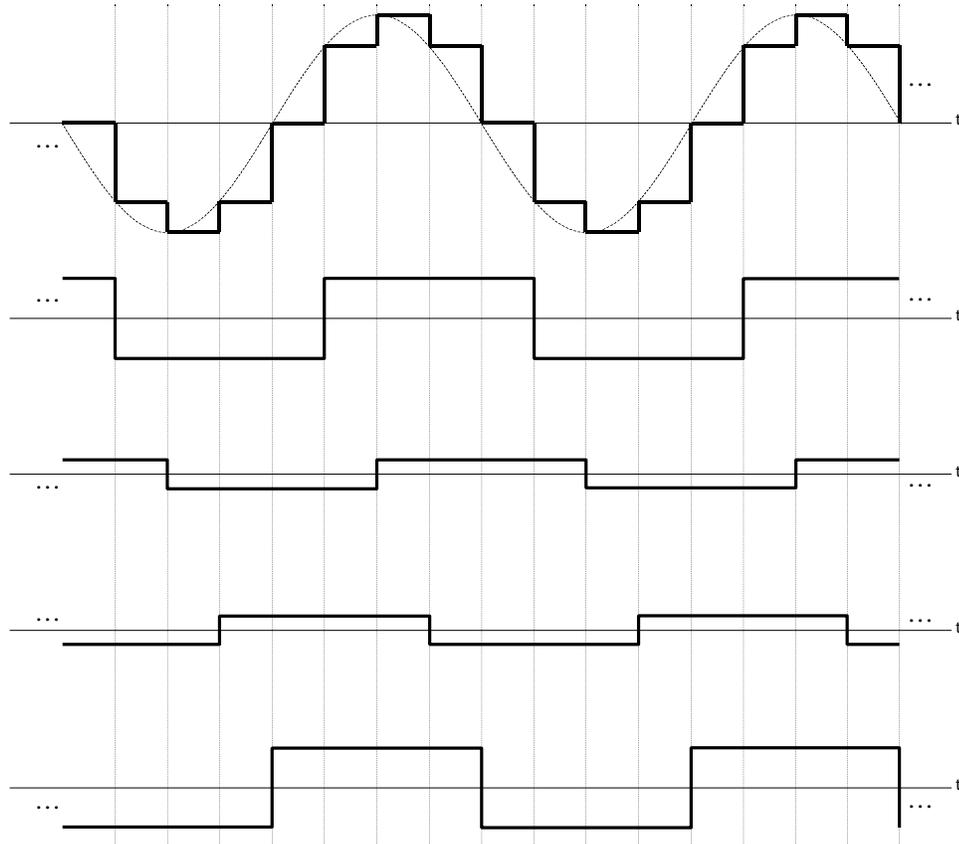


Figure 4.13 Four square waves which can be summed to generate the SHS waveform.

Applying a sample-and-hold operation, with eight samples per period, to a sinusoid can result in the waveform shown in Figure 4.10c and again in Figure 4.13. In this case, the sampling occurred at the both the positive and negative peaks of the sinusoid and equal intervals in between. Alternatively, the sampling position can be shifted with almost no effect on the spectrum of the resulting waveform. Figure 4.14 shows an example of this in which the sampling point was shifted by $-T/16$, where T is the period of the sinusoid. The resulting waveform is very similar but in can be composed of three square waves instead of four. Three square waves can be used because the samples at $t=0$ and $t=-T/8$ are equal in amplitude. This occurs at both the

negative and positive peaks and consequently, the waveform contains two fewer edges in one period leading to one less square wave. It will be shown in later sections that this reduces the hardware requirement in the implementation. Generation of the signal, $ss(t)$, which is an SHS waveform, is the key element to the rejection of unwanted harmonics in the harmonic-rejection mixer presented in the following section.

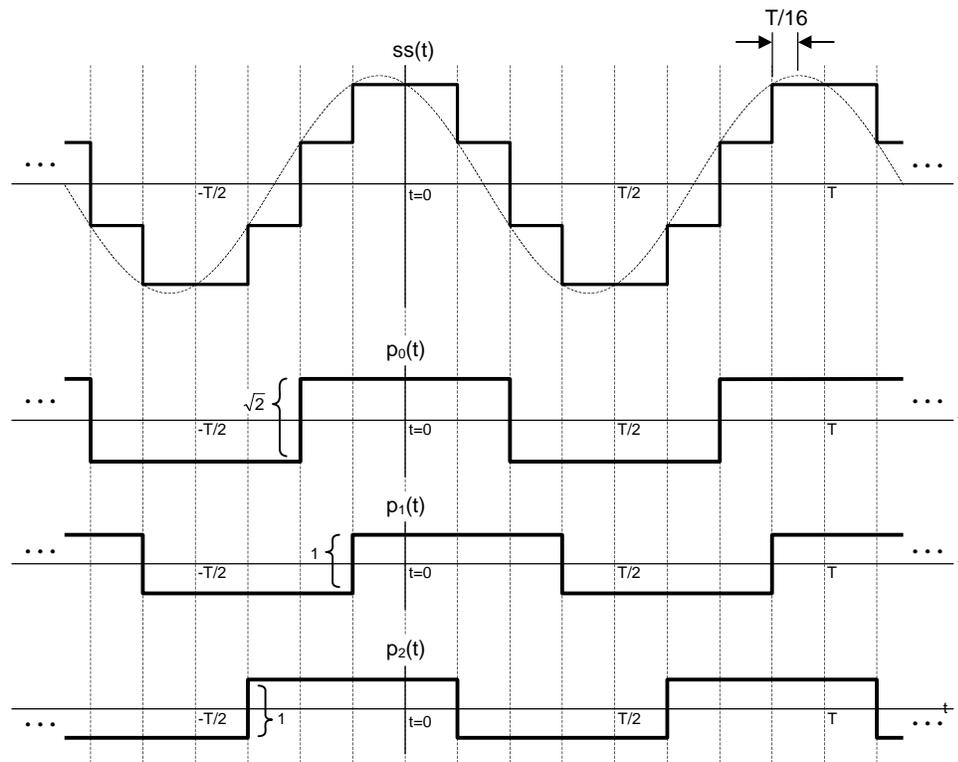


Figure 4.14 SHS waveform resulting from shifting of sampling position and the square waves which compose the SHS waveform.

The square waves illustrated in Figure 4.14 need to be scaled in amplitude and phase shifted appropriately to reject the third and fifth harmonics. Signal $p_0(t)$ is scaled by a factor of $\sqrt{2}$ compared to both $p_1(t)$ and $p_2(t)$. Furthermore, compared to $p_0(t)$, $p_1(t)$ is shifted in time by $T/8$ to the right and $p_2(t)$ is shifted by $T/8$ to the left. This time shift

corresponds to a phase shift in the fundamental term of $p_1(t)$ and $p_2(t)$ by -45 and 45 degrees respectively.

The square waves illustrated in Figure 4.14 may be represented by modified versions of Equation (4.5), which gives the time domain representation of a square wave with unit amplitude. The first square wave, $p_0(t)$, is scaled only in amplitude and is given by

$$p_0(t) = \frac{4\sqrt{2}}{\pi} \left[\cos(\omega t) - \frac{1}{3} \cos(3\omega t) + \frac{1}{5} \cos(5\omega t) - \frac{1}{7} \cos(7\omega t) \dots \right]. \quad (4.16)$$

It is interesting to notice that $p_0(t)$ is an even function and consequently the harmonic components in Equation (4.16) are all cosines.

The second square wave, $p_1(t)$, is scaled in both amplitude and shifted in time compared to $p_0(t)$ and is represented by the following expression:

$$\begin{aligned} p_1(t) &= \frac{1}{\sqrt{2}} p_0(t - T/8) \\ &= \frac{1}{\sqrt{2}} p_0(t - \pi/4\omega) \\ &= \frac{4}{\pi} \left[\cos(\omega(t - \pi/4\omega)) - \frac{1}{3} \cos(3\omega(t - \pi/4\omega)) \right. \\ &\quad \left. + \frac{1}{5} \cos(5\omega(t - \pi/4\omega)) - \frac{1}{7} \cos(7\omega(t - \pi/4\omega)) + \dots \right] \\ &= \frac{2\sqrt{2}}{\pi} \left[(\cos(\omega t) \cos(\pi/4) + \sin(\omega t) \sin(\pi/4)) - \frac{1}{3} (\cos(3\omega t) \cos(3\pi/4) \right. \\ &\quad \left. + \sin(3\omega t) \sin(3\pi/4)) + \frac{1}{5} (\cos(5\omega t) \cos(5\pi/4) + \sin(5\omega t) \sin(5\pi/4)) \right. \\ &\quad \left. - \frac{1}{7} (\cos(7\omega t) \cos(7\pi/4) + \sin(7\omega t) \sin(7\pi/4)) + \dots \right] \\ &= \frac{2\sqrt{2}}{\pi} \left[(\cos(\omega t) + \sin(\omega t)) + \frac{1}{3} (\cos(3\omega t) - \sin(3\omega t)) \right. \\ &\quad \left. - \frac{1}{5} (\cos(5\omega t) + \sin(5\omega t)) - \frac{1}{7} (\cos(7\omega t) - \sin(7\omega t)) + \dots \right]. \end{aligned} \quad (4.17)$$

Due to the time shift, $p_1(t)$ is not an even function and consequently the harmonic components are made up of both sine and cosine terms.

The final square wave, $p_2(t)$, is also shifted in time and scaled in amplitude compared to $p_0(t)$ and is represented by the following expression:

$$\begin{aligned}
 p_2(t) &= \frac{1}{\sqrt{2}} p_0(t + T/8) \\
 &= \frac{1}{\sqrt{2}} p_0(t + \pi/4\omega) \\
 &= \frac{4}{\pi} \left[\cos(\omega(t + \pi/4\omega)) - \frac{1}{3} \cos(3\omega(t + \pi/4\omega)) \right. \\
 &\quad \left. + \frac{1}{5} \cos(5\omega(t + \pi/4\omega)) - \frac{1}{7} \cos(7\omega(t + \pi/4\omega)) + \dots \right] \\
 &= \frac{2\sqrt{2}}{\pi} \left[(\cos(\omega t) \cos(\pi/4) - \sin(\omega t) \sin(\pi/4)) - \frac{1}{3} (\cos(3\omega t) \cos(3\pi/4) \right. \\
 &\quad \left. - \sin(3\omega t) \sin(3\pi/4)) + \frac{1}{5} (\cos(5\omega t) \cos(5\pi/4) - \sin(5\omega t) \sin(5\pi/4)) \right. \\
 &\quad \left. - \frac{1}{7} (\cos(7\omega t) \cos(7\pi/4) - \sin(7\omega t) \sin(7\pi/4)) + \dots \right] \\
 &= \frac{2\sqrt{2}}{\pi} \left[(\cos(\omega t) - \sin(\omega t)) + \frac{1}{3} (\cos(3\omega t) + \sin(3\omega t)) \right. \\
 &\quad \left. - \frac{1}{5} (\cos(5\omega t) - \sin(5\omega t)) - \frac{1}{7} (\cos(7\omega t) + \sin(7\omega t)) + \dots \right].
 \end{aligned} \tag{4.18}$$

Like $p_1(t)$, $p_2(t)$ is also not an even function and consequently contains both sine and cosine terms.

The sum of the three square waves leads to the desired waveform $ss(t)$, which given by

$$\begin{aligned}
 ss(t) &= p_0(t) + p_1(t) + p_2(t) \\
 &= \frac{8\sqrt{2}}{\pi} [\cos(\omega t) - \frac{1}{7} \cos(7\omega t) + \dots].
 \end{aligned} \tag{4.19}$$

To achieve the desired cancellation of the third and fifth harmonics, notice that the sine terms in $p_1(t)$ and $p_2(t)$ are equal amplitude but opposite sign and $p_0(t)$ has no sine terms.

Therefore in $ss(t)$ the sine terms sum to zero. Similarly, the cosine terms for both the third and fifth harmonics in $p_1(t)$ and $p_2(t)$ are half the amplitude and opposite sign of the corresponding harmonic terms in $p_0(t)$. As a result, in $ss(t)$ the third and fifth cosine terms also sum to zero. Therefore the summation of the three square waves also leads to a zero sum for all of the third and fifth harmonic terms and the closest unwanted signals in $ss(t)$ are the seventh and ninth harmonics.

It has been shown that by summing square waves, a SHS waveform can be generated that has significantly fewer harmonics when compared to a single square wave. However, it remains to be seen how such a waveform might be used in a mixer application. The next section will present a modified switching mixer that essentially multiplies a baseband signal with an LO signal which is shaped like $ss(t)$.

4.5 Harmonic-Rejection Mixer

The current-commutating switching mixer, introduced in Section 4.2, is widely used in wireless communications systems. This active mixer is readily implemented on a CMOS substrate and, unlike a passive mixer, can provide conversion gain. Furthermore, high isolation between the LO port and the output is possible leading to a clean output signal. Another advantage of this mixer is that the linearity is largely determined by the transconductance stage for the baseband input. Because of this, relatively well known linearization techniques can be applied. To minimize noise from the mixer, it is best if that the switching pair behave as close to ideal switches as possible, which reduces the noise contribution from these devices. Fortunately this typically does not require an

excessively large voltage swing but the switches do create other problems. As was mentioned earlier, the switching behavior introduces harmonics of the LO signal at the output which require filtering that may not be amenable to a CMOS implementation. This section presents a modified current-commutating switching mixer that reduces the filtering requirement.

4.5.1 Applicability to Switching Mixers

The previous section illustrated that a waveform generated by summing phase shifted, and amplitude scaled, square waves contains significantly fewer harmonics than a simple square wave. This was shown in Figure 4.14 and to generate the desired waveform, $ss(t)$, required summing $p_0(t)$, $p_1(t)$ and $p_2(t)$. The goal now is to apply this concept to a switching mixer to reduce the harmonics at the mixer output. However, in a typical switching mixer, the input signal is essentially multiplied by a square wave. If the input signal was somehow multiplied by $ss(t)$, instead of the typical square wave, the following output signal would result:

$$\begin{aligned} S_{out}(t) &= ss(t)S_{in}(t) \\ &= (p_0(t) + p_1(t) + p_2(t))S_{in}(t). \end{aligned} \tag{4.20}$$

If implemented directly, the summed waveform would need to be generated first and then multiplied by the input signal. This is not well suited for a switching-mixer implementation because a switching mixer multiplies an input signal by a square wave and in this case the input signal is multiplied by a SHS waveform. Alternatively, $S_{out}(t)$,

can also be expressed in terms of time shifted and scaled version of $p_0(t)$. This substitution results in

$$\begin{aligned} S_{out}(t) &= \left(p_0(t) + \frac{1}{\sqrt{2}} p_0(t - T/8) + \frac{1}{\sqrt{2}} p_0(t + T/8) \right) S_{in}(t) \\ &= S_{in}(t) p_0(t) + \frac{1}{\sqrt{2}} S_{in}(t) p_0(t - T/8) + \frac{1}{\sqrt{2}} S_{in}(t) p_0(t + T/8) \end{aligned} \quad (4.21)$$

and scaling both sides by $\sqrt{2}$ gives

$$\sqrt{2} S_{out}(t) = \sqrt{2} S_{in}(t) p_0(t) + S_{in}(t) p_0(t - T/8) + S_{in}(t) p_0(t + T/8). \quad (4.22)$$

Implementing Equation (4.20) is not easily accomplished with a switching mixer because it requires multiplying the input signal with an SHS waveform. In contrast, inspection of Equation (4.22) reveals that the output consists of a sum of three terms, each of which is the product of a square wave and the input signal. This implies that by summing the output of three mixers, the rejection of unwanted harmonics may be possible with the use of switching mixers. However, Equation (4.22) also indicates that scaling the output of each mixer is necessary before the outputs of the three mixers are summed. Specifically, the first term in Equation (4.22) is scaled by a factor of $\sqrt{2}$ compared to the second and third terms. To understand how this is possible in a switching mixer, examining the input-output relationship of a switching mixer is helpful. This will be useful in determining if scaling the input signal, the mixer gain, or the LO signal will properly scale the output.

4.5.2 Current Commutating Mixer Gain

An ideal current-commutating switching mixer is illustrated in Figure 4.15a. The mixer consists of two current sources whose values are controlled by differential input voltage signals. The switches, which enable the current commutation, are controlled by the LO signal and finally a differential output voltage is generated by a resistive load. When the switches are in the position shown in Figure 4.15a, the current source controlled by the positive input is connected to the positive output and the current source controlled by the negative input is connected to the negative output. In this case the mixer operates like a differential amplifier with a differential voltage gain equal to $-g_m R_L$. When the LO signal changes the position of the switches to the second configuration the differential voltage gain changes to $g_m R_L$. Consequently, the gain is constantly switching between $-g_m R_L$ and $g_m R_L$ as illustrated in Figure 4.15b. As a result of the changing gain, the output voltage also is switched leading to the mixing action at the output.

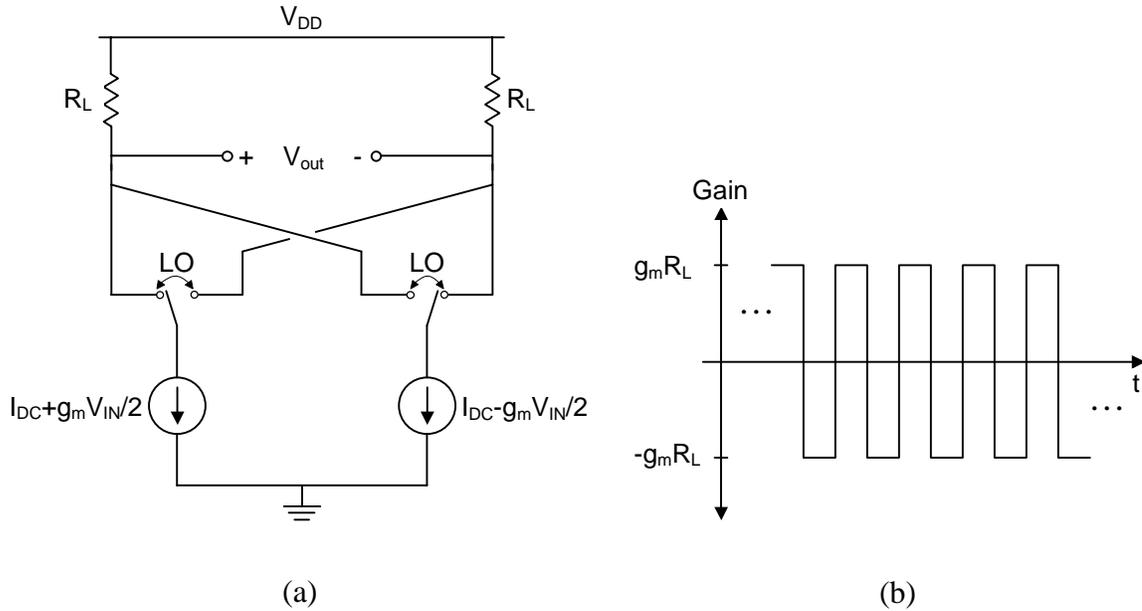


Figure 4.15 Simplified current-commutating switching mixer.

The input-output relationship of this ideal mixer can now be derived given that the gain is understood. Since the gain changes sign with each cycle of the LO signal, the output is given by the product of $g_m R_L$ and a unity amplitude square wave. This results in the following input-output relationship:

$$V_{out}(t) = V_{in}(t) g_m R_L \frac{4}{\pi} \left(\cos(\omega_{LO} t) - \frac{1}{3} \cos(3\omega_{LO} t) + \frac{1}{5} \cos(5\omega_{LO} t) \dots \right). \quad (4.23)$$

Therefore, the gain of the ideal switching mixer is $g_m R_L 4/\pi$ but harmonics are also present. If the input signal is sinusoidal, the conversion gain can be determined. Given that $V_{in}(t) = \cos(\omega_{BB} t)$ the output is given by

$$\begin{aligned} V_{out}(t) &= g_m R_L \frac{4}{\pi} \left(\cos(\omega_{BB} t) \cos(\omega_{LO} t) + \dots \right) \\ &= g_m R_L \frac{2}{\pi} \left(\cos(\omega_{LO} + \omega_{BB}) t + \cos(\omega_{LO} - \omega_{BB}) t + \dots \right). \end{aligned} \quad (4.24)$$

From this equation it is clear that the conversion gain is equal to $g_m R_L 2/\pi$.

The previous analysis was performed for an ideal switching mixer in which the switching occurs instantaneously. This results in the baseband signal being multiplied by a square wave created by the LO. However, in a practical implementation, the switching will not be instantaneous and thus the baseband signal will be multiplied by a waveform that is similar to a square wave except the transitions are non-zero. In this situation the conversion gain of a current-commutating switching mixer is given in [17]. The input-output relationship, showing only the fundamental term of the LO, is repeated here and is given by

$$V_o(t) = V_{in}(t) \frac{g_m R_L \sqrt{2} V_{LO}}{\pi (V_{GS} - V_t)_{sw}} \sin\left(\frac{\sqrt{2}(V_{GS} - V_t)_{sw}}{V_{LO}}\right) \cos(\omega_{LO} t). \quad (4.25)$$

Equation (4.23) indicates that the output signal is proportional to input signal, the g_m of the input devices and the value of the load resistance and this same conclusion can be drawn from Equation (4.25). It is worth noting that the output signal is also a function of the gate-source bias voltage on the switches.

4.5.3 HRM Schematic and Design Issues

Now that the input-output relationship of the switching mixer is understood, it is evident by inspection of Equation (4.25) that the scaling required in Equation (4.22), to reject the unwanted harmonics, can be accomplished by in one of three ways: scaling the magnitude of the input signal, scaling the value of the load resistor, or scaling g_m . Because the LO signal only switches the current back and forth, the output is not

proportional to the amplitude of the LO signal and thus the LO signal cannot be used to for scaling purposes.

It has been shown in Equation (4.22) that effectively multiplying an SHS waveform with an input signal is possible by summing the output of multiple mixers. In this case three mixers are necessary because three square waves are needed to generate the correct SHS waveform. The three mixers all receive the same input but the LO signals are shifted in phase with respect to each other. However, to achieve the proper rejection, the scaling the amplitude of the output of the three mixers is necessary. To accomplish this scaling, inspection of the input-output relationship of a CMOS Gilbert cell reveals that the scaling could be accomplished by either scaling the load resistor, the magnitude of the input or the g_m of the transconductance stage. These ideas are incorporated in the harmonic-rejection mixer and a simplified circuit diagram of the HRM is illustrated in Figure 4.16. The HRM consist of three current-commutating switching mixers, termed sub-mixers, connected in parallel to a common load. Each sub-mixer receives the identical input signal but the switches for each sub-mixer are driven three phase-shifted LO signals: LO_{-45} , LO_0 and LO_{45} . As the names imply, LO_{-45} is phase shifted by negative forty-five degrees relative to LO_0 and LO_{45} is phase shifted by positive forty-five degrees. The output current from each sub-mixer is summed in a resistive load.

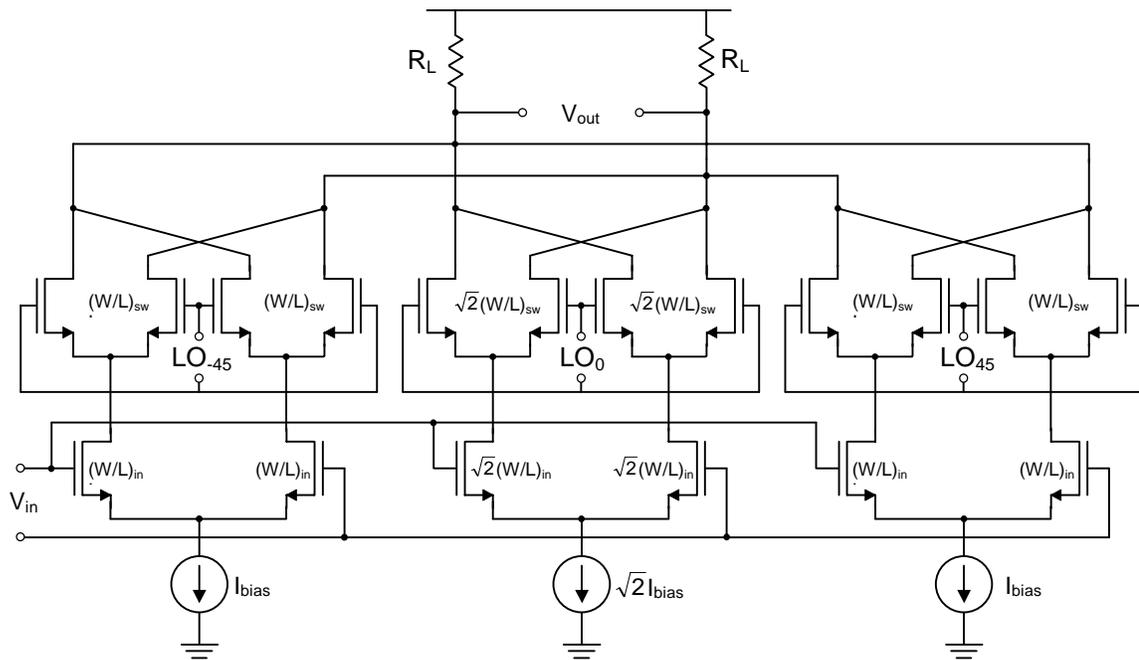


Figure 4.16 Simplified circuit diagram of the harmonic-rejection mixer.

The scaling potentially could have been achieved by scaling the load of the three mixers. However, in this case the current from each mixer is summed and an output voltage is generated through a single shared load. Therefore, scaling the load is not possible.

Another potential method to scale the output would have been scaling the magnitude of the input signal. This would have required an additional amplifier thus adding to the overall complexity.

The scaling was accomplished by adjusting the g_m of the input device. Recall from Equation (4.22) that $\sqrt{2}$ scaling is needed on one sub-mixer. This is accomplished in the HRM by scaling the bias current source and the W/L ratio of the input devices. To

understand how this properly scales the g_m , a better understanding of the transconductance of a MOS device is needed.

The drain current of a long-channel MOS device in saturation is given by

$$I_D = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_t)^2 \quad (4.26)$$

and thus g_m is equal to

$$g_m \equiv \frac{\partial I_D}{\partial V_{GS}} = \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_t). \quad (4.27)$$

Consequently, the value of g_m is directly proportional to the W/L ratio of a transistor and it can be scaled appropriately, given that the value of V_{GS} is constant. It should be noted that the g_m is also proportional to $(V_{GS} - V_t)$. From Equation (4.26) it is evident that to maintain a constant $(V_{GS} - V_t)$ voltage while scaling W/L , requires that the drain current is also scaled. As a result, the W/L ratio of the input devices in the HRM has been scaled by $\sqrt{2}$ to scale the g_m and the current source for this sub-mixer has also been scaled by $\sqrt{2}$ to maintain a constant gate-source voltage.

Alternatively, the scaling of the g_m could have been accomplished without scaling the current in the sub-mixer. An alternative equation for g_m of saturated long-channel device is given by

$$g_m = \sqrt{2 \mu_n C_{ox} \frac{W}{L} I_D}. \quad (4.28)$$

Doubling the W/L ratio while maintaining a constant drain current would have achieved required $\sqrt{2}$ scaling. However, this relationship is only valid for long channel devices

and in practice, shorter channel length transistors may be used, which will not have the same relationship between g_m and W/L .

A more general model for the drain current, which includes the effect of velocity saturation and is thus more appropriate for shorter channel lengths, is given by

$$I_D = \frac{1}{2} \mu_n C_{ox} W E_c (V_{GS} - V_t) \left(1 + \frac{L E_c}{V_{GS} - V_t} \right)^{-1} \quad (4.29)$$

where E_c is the critical electric field strength defined as the point where the carrier velocity has been reduced by a factor of two compared to the low-field situation. For a short-channel device, the value of L is small and thus the term $L E_c$ is also small compared to $(V_{GS} - V_t)$. In this case the drain current expression reduces to

$$I_D = \frac{1}{2} \mu_n C_{ox} W E_c (V_{GS} - V_t). \quad (4.30)$$

The drain current for a short-channel device does not follow the standard square-law behavior of a long-channel device and the current is independent of the channel length.

The g_m for a short-channel device is therefore given by the following expression:

$$g_m \equiv \frac{\partial I_D}{\partial V_{GS}} = \frac{1}{2} \mu_n C_{ox} W. \quad (4.31)$$

Clearly, for short channel devices the only way to scale the g_m is to scale the width of the transistor. Comparing Equation (4.31) to the g_m of a long-channel device given by Equations (4.27) and (4.28), reveals that the only method that works for both long and short channel devices is to scale the both width of the transistor and the bias current proportionally. Leaving the bias current fixed and scaling only scaling the W/L , as might

be inferred from Equation (4.28), relies heavily on the square-law relationship of a long channel device, which is not an good solution for shorter channel lengths.

If a device is behaving in the transition area between long and short channel behavior, it is evident from Equation (4.29) that scaling the width will remain an effective solution. If the drain current and the transistor width are scaled identically, the value of gate-source voltage will remain unchanged according to equation (4.29). As a result, even if the g_m is dependant on V_{GS} , as is the case when a device operating in the long-channel regime or the transition area between long and short channel behavior, the g_m will still scale properly.

From the previous discussion the reasons for scaling the current sources and the W/L ratio of the input devices should be clear. However, as can be seen in Figure 4.16, the size of the switches in the HRM was also scaled. This was done for two reasons. First, recall from Equation (4.25) that the output signal is a function of the gate-source bias voltage on the switches. Therefore, by scaling the W/L of the switches, $(V_{GS}-V_t)_{sw}$ does not change when the bias current is scaled and thus the gate-source bias voltage is equal for all of the switching transistor in the three sub-mixers. Therefore, the gain of each sub-mixer depends only on the g_m scaling of the input devices and better gain matching is achieved.

The second reason for scaling the switch size relates to the switching characteristics of the differential pair formed by the switching transistors. For square-law devices, the current is completely switched in a differential pair when the differential control voltage is greater than $\sqrt{2}(V_{GS} - V_t)_{sw}$ [18]. As was previously mentioned, this

causes a finite transition time in the waveform created by the switches. The transition time for the switching transistor in a CMOS Gilbert cell is given in [17] by

$$t_{tr} = \frac{2\sqrt{2}(V_{GS} - V_t)_{sw}}{V_{LO}\omega_{LO}} \quad (4.32)$$

where V_{LO} and ω_{LO} are the peak LO voltage and the angular frequency of the LO respectively. From this equation it is clear that the transition time is directly proportional to the gate-source bias voltage. As a result, if the gate-source bias voltages are not equal on the switches of the three sub-mixers, the transition times will also be unequal. Therefore the waveforms created by the switching transistors will have a different shape and the level of harmonic rejection will diminish. Therefore, by maintaining a constant gate-source bias voltage on all of the switches in all three sub-mixers, the matching between the sub-mixers is better and thus the harmonic-rejection is improved.

Another benefit of maintaining a constant gate-source bias voltage on the input transistors relates to the linearity of the input differential pairs. The HRM contains three input differential pairs that convert the input voltage to a current. Ideally, the distortion from each differential pair would be the same and thus they would contribute equally to the overall distortion. The distortion is evident in the differential current generated by the input differential pair. For long channel devices, the differential output current from a differential pair is given by [19]

$$I_O = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} V_{in} \sqrt{\frac{4I_{bias}}{\mu_n C_{ox} \frac{W}{L}} - V_{in}^2}. \quad (4.33)$$

where I_{bias} is the current bias current drawn by the current source. With zero input, the current in each transistor of the differential will be half of I_{bias} . If a new parameters K_1 is defined as

$$K_1 = \frac{4I_{bias}}{\mu_n C_{ox} \frac{W}{L}} \quad (4.34)$$

then the output current will be given by

$$I_o = \sqrt{\mu_n C_{ox} \frac{W}{L} I_{bias} V_{in} \sqrt{1 - \frac{V_{in}^2}{K_1}}} \quad (4.35)$$

Defining another parameter K_2 as

$$K_2 = \sqrt{\mu_n C_{ox} \frac{W}{L} I_{bias}} \quad (4.36)$$

further simplifies the output current to

$$I_o = K_2 V_{in} \sqrt{1 - \frac{V_{in}^2}{K_1}} \quad (4.37)$$

Equation (4.37) can be expanded into the following power series

$$\begin{aligned} I_o &= K_2 V_{in} \left(1 - \frac{1}{2} \frac{V_{in}^2}{K_1} - \frac{1}{8} \frac{V_{in}^4}{K_1^2} + \dots \right) \\ &= K_2 V_{in} - \frac{K_2}{2K_1} V_{in}^3 - \frac{K_2}{8K_1^2} V_{in}^5 + \dots \end{aligned} \quad (4.38)$$

The coefficient of the linear term of Equation (4.38) is K_2 and this represents the small-signal differential transconductance of the differential pair. This result agrees with the transconductance of a single device given by Equation (4.28). The higher order

terms in Equation (4.38) give rise to harmonic distortion and the fractional third harmonic distortion or HD_3 can be determined by applying Equation (4.38) to Equation (2.23). Recall that with a single sinusoidal input the fractional harmonic distortion represents the ratio of amplitude of the third harmonic to the amplitude of the fundamental and this is given by [19]

$$\begin{aligned}
 HD_3 &= -\frac{\frac{K_2}{4} \frac{2K_1}{K_2} V_{in}^2}{8 K_1} \\
 &= -\frac{1}{8} \frac{V_{in}^2}{K_1} \\
 &= -\frac{1}{32} \frac{V_{in}^2}{(V_{GS} - V_t)^2}.
 \end{aligned} \tag{4.39}$$

The results given by Equation (4.39) indicate that the harmonic distortion is a function of the amplitude of the input signal and the gate-source bias voltage. The earlier discussion showed that maintaining a constant V_{GS} allowed for the scaling of the g_m to be controlled only by the width of the input transistors. The three differential input pairs in the HRM are all driven by the same input voltage and thus, it is clear from Equation (4.39) that keeping the value of V_{GS} equal for all of the input devices also indicates that the harmonic distortion caused by these devices is also equal. This is the desired condition because in this situation each of the sub-mixers contributes equally to the distortion of the output.

4.5.4 Gain of the HRM

The HRM essentially consists of three mixers in parallel with each mixer receiving the identical input signal. The overall gain of the HRM is therefore the sum of gains of each of the three sub-mixers. However, because the phase of the LO signal applied to each sub-mixer is different compared with the other sub-mixers, the contribution to the desired output signal will also vary. This is evident by examining the fundamental components of waveforms $p_0(t)$, $p_1(t)$ and $p_2(t)$. Although both $p_1(t)$ and $p_2(t)$ both contain a sine term in their respective fundamentals, when these three waveforms are summed in Equation (4.19), the summation only contains a cosine fundamental. Therefore, only the cosine terms in $p_1(t)$ and $p_2(t)$ contribute to the overall gain of the HRM. In contrast, $p_0(t)$ contains only a cosine term in its fundamental and thus contributes all of the energy from its fundamental is transferred to the output. Therefore, the effective gain or the gain contribution to the output from a particular sub-mixer is significantly less when the phase of the LO does not match the phase of the output.

To quantify the effects of LO phase differences on the gain of the complete HRM a metric needs to be established. For this metric, the gain of a single current-commutating mixer is defined as A_m . Therefore for this single mixer the output is given by

$$S_o = A_m S_i. \quad (4.40)$$

It is assumed in this case that the LO and mixer output are in phase with each other. In the case of the HRM the gain of each sub-mixer relative to the metric A_m will be

examined. The sub-mixer which receives the LO_0 signal with zero phase will have a gain contribution of A_m because in this case the LO and the output are in phase.

The gain contribution of the sub-mixers driven by LO_{45} and LO_{-45} is reduced compared to the first sub-mixer for two reasons. First, the g_m of these sub-mixers has been scaled by a factor of $1/\sqrt{2}$ compared with the other sub-mixer, which reduces the gain contribution of each sub-mixer by this factor. The second reason stems from the phase relationship between the LO and the output. It is evident from Equation (4.17) that the gain contribution from the sub-mixer driven by LO_{45} is reduced by a factor of $\cos(\pi/4)$, which equals $1/\sqrt{2}$. Similarly, from Equation (4.18) it is clear that the gain contribution of the sub-mixer driven by LO_{-45} is reduced by a factor of $\sin(\pi/4)$, which is also $1/\sqrt{2}$. This gain reduction stems from the 45-degree phase difference between the LO and the output. Therefore, combining the gain reduction from the g_m scaling with the gain reduction from the phase difference results in a gain contribution that is reduced by a factor of two compared with the gain contribution of the sub-mixer driven by the LO that is in phase with the output.

The total gain of the HRM can now be established relative to the gain a single current commutating mixer, whose gain is defined as A_m . The gain of the HRM is the sum of the individual gain contributions from the three sub-mixers and is given

$$\begin{aligned}
 A_{HRM} &= A_{sub-mixer1} + A_{sub-mixer2} + A_{sub-mixer3} \\
 &= A_m + \frac{A_m}{2} + \frac{A_m}{2} \\
 &= 2A_m.
 \end{aligned} \tag{4.41}$$

Thus, the gain of the HRM is double the gain of a single mixer. Applying the results of Equation (4.41) to Equation (4.25) readily yields a complete expression for the gain of the HRM. It is worth noting that the conversion gain of the HRM is also double the conversion gain of a single mixer.

4.6 Matching in HRM

The ability of the HRM to attenuate the unwanted harmonics created in the mixing process depends on a number of factors, but two of the most critical are the phase matching of the LO signals and the gain matching of the sub-mixers. Depending on the method that is used to generate the multiple phases, phase mismatch generally stems from device mismatch. Similarly, the gain mismatch of the sub-mixers is dominated by the device mismatch between the input differential pairs and the current sources.

To model both types of mismatch, Figure 4.17 illustrates the summation of the three square waves needed to generate an SHS waveform. One of the square waves, $p_o(t)$, contains both an amplitude mismatch term, α , and a phase mismatch term, θ . The spectrum of the resulting waveform contains the expected fundamental term and seventh harmonic, but it also contains unwanted the third and fifth harmonics. This spectrum is then overlaid on the spectrum of square wave whose fundamental is of equal magnitude compared with the SHS waveform with mismatch. The harmonic components of the square wave are shown in gray dashed lines while they spectrum from the SHS waveform is in solid black. Note that the fundamental and seventh harmonic of the square wave are not visible because they are equal to the corresponding terms in the SHS waveform. The

amount by which the third and fifth harmonics are suppressed, as compared with the harmonics from the single square wave, is defined as the third harmonic rejection or HR_3 , and the fifth harmonic rejection or HR_5 . This is the relevant figure of merit because in typical switching mixer the input is multiplied by a square wave but in this case the input is multiplied by the SHS waveform. Therefore comparing the output of the HRM to square wave represents the relative improvement in the suppression of the harmonics.

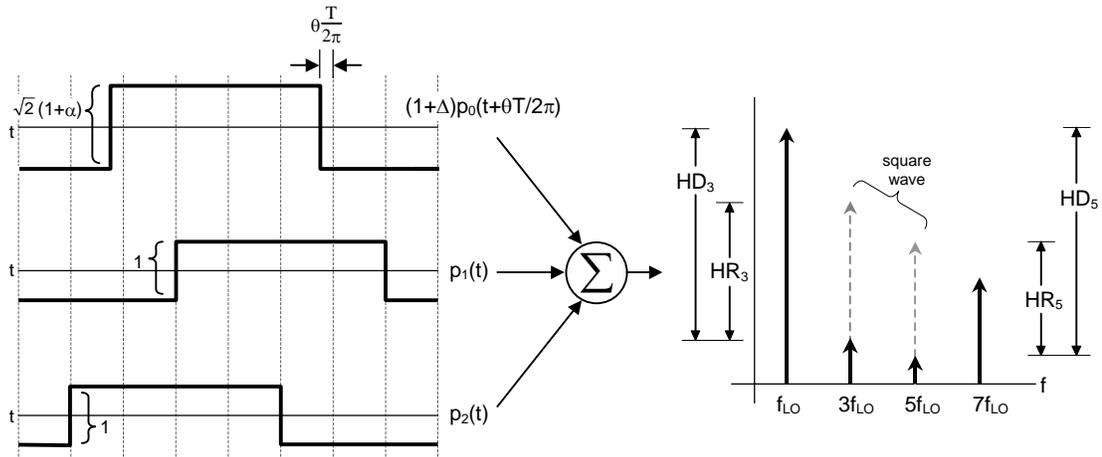


Figure 4.17 Graphical representation of HR_3 and HR_5 .

Based on the harmonics of a square wave from Equation (4.5), the relationship between HD_3 and HR_3 in dB is given by

$$\begin{aligned} HD_3 &= HR_3 + 20 \log\left(\frac{1}{3}\right) \\ &= HR_3 + 9.5 \text{ dB} \end{aligned} \quad (4.42)$$

and similarly

$$\begin{aligned} HD_5 &= HR_5 + 20 \log\left(\frac{1}{5}\right) \\ &= HR_5 + 14 \text{ dB}. \end{aligned} \quad (4.43)$$

To quantify the effects of gain and phase mismatch on HR_3 and HR_3 requires performing the summation illustrated Figure 4.17. Recall that the Fourier series representation for $p_0(t)$, $p_1(t)$ and $p_2(t)$ are given by Equations (4.16), (4.17) and (4.18), respectively. Applying the gain and phase mismatch to $p_0(t)$ results in $p_{0M}(t)$ which is given by

$$\begin{aligned}
 p_{0M}(t) &= (1 + \alpha) p_0\left(t + \theta \frac{T}{2\pi}\right) \\
 &= (1 + \alpha) \frac{4\sqrt{2}}{\pi} \left[\cos(\omega t + \theta) - \frac{1}{3} \cos(3\omega t + 3\theta) + \frac{1}{5} \cos(5\omega t + 5\theta) + \dots \right] \\
 &= (1 + \alpha) \frac{4\sqrt{2}}{\pi} \left[(\cos \theta \cos \omega t - \sin \theta \sin \omega t) - \frac{1}{3} (\cos 3\theta \cos 3\omega t \right. \\
 &\quad \left. - \sin 3\theta \sin 3\omega t) + \frac{1}{5} (\cos 5\theta \cos 5\omega t - \sin 5\theta \sin 5\omega t) + \dots \right].
 \end{aligned} \tag{4.44}$$

Substituting $p_{0M}(t)$ for $p_0(t)$ in Equation (4.19) results in

$$\begin{aligned}
 ss_M(t) &= p_{0M}(t) + p_1(t) + p_2(t) \\
 &= \frac{4\sqrt{2}}{\pi} \left[\cos \omega t (1 + (1 + \alpha) \cos \theta) - \sin \omega t (1 + \alpha) \sin \theta \right. \\
 &\quad \left. + \frac{1}{3} (\cos 3\omega t (1 - (1 + \alpha) \cos 3\theta) + \sin 3\omega t (1 + \alpha) \sin 3\theta) \right. \\
 &\quad \left. - \frac{1}{5} (\cos 5\omega t (1 - (1 + \alpha) \cos 5\theta) + \sin 5\omega t (1 + \alpha) \sin 5\theta) \right].
 \end{aligned} \tag{4.45}$$

From Equation (4.45) the third harmonic rejection can be determined by summing the power of the third harmonic resulting in

$$\begin{aligned}
 HR_3 &= \frac{\text{power of third harmonic from SHS}}{\text{power of third harmonic from square wave}} \\
 &= \frac{\left(\frac{1}{3}\right)^2 (1 - (1 + \alpha) \cos 3\theta)^2 + \left(\frac{1}{3}\right)^2 ((1 + \alpha) \sin 3\theta)^2}{\left(\frac{1}{3}\right)^2} \\
 &= (1 - (1 + \alpha) \cos 3\theta)^2 + ((1 + \alpha) \sin 3\theta)^2 .
 \end{aligned} \tag{4.46}$$

A similar relationship can be derived for the fifth harmonic rejection, which results in

$$HR_5 = (1 - (1 + \alpha) \cos 5\theta)^2 + ((1 + \alpha) \sin 5\theta)^2 . \tag{4.47}$$

The third harmonic rejection is plotted in Figure 4.18 for three different values of gain mismatch α . The third harmonic rejection for one degree of phase mismatch and one percent gain error is approximately 25 dB. Therefore, applying Equation (4.42) results in a value for HD_3 of about 35 dB.

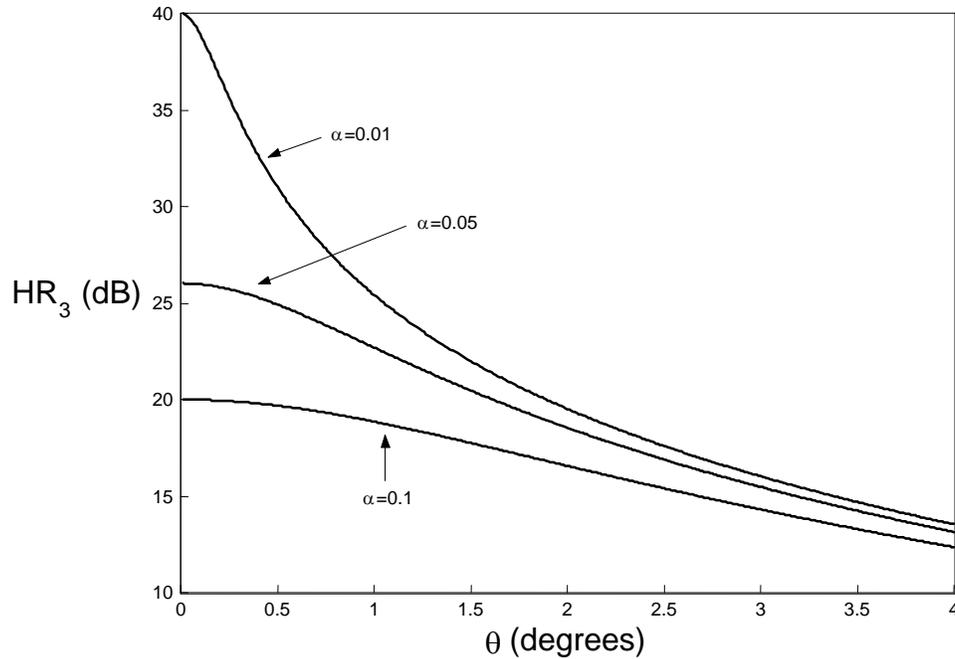


Figure 4.18 Third harmonic rejection as a function of gain and phase error.

A plot of the fifth harmonic rejection as function of phase and gain error is shown in Figure 4.19. For low values of phase error, HR_5 approaches HR_3 . For small phase errors, the gain mismatch is the dominant source of error. From Equations (4.46) and (4.47) it is clear that as the phase error approaches zero, the values of HR_3 and HR_5 converge. However, as the phase error gets larger, the HR_5 decreases faster than HR_3 due to its greater dependence on phase error. This greater dependence is evident by the fact that HR_5 is proportional to the cosine of 5θ while the third harmonic rejection is proportional to the cosine of 3θ .

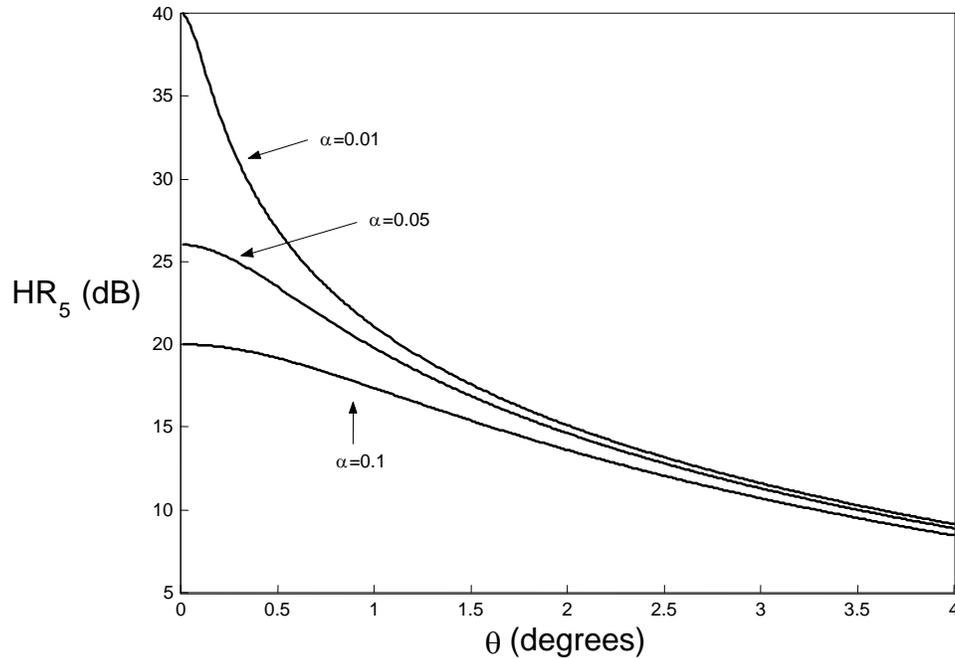


Figure 4.19 Fifth harmonic rejection as a function of gain and phase error.

The specifications of the system will determine the tolerances in both the gain and phase mismatch. Gain mismatches in the sub-mixers are a function of device mismatch. The two main sources of gain mismatch are the current sources and the input differential pairs. The overall mixer gain does not depend strongly on matching in the switches and therefore mismatch between sets of switches is less of a problem.

Device mismatch is typically caused by mismatch in the device geometry and threshold voltage mismatch. Large device sizes can limit mismatch caused due to device geometry and large gate to source bias voltages can reduce the effects of threshold voltage mismatch. Unfortunately, having both large device geometries and large bias voltages requires a larger current and thus more power.

Phase errors are also caused by device mismatch but to this point, the generation of multiple LO phases has not been discussed. The following section will discuss another potential limitation of the HRM and Section 4.8 will present two methods to generate the multiple LO phases that are needed for the HRM.

4.7 Harmonic Rejection in the General Case

To this point the majority of the analysis has assumed that a switching mixer effectively multiplies an input signal with a square-wave version of the LO signal. This assumption, while helpful for analytical purposes assumes instantaneous switching, which is clearly not possible. As was mentioned in Section 4.5 the switches require a finite transition time to completely switch the current. Therefore a better model of the waveform created by the switches is a waveform that is similar to a square wave except the transition times are non-zero [17][20][21]. An example of such a waveform is illustrated in Figure 4.20 showing a sinusoidal LO signal and the resulting ideal square wave and the more realistic waveform. This realistic waveform may still have large harmonics that need filtering but it is not clear the methods used by the HRM are still applicable for waveforms other than square waves. This section will propose a more general model for the waveform that is generated by the switches and show that the principles used in the HRM still apply.

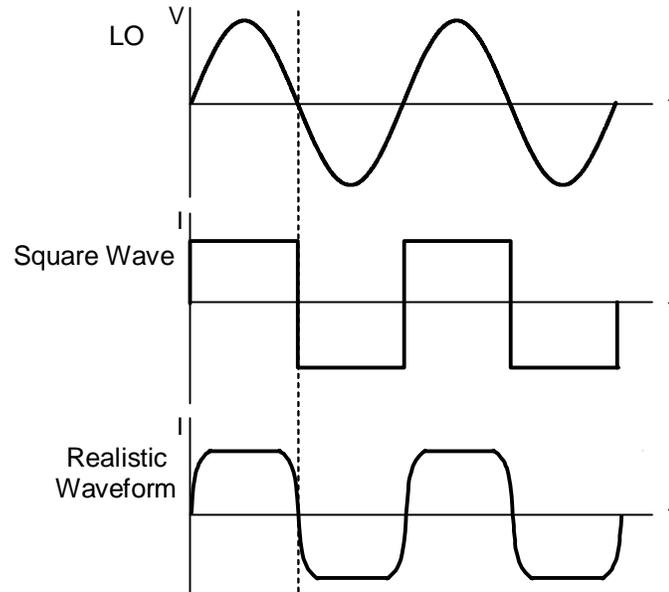


Figure 4.20 Square wave and a more realistic current waveform generated by a sinusoidal LO signal.

The rejection of the unwanted harmonics in the HRM relies on the cancellation of harmonics in square waves. This brings up the obvious question; if each sub mixer multiplies the input signal by a waveform that is not a square wave, yet still has significant harmonics, will the rejection of the unwanted harmonics still occur and if so, to what extent? To answer this question, a more general model of the waveform created by the mixer switches is necessary.

The waveform created by the switches, which is effectively multiplied by the input signal of each sub-mixer, can be modeled by a Fourier series expansion resulting in

$$x_0(t) = \sum_{k=-\infty}^{\infty} a_k e^{jk\omega_0 t} \quad (4.48)$$

where the coefficients a_k are complex Fourier coefficients. Limiting the discussion to the only positive frequencies for the sake of simplicity results in the following:

$$x_{p0}(t) = a_0 + a_1 e^{j\omega_0 t} + a_2 e^{j2\omega_0 t} + a_3 e^{j3\omega_0 t} + a_4 e^{j4\omega_0 t} + a_5 e^{j5\omega_0 t} + \dots \quad (4.49)$$

The use of differential circuits effectively eliminates the DC component and all the components with even Fourier coefficient subscripts. As a result, for the positive frequencies and differential signaling, the waveform can be describe by

$$x_{pd0}(t) = a_1 e^{j\omega_0 t} + a_3 e^{j3\omega_0 t} + a_5 e^{j5\omega_0 t} + a_7 e^{j7\omega_0 t} \dots \quad (4.50)$$

where the subscripts p and d represent positive frequency and differential signaling respectively. Equation (4.50) is a more general expression for the waveform that multiplies the input signal in a switching mixer. A general expression for this waveform is critical because in an actual mixer the waveform shape may not be well described. If the harmonic rejection depends on the switches generating pure square waves, the circuit may not practical.

Recall from Figure 4.14 that to generate the SHS waveform that was needed to reject the unwanted harmonics, three square waves were summed. Compared to the first square wave, $p_0(t)$, the second and third square waves are shifted in time by $T_0/8$ for $p_1(t)$ and $-T_0/8$ for $p_2(t)$. Furthermore, the amplitude of these two square waves was scaled by a factor $1/\sqrt{2}$ compared to $p_0(t)$. It was shown in Equation (4.19) that summing the three square waves resulted in a waveform in which the third and fifth harmonics were canceled. The logical question that follows is if summing time-shifted and amplitude scaled versions $x_{pd0}(t)$ will also result in the cancellation of the third and fifth harmonics? To answer this question it is useful to cite a well-known property of Fourier series

relating a shift in time to the values of the Fourier series coefficients. This property of Fourier series is described by the following:

$$x(t - t_0) \xleftrightarrow{\mathfrak{F}} a_k e^{-jk(2\pi/T_0)t_0} . \quad (4.51)$$

Equation (4.51) states that a shift in time by t_0 corresponds to the Fourier coefficients getting multiplied by the complex exponential that is shown.

In Section 4.4 the expressions for $p_1(t)$ and $p_2(t)$ were derived from the expression for $p_0(t)$. In this, the more general case, $x_{pd0}(t)$ will function as the equivalent of $p_0(t)$ for the square-wave case and thus the other time shifted waveforms will be derived from $x_{pd0}(t)$. Applying the property from Equation (4.51) to $x_{pd0}(t)$ with a time shift of $T_0/8$ and amplitude scaling of $1/\sqrt{2}$ results in

$$\begin{aligned} x_{pd1}(t) &= \frac{1}{\sqrt{2}} x_{pd0}(t - T_0/8) \\ &= \frac{1}{\sqrt{2}} \left(a_1 e^{-j\pi/4} e^{j\omega_0 t} + a_3 e^{-j3\pi/4} e^{j3\omega_0 t} + a_5 e^{-j5\pi/4} e^{j5\omega_0 t} + a_7 e^{-j7\pi/4} e^{j7\omega_0 t} \dots \right) \quad (4.52) \\ &= a_1 (0.5 - 0.5j) e^{j\omega_0 t} + a_3 (-0.5 - 0.5j) e^{j3\omega_0 t} \\ &\quad + a_5 (-0.5 + 0.5j) e^{j5\omega_0 t} + a_7 (0.5 + 0.5j) e^{j7\omega_0 t} \dots . \end{aligned}$$

The signal $x_{pd1}(t)$ can be thought of as the equivalent of $p_1(t)$ but for the more general case.

Applying a time shift of $-T_0/8$ and amplitude scaling of $1/\sqrt{2}$ results in the more general equivalent of $p_2(t)$ and this is given by

$$\begin{aligned} x_{pd2}(t) &= \frac{1}{\sqrt{2}} x_{pd0}(t + T_0/8) \\ &= \frac{1}{\sqrt{2}} \left(a_1 e^{j\pi/4} e^{j\omega_0 t} + a_3 e^{j3\pi/4} e^{j3\omega_0 t} + a_5 e^{j5\pi/4} e^{j5\omega_0 t} + a_7 e^{j7\pi/4} e^{j7\omega_0 t} \dots \right) \quad (4.53) \\ &= a_1 (0.5 + 0.5j) e^{j\omega_0 t} + a_3 (-0.5 + 0.5j) e^{j3\omega_0 t} \\ &\quad + a_5 (-0.5 - 0.5j) e^{j5\omega_0 t} + a_7 (0.5 - 0.5j) e^{j7\omega_0 t} \dots . \end{aligned}$$

To complete the general model, the three general expressions given by Equations (4.51), (4.52), and (4.53) need to be summed just as the square waves were summed in Equation (4.19). This summation results in

$$x_{pd0}(t) + x_{pd1}(t) + x_{pd2}(t) = 2a_1e^{j\omega_o t} + 2a_7e^{j7\omega_o t} + \dots \quad (4.54)$$

The results of Equation (4.54) are very interesting and indicate that the third and fifth harmonics will be cancelled regardless of the shape of the actual waveform that is created by the switches. This result is absolutely critical to the practicality of the HRM technique because the shape of the switching waveform may not be well described. Recall that this derivation was done only for the positive frequencies of Equation (4.48). A similar derivation can be performed for the negative frequencies with identical results, which also shows the rejection of the third and fifth harmonics.

Although complex exponentials were used to demonstrate the cancellation of the third and fifth harmonics for the general case, the trigonometric forms used for the case of square waves may provide some added intuition. If the waveform created by the first switching mixer is symmetric about the axis when $t=0$, it can be shown that it will only contain cosine terms. For example, the Fourier series representation of $p_0(t)$ in Equation (4.16) contains only cosine terms. If the other two time-shifted waveforms are generated from this original waveform, it becomes clear how the harmonic cancellation occurs. Shifting in time by $T/8$ and $-T/8$ creates waveforms that have both sine and cosine terms. However, the sine components of the two new waveforms are equal amplitude and opposite in sign for the third and fifth harmonics. Furthermore, the cosine terms for the third and fifth harmonic are the same sign as each other but are opposite sign compared

with original waveform. Therefore, when summed, the cosine terms and the sine terms of the third and fifth harmonics will cancel if scaled properly and this independent of the shape of the waveforms. This cancellation was apparent when $p_0(t)$, $p_1(t)$ and $p_2(t)$ were summed. It can be concluded that in the more general case, the same type of cancellation occurs except the coefficients of the third and fifth harmonics may not be $1/3$ and $1/5$ respectively. These values for the coefficients represent the worst-case scenario for switching mixers. That is to say that deviation from the square wave approximation will make these values smaller. The values of $1/3$ and $1/5$ represent the upper bound for the coefficients of the third and fifth harmonic created by the switches of a switching mixer.

The mismatch analysis from the previous section also relies on the square-wave assumption and therefore, this analysis also represents the worst case. In practice, the rejection of the third and the fifth harmonics, for a given phase and gain mismatch, will be at worst those values that are determined by Equations (4.46) and (4.47). The rejection will probably be somewhat better than these equations predict but this is dependant on the exact shape of the waveform created by the switches.

4.8 LO Phase Generation

The generation of multiple LO phases is critical to the operation of the HRM. In the HRM presented in the previous section, eight LO phases were needed to reject the third and fifth harmonics. The method by which these eight phases are generated is the focus of this section.

Many methods have been presented to generate different LO phases but these methods have primarily focused on quadrature phase generation needed by image-reject mixers[22][23][24]. In this section, two methods to generate eight LO phases will be discussed. The first method uses a frequency divider to generate the necessary phases. The second method creates the phases with the use of a passive poly-phase filter similar to the poly-phase filters used for quadrature generation.

4.8.1 Generation of Eight Phases using a Frequency Divider

As mentioned eight LO phases used by needed in the HRM. Consequently, the resulting LO signals are offset in phase from one another by 45 degrees. Differential signaling inherently provides two phases. As a result, a frequency divider, which divides by four, will provide the necessary 45-degree phase shifting. A divide-by-four schematic diagram is illustrated in Figure 4.21 using two edge-triggered master-slave (MS) D flip-flops. The LO signal is applied to the clock input and the complementary output from the second flip-flop, designated D_2 is fed back to the input of the first flip-flop, D_1 . Signals Q_1 and Q_2 are 90 degrees out of phase, but a more detailed view of the circuit illustrates how the 45-degree phase shifts are generated.

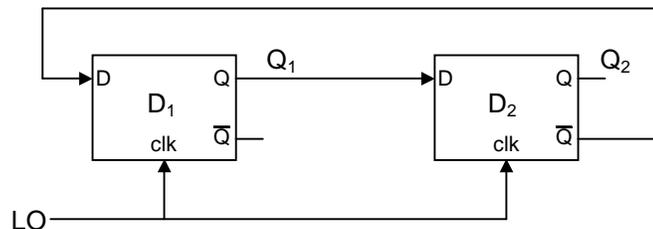


Figure 4.21 Divide-by-four frequency divider.

A master-slave flip-flop, as the name implies, is composed of a master flip-flop and slave flip-flop. Expanding each D flip-flop, by showing both the master and the slave flip-flops, and including the differential signal path, results in the diagram shown in Figure 4.22. Each MS D flip-flop is composed of two negative edge triggered D flip-flops, a master and a slave. Therefore, internal to the MS flip-flop, are internal nodes, designated, Q_{1M} and Q_{2M} , which are the outputs of each master flip-flop. The outputs from the slave flip-flops are simply the output of the composite MS D flip-flop and are given by Q_1 and Q_2 . The complements of these four signals are also present but are not labeled for the sake of simplicity.

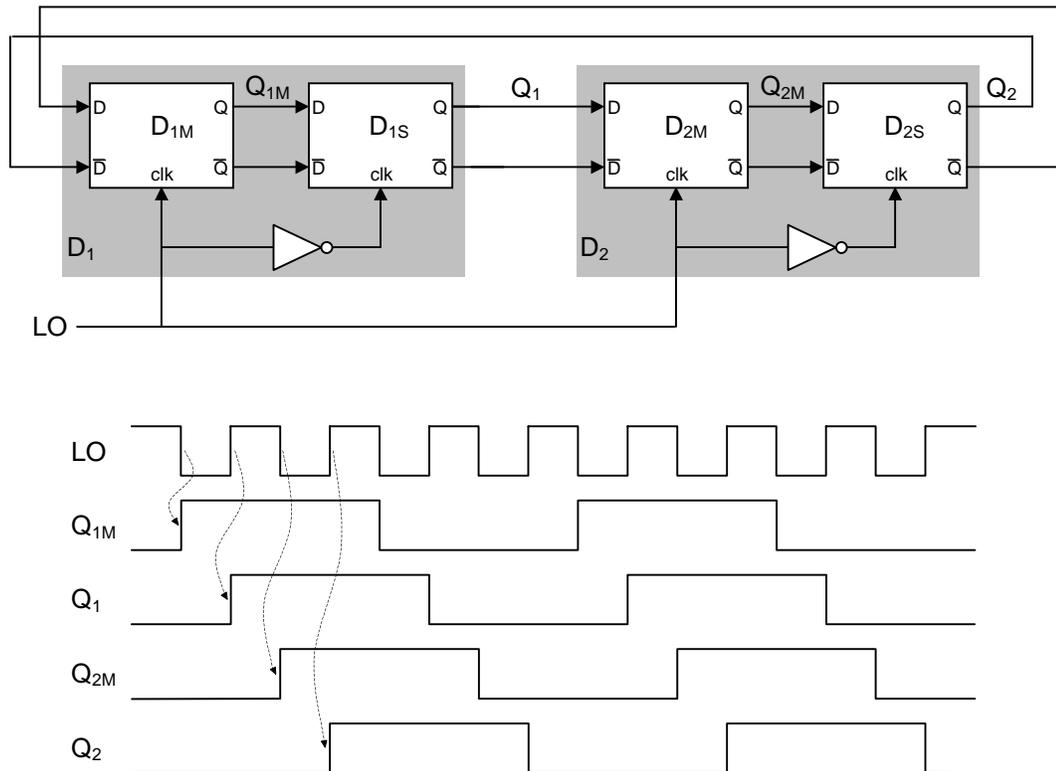


Figure 4.22 Expanded view of divide-by-four and timing diagram.

A timing diagram for the divide-by-four circuit is also shown in Figure 4.22. The diagram shows the LO signal and the positive output from each of the four negative edge triggered D flip-flops. The first falling edge of the LO changes the value of Q_{IM} from low to high. This transition from low to high is then propagated all the way to Q_2 by each falling or rising edge. The process is then repeated as Q_{IM} changes from high to low and this change gets propagated. It is clear from the timing diagram that frequency division has occurred because each of the four flip-flop outputs is one fourth of the LO frequency. Furthermore, the timing diagram shows that a 45-degree phase shift is present between each successive different flip-flop output. Combined with the complementary signals, eight phases of the frequency-divided signal have been generated.

It was shown that the frequency divider in Figure 4.22 is capable of generating the phases that are needed in an HRM. One drawback of this method is that the LO must operate at four times the frequency that is needed by the LO port of the HRM. Depending frequency of the LO that is needed in a particular application, this may represent a serious problem. However, as the frequency needed by the mixer LO port gets higher, the spacing between the harmonics also increases. As a result, the filtering becomes easier and a HRM might be unnecessary. In a two-step mixer the IF mixers are generally operating well below the RF frequency and therefore the problem of the LO operating at a frequency, which is four times the mixer frequency, is minimized. Consequently, this frequency divider does not present a serious problem when HRMs are operated at IF.

4.8.2 Generation of Eight Phases Using a Passive RC Filter

Another method, which may be used to generate the eight phases needed by the HRM, is based on asymmetric polyphase filters [25][26]. Asymmetric polyphase filters have long been used to generate quadrature signals that are needed for image rejection in both transmitters and receivers.

These filters are based on the basic principles of a single pole created by a RC low-pass filter and the pole-zero combination created by a CR high-pass filter. A circuit diagram of a RC low-pass filter is illustrated in Figure 4.23a, which also shows a phasor representation of the input and output at the 3 dB frequency. At the 3 dB frequency, that is when the angular frequency of the input is equal to the reciprocal of the RC product, the phase of the output is retarded by 45 degrees compared with the input. This is evident by inspection of the transfer function, which at the 3 dB frequency is given by

$$H(j\omega) = \frac{1-j}{2}. \quad (4.55)$$

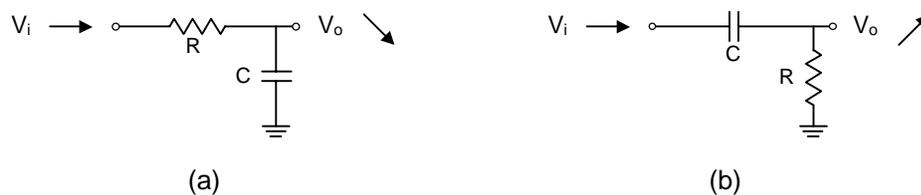


Figure 4.23 (a) RC low-pass filter. (b) CR high-pass filter.

Also shown in Figure 4.23 is the circuit diagram of a high-pass CR filter along with the input and output phasors at the 3 dB frequency. This filter advances the phase of

the output by 45 degrees and this is clear by its transfer function at the 3 dB frequency, which is given by

$$H(j\omega) = \frac{1+j}{2}. \quad (4.56)$$

Asymmetric polyphase filters used for quadrature generation are based on the two circuits shown in Figure 4.23. An example of such a filter is illustrated in Figure 4.24 again showing the phasor representation of the input and output. The two inputs, V_{ip} and V_{in} , make up a differential input signal. Each input then passes through a RC low-pass filter to one output and also a CR high-pass filter to a different output. As a result, the phase of each of the four outputs is rotated by 45 degrees or -45 degrees compared to its input. Consequently, the output consists of four signals that are 90 degrees out of phase with one another. These four phases form a pair of differential quadrature signals.

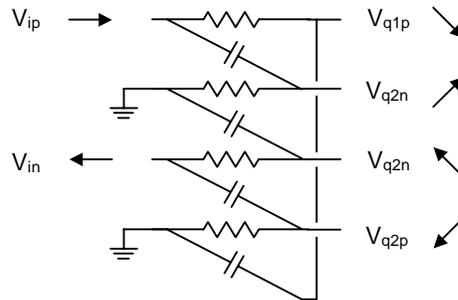


Figure 4.24 Polyphase filter used for quadrature generation.

The 45-degree phase shifts created by the RC and CR filters create four phases of the input signal that are 90 degrees out of phase. However, the HRM requires eight phases to successfully cancel the unwanted harmonics. To create these eight phases, a circuit that creates a 22.5-degree phase shift is needed. Shown in Figure 4.25 are two

circuits that are very similar to the circuits in Figure 4.23 but an extra resistor has been added that is in series with the capacitor. To see how these RCR circuits create a 22.5-degree phase shift it is useful to examine the transfer function of these two circuits. Furthermore, to achieve the desired phase shift at a certain frequency, the relationship between R_1 , R_2 and C needs to be defined.

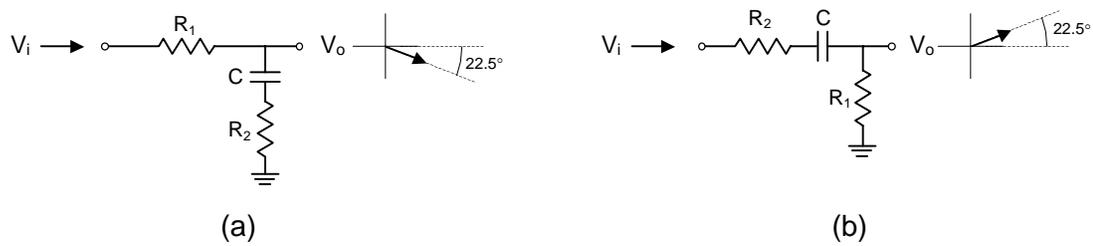


Figure 4.25 RCR circuits giving 22.5 degree phase shift.

The RCR circuit shown in Figure 4.25a retards the phase angle of the input signal by 22.5 degrees given that the values of R_1 , R_2 and C are chosen correctly. The transfer function of this circuit is given by

$$H_1(j\omega) = \frac{1 + j\omega CR_2}{1 + j\omega C(R_1 + R_2)}. \quad (4.57)$$

At $\omega = \omega_z = 1/R_2C$, the transfer function reduces to

$$H_1(j\omega) \Big|_{\omega = \frac{1}{R_2C}} = \frac{1 + j_2}{1 + j \left(\frac{R_1 + R_2}{R_2} \right)} \quad (4.58)$$

and at this frequency the phase is given by

$$\angle H_1(j\omega_z) = 45^\circ - \tan^{-1} \left(\frac{R_1 + R_2}{R_2} \right). \quad (4.59)$$

Therefore, if $R_1 = R_2\sqrt{2}$, the phase is equal to

$$\begin{aligned}\angle H_1(j\omega_z) &= 45^\circ - \tan^{-1}(\sqrt{2} + 1) \\ &= -22.5^\circ.\end{aligned}\tag{4.60}$$

Therefore a phase shift of -22.5 degrees occurs at ω_z when $R_1 = R_2\sqrt{2}$.

Now that the relationship between R_1 and R_2 has been determined to achieve the proper phase shift, the magnitude of the transfer function at ω_z can be determined from Equation (4.58) and it is given by

$$|H_1(j\omega_z)| = \frac{1}{\sqrt{2 + \sqrt{2}}}.\tag{4.61}$$

The focus now shifts to the circuit shown in Figure 4.25b. The transfer function of this RCR circuit is given by

$$H_2(j\omega) = \frac{j\omega CR_1}{1 + j\omega C(R_1 + R_2)}\tag{4.62}$$

At ω_z the phase angle is equal to

$$\begin{aligned}\angle H_2(j\omega_z) &= 90^\circ - \tan^{-1}(\sqrt{2} + 1) \\ &= 22.5^\circ\end{aligned}\tag{4.63}$$

and the magnitude is equal to

$$|H_2(j\omega_z)| = \frac{1}{\sqrt{2 + \sqrt{2}}}.\tag{4.64}$$

It was shown that the circuits in Figure 4.25 generate a 22.5 degree and a -22.5 phase shift at a particular frequency, ω_z . Furthermore, at this frequency the magnitudes of the two transfer functions are equal. Therefore, these RCR circuits may be used in a similar fashion to the way the RC filters in Figure 4.23 are applied to the quadrature polyphase filter of Figure 4.24. A circuit diagram of such a configuration using the RCR filters is illustrated in Figure 4.26 showing the generation of eight output phases from two input phases. A quadrature polyphase filter first generates quadrature signals from the differential input. These quadrature signals are then applied to an eight-phase polyphase filter that generates the eight phases. The eight-phase polyphase filter is constructed very much like a quadrature polyphase filter except the basic structure uses RCR filters as apposed to RC filters. Furthermore, the quadrature polyphase filter contains four core RC cells while the eight-phase polyphase is comprised of eight core RCR structures. Each output is composed of two signals because there are two paths to every output. One path is through resistor R_1 and this provides a -22.5-degree phase shift. The second path to each output is through the R_2 - C series combination and this provides a 22.5-degree phase shift. Phasor representations of the two components of each output are illustrated as well as the actual output, which is the vector sum of these two components.

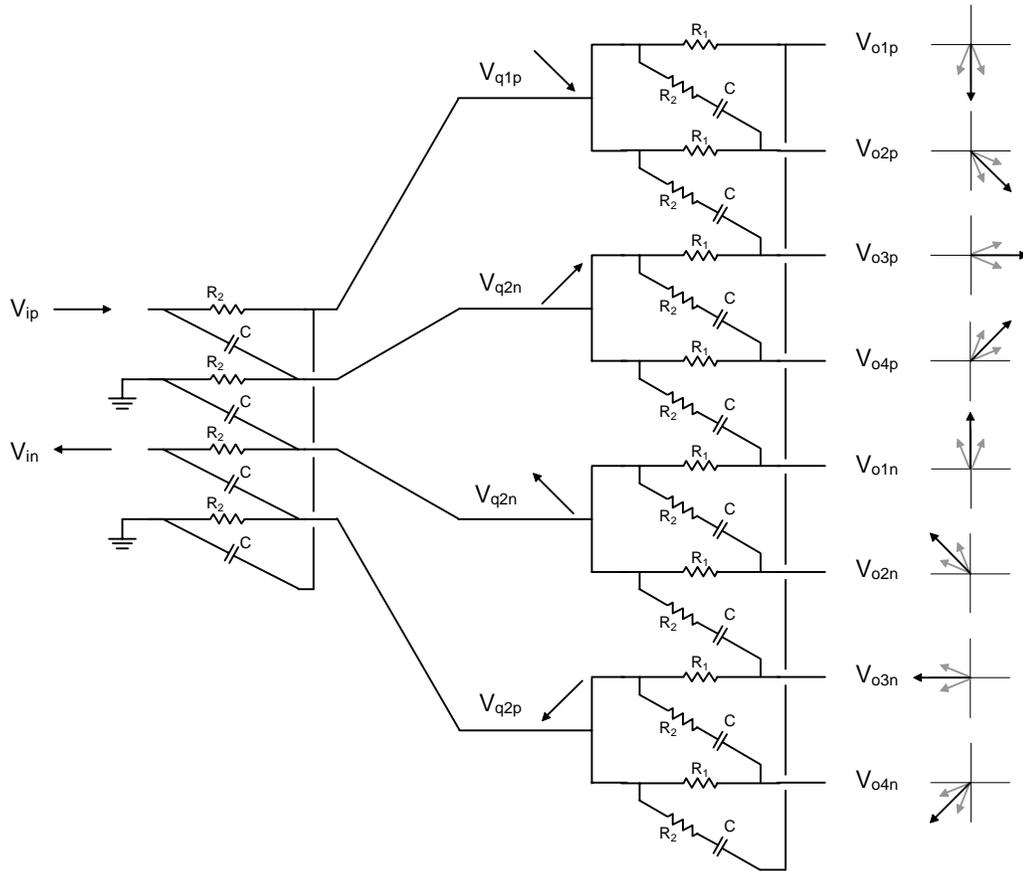


Figure 4.26 Eight-phase polyphase filter.

The filter configuration illustrated in Figure 4.26 contains both a quadrature polyphase filter and an eight-phase polyphase filter. The eight-phase filter requires quadrature inputs to operate correctly. If quadrature signals are available and have thus been generated by other means, the quadrature polyphase filter is not necessary.

One significant drawback of the eight-phase polyphase filter stems from passive device variability on integrated circuits. The eight-phase polyphase filter generates the proper phase shift and equal output magnitudes at one particular frequency and thus the response is very narrow band. Therefore, variation in the value of the resistors and capacitors will cause errors in the phase and magnitude of the output signals. Because

these signals will be applied to the switching transistors in a HRM, the magnitude variation may not be a serious problem. However, phase error, as discussed in the previous section, can significantly affect the rejection of unwanted harmonics. This problem also occurs in quadrature polyphase filters and has been lessened by using multiple filters in series. The frequency of operation for each stage is then staggered to effectively broaden the band of operation [25][26][27][28]. This approach can be applied to the eight-phase polyphase filter resulting in the circuit diagram illustrated in Figure 4.27.

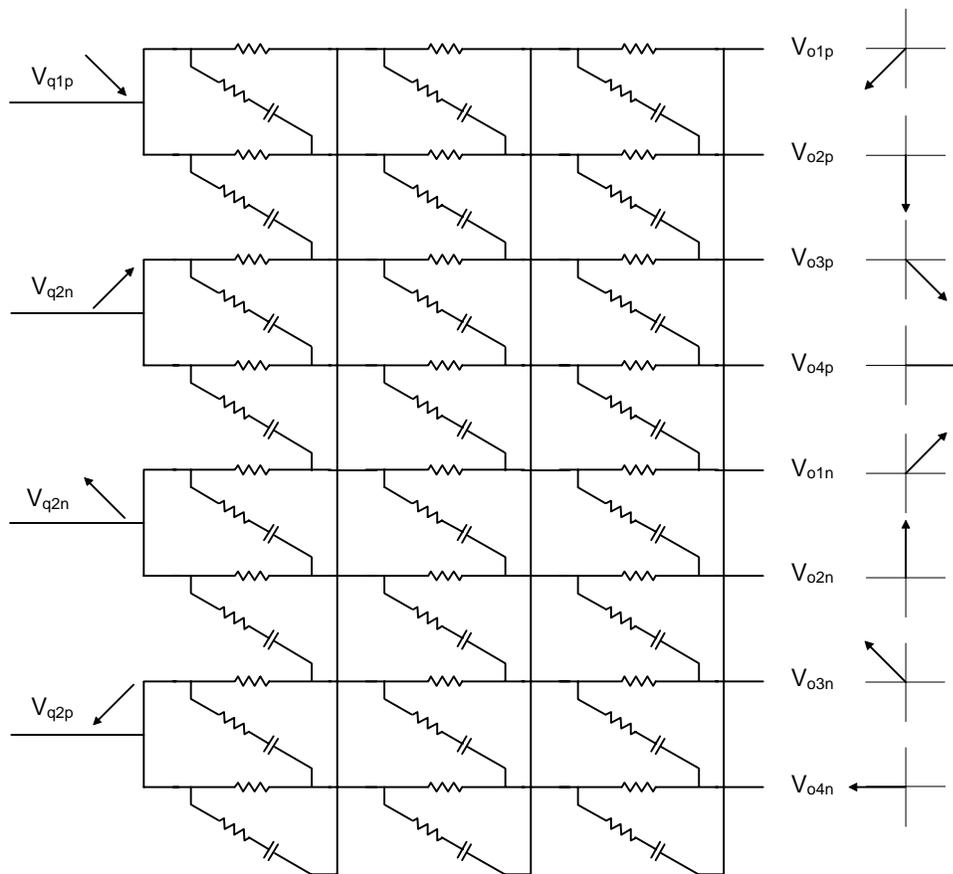


Figure 4.27 Three stage eight-phase polyphase filter.

4.9 Summary

In this chapter the harmonic-rejection mixer was presented as method to reduce the filtering requirements in transmitter applications. The HRM is a mixer that can reject unwanted harmonics created by the switches of switching mixer. A method to achieve the proper scaling necessary for cancellation was also presented which involved scaling both the size of the input devices, the switching transistors and the bias current source. Critical limitations of the HRM were analyzed such as mismatch and short-channel effects. It was shown that the rejection of the unwanted harmonics by the HRM is independent of the actual waveform that is created by the switches in a switching mixer. In addition to the HRM, two methods to generate the multiple LO phases needed by the HRM were discussed.

The HRM facilitates high integration in transmitters by potentially eliminating the need for IF filtering while allowing for a high performance two-step transmitter. The following chapter will present a transmitter architecture, based on the HRM, that has the potential for high integration in CMOS, high performance and the potential to operate with multiple RF standards.

Chapter 5

Harmonic Rejection Transmitter

5.1 Introduction

While the previous chapter presented the HRM as a method to frequency translate signals and simultaneously reject unwanted harmonics created in the mixing process, the focus now shifts to applying the HRM to a complete RF transmitter. This chapter will present a transmitter architecture that takes advantage of the properties of the HRM with regards to decreased filtering. In addition to the use of the HRM, the architecture will be evaluated for three key properties: the potential for integration on single CMOS substrate, the potential for high performance in CMOS, and the ability of operate with multiple RF standards.

A number of transmitter architectures were introduced in Chapter 3 and discussed with regard to integration, performance and multi-standard capability. Transmitters generally fall into two broad categories: VCO based and mixer based. A summary of the conclusions from Chapter 3 begins with VCO based transmitters. VCO based transmitters such as the OPLL allow for high single chip integration and high performance but they are limited to constant-envelope modulation schemes. Therefore they are not well suited for multi-standard operation because of the prevalence of non-constant envelope modulation. To address this issue, similar VCO based transmitters have been used which apply phase and amplitude modulation in two separate signal paths. Depending on the details of how this is accomplished this can lead to two problems. If the amplitude modulation is applied open loop, then precise characterization of the PA is necessary for proper operation. The need for PA characterization is not practical if a single chip solution is the goal. Alternatively, the amplitude modulation can be applied in a closed loop and this presents problems for larger signal bandwidths because of the precise timing matching that is needed between the phase and amplitude signal paths. As a result, VCO based architectures are not well suited for multi-standard operation because some of the most heavily used RF standards use large channel bandwidths for high data rates.

The second broad category of transmitters, mixer based transmitters, generally use quadrature modulators to both modulate and frequency translate the baseband signal. Direct-conversion transmitters frequency translate the baseband signal to RF in a single step. These transmitters suffer from LO pulling, large LO feedthrough and quadrature inaccuracy because it is necessary to generate high frequency quadrature LO signals.

Although certain techniques have been used to minimize the effects of LO pulling, the performance of these transmitters is generally limited. Another performance concern is intermodulation in the PA which can lead to a third order component of the baseband signal.

To alleviate some of the performance issues with homodyne transmitters, the up-conversion is often performed in two steps. These heterodyne transmitters avoid the LO pulling problem because neither LO is operating at the same frequency as the output. This is a critical issue especially when the goal is the integration of a relatively high power PA with low phase noise frequency synthesizers. Other benefits of heterodyne transmitters are the higher quadrature accuracy that is possible because the generation of quadrature LO signals occurs at a lower frequency and LO feedthrough is reduced because of the lower frequency used by the mixers in the quadrature modulator. Even with these performance advantages, heterodyne transmitters suffer from a significant drawback because an IF filter is needed to filter the harmonics created by the mixers in the quadrature modulator. Therefore, integration in a single IC is a significant challenge in heterodyne transmitters.

The architecture presented in this chapter is a mixer based transmitter which uses two steps to perform the up-conversion and is termed the Harmonic Rejection Transmitter (HRT). The use of HRMs mitigates the need for a discrete IF filter thus making the architecture well suited for single chip integration. The next section will present the architecture and some of its key characteristics. Section 5.3 will discuss the image-rejection properties of the transmitter and will present a tuning scheme to increase

performance that is easily implemented. The advantages of this architecture with regard to synthesizer integration will then be discussed in Section 5.4.

5.2 Harmonic Rejection Transmitter Architecture

The HRT uses two steps to frequency translate the baseband signal to RF and therefore retains the performance advantages of conventional heterodyne transmitters in that both LO pulling and LO feedthrough are reduced and quadrature accuracy is improved because the quadrature generation of the IF LO occurs at a lower frequency compared to direct-conversion transmitters. In addition the HRT has a number of advantages with respect to integration and performance compared with a traditional heterodyne transmitter. The HRT will be analyzed with respect to the three critical issues: integration, performance, and multi-standard operation. A block diagram of the HRT transmitter is illustrated in Figure 5.1 where the HRMs are designated by a symbol that is similar to a standard mixer except extra lines have been added to indicate the use of multiple phases of the LO. Furthermore, a thick line is used to indicate multiple lines. This is necessary because each HRM uses three differentials LO signals. The thick line coming from the eight phase generator represents four pairs of differential signals.

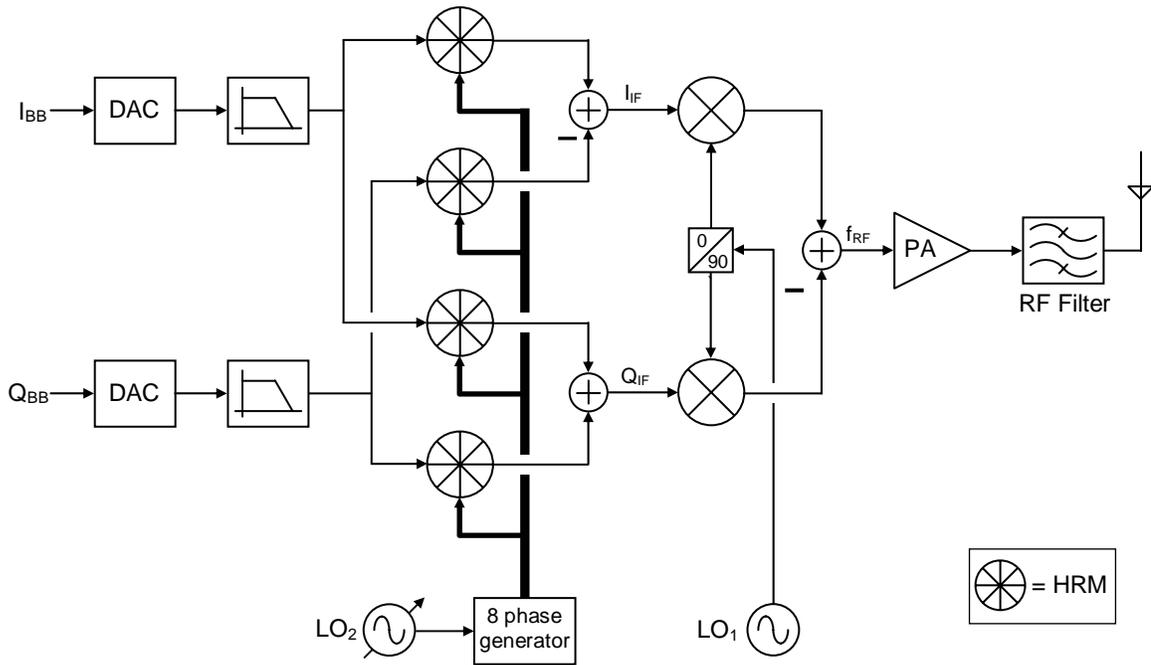


Figure 5.1 Block diagram of a Harmonic Rejection Transmitter.

The basic function of the HRT is as follows. The baseband is identical to the homodyne and heterodyne transmitter which were discussed in Chapter 3 as digital in-phase and quadrature baseband signals pass through a DAC and then a low-pass filter. The analog I and Q baseband signals are then up-converted to IF with four HRMs generating I and Q IF signals. By generating I and Q IF signals the image created in the up-conversion to RF can be attenuated through the use of an image-rejection RF mixer. The image-reject configuration uses two RF mixers and I and Q LO_1 signals to complete the up-conversion to RF. By generating I and Q IF signals and then using an image-reject mixer, the RF filtering requirements are reduced compared to the conventional heterodyne transmitter. Finally, the RF signal is then amplified by a PA before it passes through an RF bandpass filter and on to the antenna.

Another difference between the HRT and a conventional heterodyne transmitter is the method of channel selection. In a conventional heterodyne transmitter the RF LO performs channel selection and IF LO is fixed in frequency. In contrast, in the HRT the IF LO performs the channel selection and the RF LO is fixed in frequency. The reasons for this reversal will be discussed in more detail in a later section but in short, by reversing the roles of the two synthesizers, the total phase noise of both synthesizers can be reduced, which eases the RF filtering requirements and improves the overall performance. With the basic operation of the HRT complete, the HRT will be discussed with respect to each of the three important issues: integration, performance and multi-standard capability.

5.2.1 Integration Advantages of the HRT

The first of these to be discussed will be integration. It should be noted that discrete components are typically used to increase performance whether it be modulation accuracy or more typically spectral emissions. Therefore, the comments on the potential for high integration all are made with the underlying assumption that performance will not be significantly degraded.

One of the major drawbacks of heterodyne transmitters is the IF harmonics that are created with switching mixers. The HRT uses HRMs to up-convert the baseband signal to IF which eases the filtering requirements by attenuating the third and fifth harmonics. The remaining higher order harmonics are then filtered with an RF filter after the PA. In this way the IF filter used by a conventional heterodyne transmitter, as shown

in Figure 3.7, is no longer necessary and integration in CMOS is made easier. However, before integration of the entire signal path is complete, the RF filter prior to the PA needs to be addressed.

In addition to the relaxation of the IF filtering requirement, the requirements on the pre-PA RF filter have also been relaxed through use of HRMs. In a conventional heterodyne transmitter, the pre-PA filter is needed to reduce potential third order intermodulation and to attenuate transmitted noise and spurs. As previously mentioned in Chapter 3, energy located at three times the carrier frequency can intermodulate in the PA and degrade the modulation accuracy. In a heterodyne transmitter energy is generated at three times the carrier frequency from the $3LO_2 \times 3LO_1$ mixing product. However, due to the rejection of the third IF harmonic, the degradation of modulation accuracy by third order intermodulation in the PA is significantly reduced in the HRT. Consequently, the pre-PA RF filter is no longer needed to prevent significant third order intermodulation. Additionally, since the transmitted wideband noise is typically dominated by the phase noise of the synthesizers, switching the roles of the synthesizers potentially allows for lower phase noise with integrated synthesizers, which in turn reduces the pre-PA filtering requirement. As a result PA integration is now possible because the RF filtering can be performed with one RF filter after the PA without degradation to the modulation accuracy.

Another aspect of the HRT that was briefly mentioned earlier is the generation of I and Q IF signals, which also reduces the RF filtering requirement. Quadrature IF signals were generated essentially using two quadrature modulators. The first quadrature modulator generates $I_{BB}I_{LO2} - Q_{BB}Q_{LO2}$, which is the standard quadrature modulator

configuration resulting in I_{IF} . The second quadrature modulator generates $I_{BB}Q_{LO2} + Q_{BB}I_{LO2}$ and this in turn generates Q_{IF} . By generating quadrature IF signals, the image created when the IF signal is up-converted to RF can be attenuated with an image-rejection mixer assuming quadrature LO_1 signals are available. This is readily apparent when the RF output is represented by the following:

$$\begin{aligned}
 RF_{out} &= I_{IF} \times I_{LO1} - Q_{IF} \times Q_{LO1} \\
 &= \cos(\omega_{LO2}t + \phi_m(t))\cos(\omega_{LO1}t) - \sin(\omega_{LO2}t + \phi_m(t))\sin(\omega_{LO1}t) \\
 &= \cos((\omega_{LO1} + \omega_{LO2})t + \phi_m(t)).
 \end{aligned} \tag{5.1}$$

Clearly the RF output contains only the sum of the two terms while the difference has been rejected. By rejecting the image the RF filtering requirements are relaxed and sufficient RF filtering can be accomplished with a single RF filter after the PA.

The ability to integrate low phase noise frequency synthesizers is another important element to transmitter integration because the phase noise affects the amount of filtering that is needed at RF and the overall performance of the transmitter. Therefore, for complete integration the phase noise of the frequency synthesizers needs to be minimized given the relatively low-Q passive components that are available in a standard CMOS processes. By switching the roles of the synthesizers the overall phase noise can be reduced and this will be discussed in more detail in Section 5.4.

5.2.2 Performance

The performance of a transmitter is dependant to some extent on the level of integration of the transmitter because discrete components are typically used to increase

performance. However, aspects other than the discrete components affect the performance and these will now be the focus. The performance of a transmitter can generally be broken down into two components: modulation accuracy and spectral emissions. The modulation accuracy is affected by in-band signals while the spectral emissions are a measure of signals that are transmitted both in and out of band.

The modulation accuracy is primarily limited by the following problems: gain matching in the quadrature modulator, phase matching in the quadrature modulator, DC offset in the quadrature modulator, LO feedthrough, third order distortion and phase noise of the frequency synthesizers. The gain and phase matching are typically lumped together into the image rejection in a SSB test. The image rejection in this test is a function of both the gain and phase mismatch and it will be shown in the Section 5.3 that the image rejection of the HRT is improved compared with a conventional heterodyne transmitter. Furthermore, this section will also present a low-frequency tuning method which can further improve the image rejection and lower the DC offset.

The benefit of heterodyne transmitters with respect to LO feedthrough has already been discussed with regard to the HRT. In addition, the use of Gilbert cell mixers further lowers the LO feedthrough because of the high isolation between the LO port and the output in these mixers. It has also already been shown that third order intermodulation in the PA is lowered in the HRT due to the rejection of the third IF harmonic. Therefore the third order distortion is limited by the baseband circuits and the IF mixers and care must be taken when designing these circuits.

Another key performance issue is the phase noise of the frequency synthesizers. The phase noise affects both the modulation accuracy and the spectral emissions and

therefore limiting the phase noise is critical for high performance. The in-band phase noise affects the modulation accuracy while the out-of-band phase noise is often the limiting factor with respect to spectral emissions. The phase noise of frequency synthesizers will be discussed in more detail in Section 5.4.

Other factors which can affect the spectral emissions include the IF image and LO mixing products. The problem with the IF image was addressed with the use of an RF image-reject mixer. The problem with LO mixing products is caused when order higher harmonics of the two frequency synthesizers mix and the resulting mixing product falls in-band where it cannot be filtered. In typical heterodyne transmitters mixing products are mapped out for the set of frequencies used by the two synthesizers. The IF frequency is then chosen such that no large mixing products fall in-band and cause a spectral mask violation. This is made significantly easier by the HRMs because of the rejection of some of not only the third and fifth harmonics but also the 11th and 13th harmonics.

5.2.3 Multi-Standard Capability

With regard to multi-standard operation, mixer based transmitters do not suffer from the limitations that affect VCO based transmitters. As a result mixer based transmitters are well-suited for multi-standard operation but some issues still need to be addressed. To operate with multiple RF standards that employ different channel bandwidths, the bandwidth of the baseband circuits must be either reconfigurable or wide enough to accommodate the largest bandwidth. Reconfigurable baseband circuits would be desirable from a power perspective because the lower bandwidth signals would

potentially require less power, but this may increase the overall complexity. The trade-off between complexity and power will depend on other factors including the actual standards that are being implemented and the feasibility of designing reconfigurable circuits

The linearity of the transmitter is also an issue when considering multi-standard operation. Constant and non-constant envelope modulation schemes place significantly different linearity requirements on the transmitter. The linearity requirements of a NCE modulation scheme are much higher than CE modulation. As result, similar to the bandwidth requirements the linearity of the circuit blocks either needs to be reconfigurable or linear enough to accommodate NCE modulation. Again reconfigurable circuits are desirable from a power consumption perspective but this may add to the overall complexity.

5.3 Image Rejection

To begin the discussion of image rejection in the HRT it is instructive to understand image rejection in a direct conversion transmitter, which is essentially a quadrature modulator. In normal operation, a transmitter modulates and up-converts a baseband signal, which is often this is often done simultaneously by a quadrature modulator. To gain insight into the performance of a quadrature modulator sinusoidal signals are applied to the baseband inputs as apposed to modulated signals. In this case, when gain and phase mismatch is present, the output will be composed of two components: a term at the sum of the LO and baseband signals and a term at the difference of the two. The ratio of

these two components is termed the image rejection or sideband suppression of the modulator. The image rejection is primarily affected by phase mismatch in the quadrature LO signals and gain mismatch in the I and Q signal paths. In general the image rejection is not a strong function of gain mismatch in the LOs because the gain of the mixers is not a strong function of the LO amplitude. Similarly, phase mismatch between the I and Q baseband signal paths is usually small and therefore ignored. A model of quadrature modulator used for evaluating the effects of gain mismatch and phase error on the image rejection is shown in Figure 5.2. The gain error is modeled by the α and LO phase error modeled by θ .

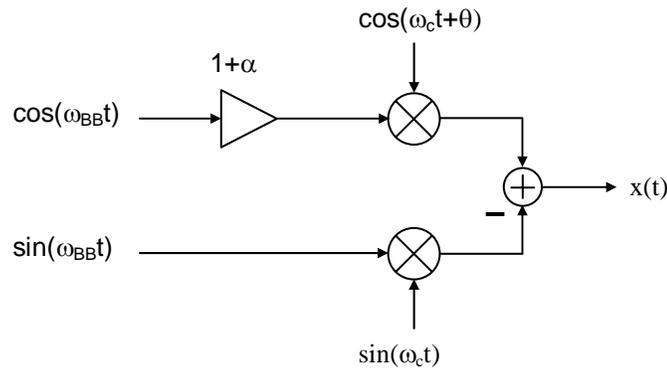


Figure 5.2 Gain and phase mismatch in a quadrature modulator.

To determine the image rejection the power of the desired term needs to be compared with the power of the unwanted sideband. The image rejection of this quadrature modulator was given in [29] and is equal to

$$IR_{QM} = 10 \log \left[\frac{1 + (1 + \alpha)^2 + 2(1 + \alpha) \cos(\theta)}{1 + (1 + \alpha)^2 - 2(1 + \alpha) \cos(\theta)} \right]. \quad (5.2)$$

The image rejection of the quadrature modulator sets the image rejection for direct conversion transmitters and conventional heterodyne transmitters. Although Equation (5.2) will give the level of image rejection, it does not provide much insight into the effect on the transmitted signal.

To better understand how gain mismatch and phase errors effect the transmitted signal a constellation diagram is often useful. The effect of phase error and gain mismatch and on the constellation of a GMSK modulated signal was shown in Figures 2.9 and 2.13 respectively. A more detailed view of the effect phase error on one point in the constellation was illustrated in Figure 2.10. Notice that a phase error in the LOs does not create a phase error in the constellation point but it does cause a magnitude error. The point on the constellation either moves closer to the origin or farther away depending on the sign of the phase error and the quadrant of interest. Looking back at the GMSK constellation with pure phase error it is now evident how the constellation becomes distorted. If the first quadrant is defined as the upper right quadrant, and rotating counterclockwise moves to the second, third and fourth quadrants, the points in the first and third quadrants moved farther away from the origin while the points in the second and fourth quadrant moved closer in.

In contrast to a phase error, amplitude mismatch in the baseband input causes both a phase and magnitude error in the transmitted signal, which is clear from the GMSK constellation shown in Figure 2.13. Understanding how the points move on the constellation will be helpful with regard to the affect of phase error in the HRT.

To model the effects of gain mismatch and phase error in the HRT, phase errors in both LOs need to be accounted for as well as gain mismatch. Shown in Figure 5.3 is such

a model that includes a baseband gain mismatch, α , LO₂ phase error, θ_2 , and LO₁ phase error, θ_1 . This transmitter model essentially consists of two quadrature modulators: one to generate an in-phase IF signal given by I_{IF} and one to generate a quadrature phase IF signal, Q_{IF} . The I and Q IF signals are then up-converted by an RF image reject mixer, which attenuates the unwanted IF image. However, this second up-conversion also affects the baseband image rejection as will be shown.

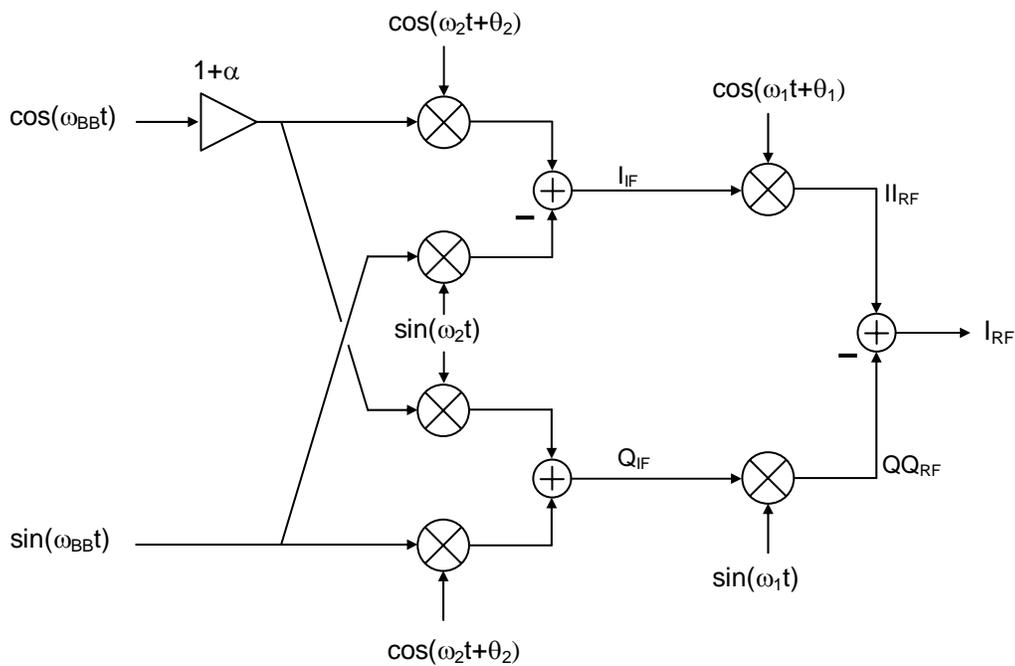


Figure 5.3 Model used to evaluate image rejection in the HRT.

A complete derivation of the baseband image rejection is given in Appendix A.

The baseband image rejection for the HRT is given by

$$IR_{HRT} = 10 \log \left[\frac{1 + \cos(\theta_1 - \theta_2) + (1 + \alpha)^2 (1 + \cos(\theta_1 + \theta_2)) + 2(1 + \alpha)(\cos \theta_1 + \cos \theta_2)}{1 + \cos(\theta_1 - \theta_2) + (1 + \alpha)^2 (1 + \cos(\theta_1 + \theta_2)) - 2(1 + \alpha)(\cos \theta_1 + \cos \theta_2)} \right] \quad (5.3)$$

where gain mismatch and phase errors in both LOs have been included. Although this expression does not provide much intuition about the behavior of the HRT, it has many similarities to the image rejection of a quadrature modulator shown in Equation (5.2). A plot of the image rejection of the HRT as function of θ_2 is shown in Figure 5.4 with θ_1 fixed at two degrees and for two different values of gain mismatch, α : zero and 0.01. Also shown is the image rejection of a quadrature modulator as a function of phase error for the same values of gain mismatch. The levels of image rejection for HRT are considerably higher than the quadrature modulator for equal levels of phase error and gain mismatch. For example, with two degrees of phase error, the image rejection of quadrature modulator is approximately 35 dB for both values of gain mismatch while the HRT achieves approximately 46 dB with one percent gain mismatch and 70 dB with perfect gain matching. Notice that the image rejection is double in dB for the case of perfect gain matching. The reason for this will be apparent later in this section. It also interesting to note that for case with the one percent gain matching, the image rejection of the HRT does not change significantly as the phase error of the LO_2 increases. This is a very significant observation because it implies that the HRT is insensitive to phase mismatch and thus can potentially achieve better image rejection, and thus better overall performance, with large variations in the phase error. Because the phase error of LO_1 is fixed in this case, it remains to be seen if the image rejection is also insensitive to variations in LO_1 phase error.

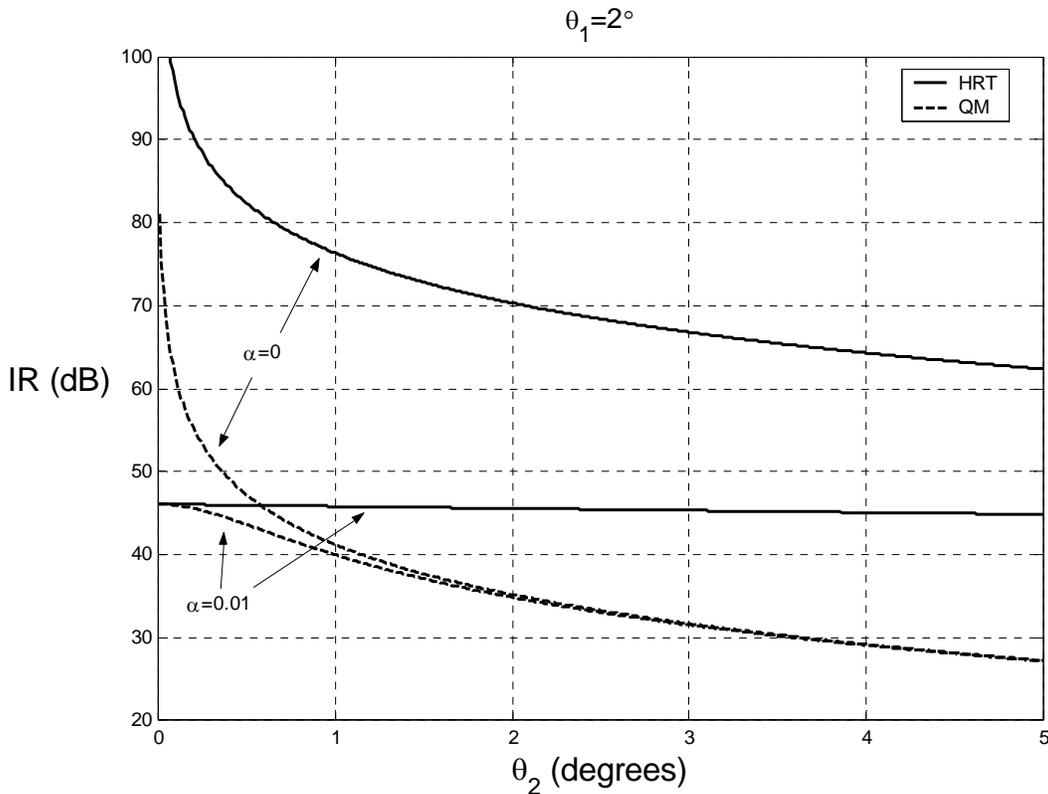


Figure 5.4 Image rejection for HRT and quadrature modulator as function of θ_2 for two values of α and with $\theta_1=2^\circ$.

The image rejection of the HRT is plotted as function of both θ_1 and θ_2 in Figure 5.5 for one percent and five percent gain mismatch. From the contours it is clear that the image rejection is fairly flat in both cases indicating that it is insensitive to phase mismatch in both LO_1 and LO_2 . The image rejection is a stronger function of gain mismatch than phase error for these values of gain mismatch. Although these results indicate that the image rejection of the HRT is superior to a quadrature modulator, Equation (5.3) does not explain at an intuitive level how this occurs. To help with this, the image rejection will be examined with respect to phase error with no gain mismatch

and then with respect to amplitude mismatch with no phase error. This will isolate the effects on the output signal caused by each source of error.

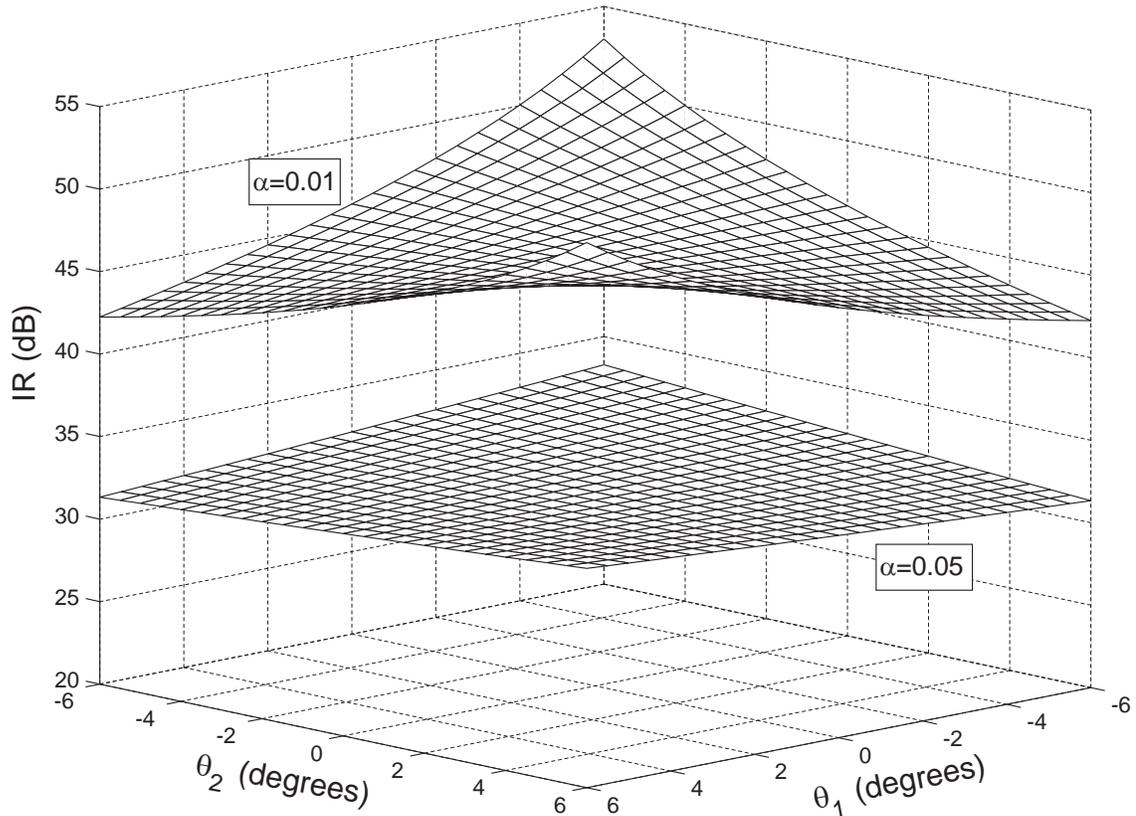


Figure 5.5 Image rejection for HRT as function of θ_1 and θ_2 for two different values of α .

A better understanding of the behavior of the HRT with respect to phase error in LO_2 is possible by examining the constellation diagrams at various points in the HRT. To isolate the effects of LO_2 phase mismatch, the gain mismatch and the phase error in LO_1 set to zero. Furthermore, for the sake of clarity, the phase error in LO_2 is assumed to be evenly divided between the I and Q LO signals. Shown in Figure 5.6 are the constellation diagrams at five points in the HRT. Figures 5.6a and 5.6b show the

constellation diagrams, along with the ideal points, for the I and Q IF signals. The constellations are identical because the errors stem from the LO signal and not the baseband signal. The IF signals are then up-converted to RF with a mixer. Because the in-phase IF signal is multiplied by an in-phase LO_1 the constellation diagram of the resulting signal is unchanged. This constellation shown in Figure 5.6c. The quadrature-phase IF signal is multiplied with a quadrature-phase LO_1 signal and this rotates the constellation by 90 degrees as shown in Figure 5.6d. It is worth noting that the down-converted terms in both I_{RF} and Q_{RF} have been ignored. When the two RF signals are then summed at the output, as shown in Figure 5.6e, the errors in the constellation effectively cancel. This explains the insensitivity to of the image rejection to a phase error in LO_2 because the errors get cancelled. A similar effect can be seen from phase errors in LO_1 .

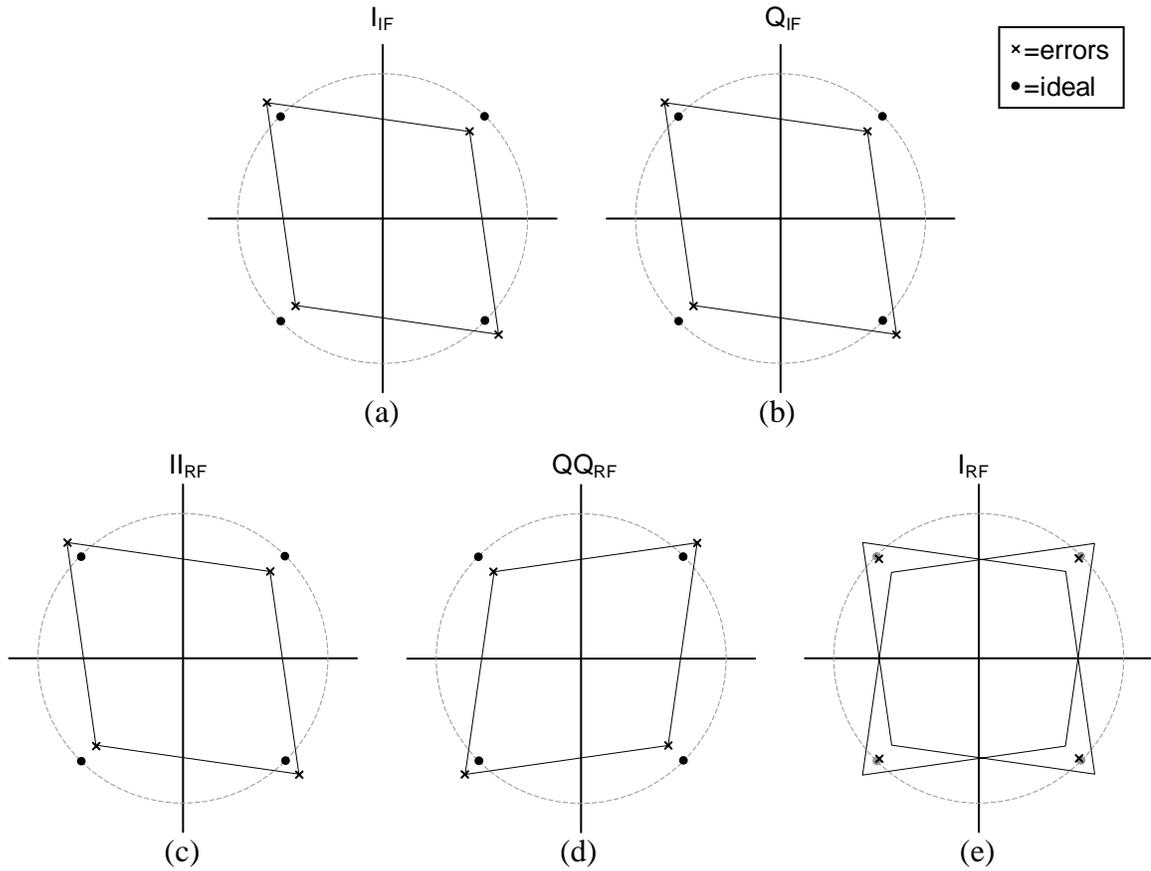


Figure 5.6 Constellation diagrams illustrating the effect of LO_2 phase error at various points in the HRT. (a) I_{IF} . (b) Q_{IF} . (c) I_{RF} . (d) Q_{QRF} . (e) I_{RF} .

If a phase error exists in both the LO_2 and LO_1 signals, the constellation point for the upper right quadrant of the output signal is shown in Figure 5.7. The diagram shows the constellation points from I_{RF} and Q_{QRF} for the case with LO_1 and LO_2 phase error and the case with only LO_2 phase error. The points marked by circles are a result of LO_2 phase errors and these points are identical to the results from Figure 5.6. An additional phase error in LO_1 rotates these points which results in the “x” constellation points. When the signals are summed, the resulting output is very close to the ideal constellation point indicating a small error and large image rejection. It is clear from this figure that

the phase errors tend to cancel in the HRT resulting in superior image rejection with respect to LO phase errors.

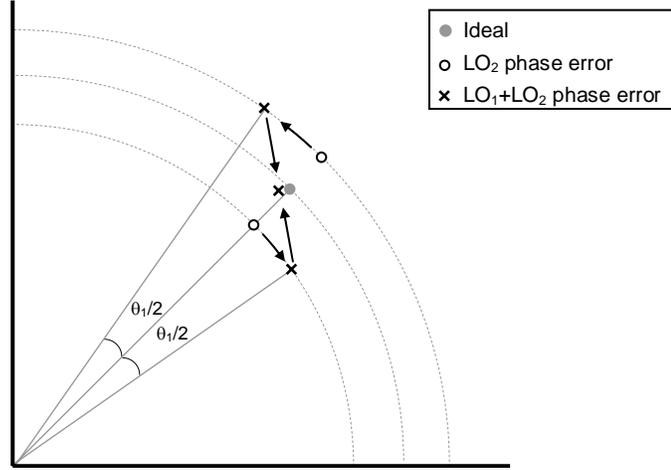


Figure 5.7 Constellation diagram of the HRT output in one quadrant with LO_1 and LO_2 phase errors.

The effect of phase errors on the HRT can be quantified by evaluating Equation (5.3) for the case with no gain error. This results in the following:

$$\begin{aligned}
 IR_{HRT} \Big|_{\alpha=0} &= 10 \log \left[\frac{1 + \cos(\theta_1 - \theta_2) + (1 + \cos(\theta_1 + \theta_2)) + 2(\cos \theta_1 + \cos \theta_2)}{1 + \cos(\theta_1 - \theta_2) + (1 + \cos(\theta_1 + \theta_2)) - 2(\cos \theta_1 + \cos \theta_2)} \right] \\
 &= 10 \log \left[\frac{(1 + \cos \theta_1)(1 + \cos \theta_2)}{(1 - \cos \theta_1)(1 - \cos \theta_2)} \right].
 \end{aligned} \tag{5.4}$$

As a point of reference, the image rejection of the quadrature modulator with no gain error can be determined by evaluating Equation (5.2), which results in

$$IR_{QM} \Big|_{\alpha=0} = 10 \log \left[\frac{1 + \cos \theta}{1 - \cos \theta} \right]. \tag{5.5}$$

Notice that the image rejection of the HRT is determined by the product of the terms from the two LOs, which results in significantly higher image rejection. Comparing the

two results, if the phase error in the quadrature modulator is equal to the phase error in LO_1 and LO_2 , the image rejection in the HRT is double in dB that of the quadrature modulator. This result agrees corroborates the effects of phase error on the transmitted constellation seen in Figure 5.7. Furthermore, if either quadrature generation of LO_1 or LO_2 is very accurate, the image rejection will be high assuming the gain mismatch is small.

The affect of gain mismatch on the HRT is also clarified with constellation diagrams. Constellations in the HRT resulting from gain mismatch with no phase mismatch are illustrated in Figure 5.8. The constellations for I and Q IF signals are shown in Figure 5.8a and 5.8b and unlike the case with pure phase mismatch, the constellations are not identical. This occurs because mismatch on the baseband I and Q signals are mapped to different axis in the I/Q plane in the I_{IF} and Q_{IF} signals. The I_{IF} signal is up-converted but unchanged in Figure 5.8c but up-conversion by the Q LO_1 signal rotates the Q_{IF} constellation by 90 degrees as shown in Figure 5.8d. Finally, when the signals are summed and normalized, the resulting constellation, Figure 5.8e, has the same shape as I_{IF} . This indicates that the HRT does not improve the image rejection with respect to gain mismatch like it did for phase error.

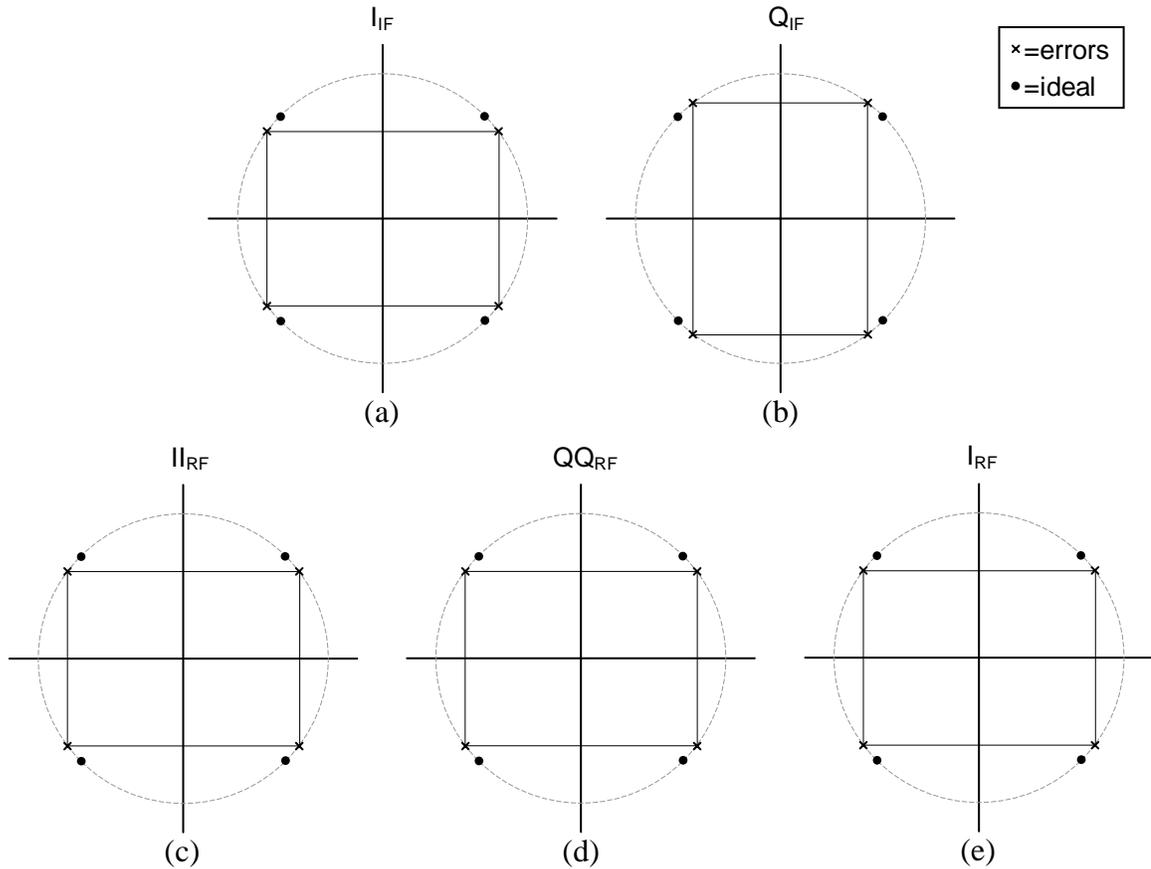


Figure 5.8 Constellation diagrams illustrating the effect of gain mismatch at various points in the HRT. (a) I_{IF} . (b) Q_{IF} . (c) I_{RF} . (d) $Q_{Q_{RF}}$. (e) I_{RF} .

Although the constellation diagrams have provided insight into the effects of a phase error or a gain mismatch they are not very helpful when both effects are considered together. Figure 5.9 shows the HRT with phasor representations of the input, output, LOs and intermediate signals. The gain mismatch is shown in green while the phase error is shown in red. In this way, the effects of the phase error and gain mismatch can be tracked through the HRT to the output. Notice that the image at the output is purely caused by the gain mismatch and the phase mismatch has been canceled, which agrees with the earlier conclusions. In this diagram the phase error of LO_1 is zero.

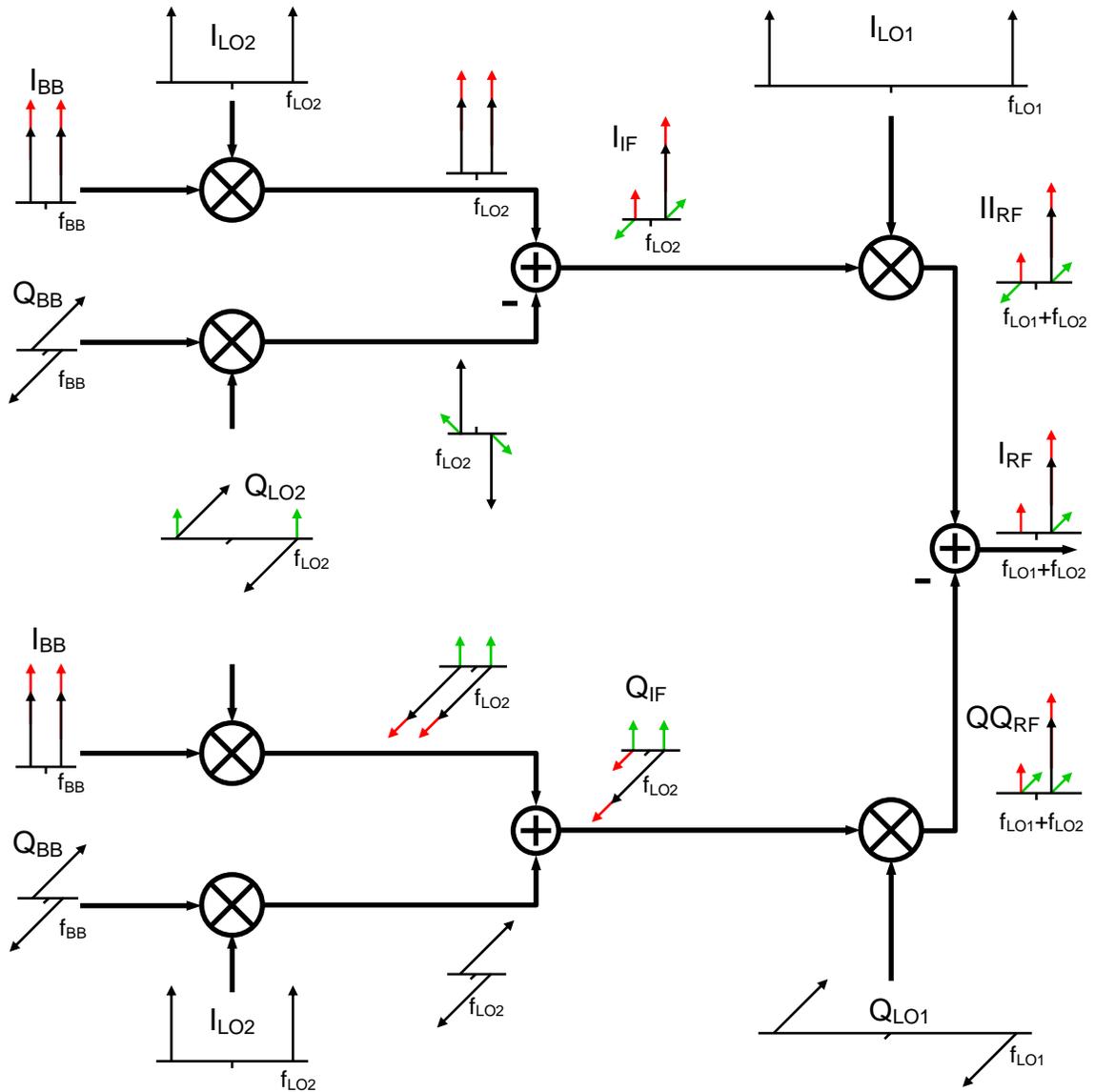


Figure 5.9 Phasor diagram of HRT showing the effects of gain mismatch and LO₂ phase error.

In summary it has been shown that the image rejection of the HRT is better than a single quadrature modulator and thus better than the image rejection of a direct conversion or conventional heterodyne transmitter. Therefore, because the modulation accuracy is reflected in the image rejection, the performance is also substantially improved. The image rejection of a transmitter is generally a function of phase error and

gain mismatch. In the HRT, the effects of phase error on the image rejection have been reduced because the phase errors tend to cancel at the output. This can be seen in Equation 5.4 where it was shown that for the case of zero gain mismatch, the image rejection is determined by the product of the phase error terms from each LO. This results in significantly better image-rejection when compared to a single quadrature modulator in this case. Because the phase errors tend to cancel in the HRT, to even further improve the performance, minimizing gain mismatch will provide maximum benefit. The following section discusses a method to lower the gain mismatch, and thus capitalize on the reduced effects of phase error, with minimal additional complexity.

5.3.1 Amplitude mismatch tuning method

It was shown in the previous section that to further improve the image rejection performance of the HRT, correcting gain mismatch is a logical method. Gain mismatch in a quadrature modulator can be caused by mismatch in any of the baseband components including DACs, filters, and mixers. Although proper layout and large devices can reduce the effects, the benefits are limited. Alternatively, gain mismatch can be corrected through use of feedback. Through the use of feedback, the baseband input amplitudes can be adjusted to compensate for gain mismatch in the signal path. From a performance perspective, this is desirable but this can add considerable complexity to the transmitter, particularly if the feedback requires high frequency circuits.

Illustrated in Figure 5.10 is a quadrature modulator with two potential feedback paths shown in gray. The first path involves feeding back baseband signals at the output

of each filter. This is desirable from a complexity view point because the signals are low frequency, but it does not take into account gain mismatch in the mixers.

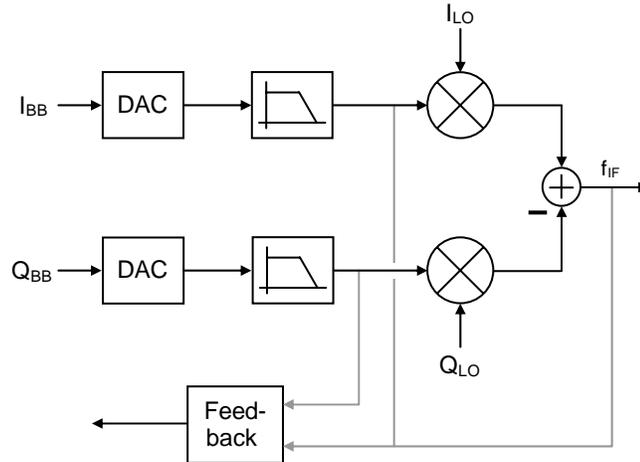


Figure 5.10 Feedback options for quadrature modulator.

The second feedback path occurs after the mixer outputs are summed. Although this path can correct for gain mismatch in the entire baseband, including the mixers, the signal is at a much higher frequency, which consequently adds to the additional complexity. To alleviate this tradeoff, the LO signal to the mixer can be changed.

In a typical Gilbert cell mixer the upper switching transistors are typically connected to the LO. A modified Gilbert cell mixer is illustrated in Figure 5.11 in which two of the switching transistors have been connected to ground and two to the supply voltage. The transistors whose gates are grounded are clearly off and these are shown in gray. By connecting the transistors in this way, the mixer is essentially converted to an amplifier which does not perform frequency conversion. This amplifier can then be incorporated into a low-frequency feedback loop which corrects for gain errors. One drawback to this method is that mismatch in the switches will not be corrected.

However, the conversion gain of Gilbert cell mixer is a relatively weak function of the switching transistor geometry and therefore the drawback is minimal.

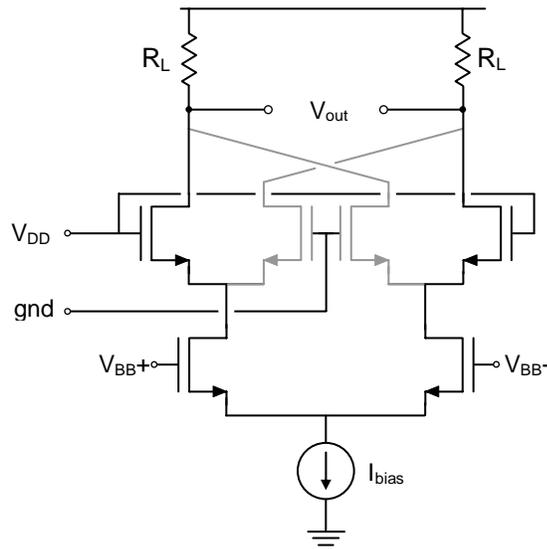


Figure 5.11 Amplifier converted from Gilbert cell mixer.

With the mixer converted to a gain stage, the feedback can now include the effects of gain mismatch in the mixers but without the need for high frequency feedback, which minimizes the additional complexity. This configuration is illustrated in Figure 5.12. To test the gain matching, identical sinusoidal inputs can be applied to the I and Q signal paths and this should produce no output because the output of the bottom path is subtracted from the top path. Any output that does occur at the baseband frequency is a result of gain mismatch in the two paths. This mismatch can then be compensated for by adjusting the amplitude of the baseband input signals.

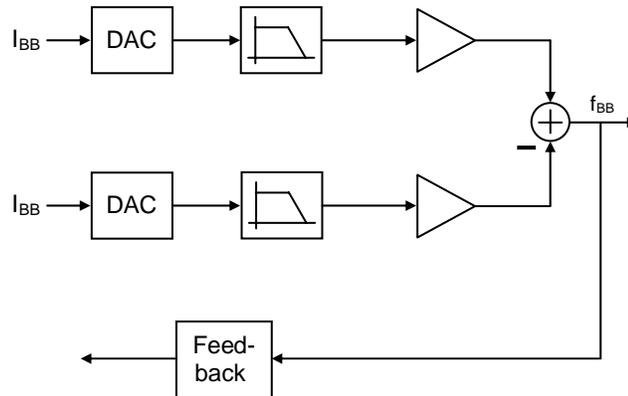


Figure 5.12 Low frequency feedback with amplifier in place of mixers.

This low frequency feedback technique can easily be applied to the HRT because each HRM consists of three Gilbert cell mixers. Consequently, by correcting for gain mismatch in the HRT, the performance of the HRT can be significantly improved. Another benefit of this tuning scheme is that DC offsets in the baseband can also be corrected with the same feedback loop and this can even further improve the performance.

5.4 Synthesizer Integration

Synthesizer phase noise is an important consideration in transmitter design because it can affect both the modulation accuracy and the unwanted spectral emissions. Phase noise close to the carrier can cause inaccuracy in the transmitted signals as was shown by the GMSK constellation in Figure 2.20. Furthermore, phase noise that is far from the carrier can violate spectral mask requirements and thus, to meet the spectral emissions requirements costly discrete filters are often necessary. Frequency synthesizer phase noise depends on a number of factors including the architecture of the transmitter. This

section will attempt to explain the rationale behind the frequency plan of HRT with respect to minimizing the phase noise in the frequency synthesizers when implemented in a CMOS process.

In the HRT the channel selection in the HRT is performed with the low frequency synthesizer while the RF synthesizer is fixed in frequency. In contrast both direct conversion and conventional heterodyne transmitters perform channel selection with an RF synthesizer. This approach was also taken in the wide-band IF receiver architecture [30]. Because there is no narrow band channel filtering in the HRT, both frequency synthesizers contribute phase noise to the output signal. Therefore, minimizing the total phase noise involves minimizing the sum of the phase noise from the two frequency synthesizers.

One critical factor in determining the phase noise of the frequency synthesizers is the quality of the passive components that are available. In a standard CMOS process, the Q of the inductors tends to be fairly low thereby raising the phase noise contribution of the VCO. This is evident by examining the SSB phase noise of the VCO as a function of offset frequency, which is given by [31]

$$L(\Delta f) = 10 \log \left[\frac{FkT}{2P_s} \left(1 + \left(\frac{f_0}{2Q\Delta f} \right)^2 \right) \left(1 + \frac{f_c}{\Delta f} \right) \right] \quad (5.6)$$

where Q is the quality factor of the tank, f_c is the 1/f noise corner, F is the noise factor, f_0 is the carrier frequency and P_s is the signal power. Clearly with everything else equal, lowering the Q of the tank will increase the phase noise of the oscillator. Therefore, when implementing a frequency synthesizer in CMOS, the phase noise contribution of the

VCO to the overall phase noise, can be very large and dominate the overall phase noise. Therefore, a technique is needed that can reduce the contribution of VCO phase noise to the overall phase noise of the frequency synthesizer.

One method to reduce the effects of VCO phase noise is through the use of a wide loop bandwidth PLL [32]. In this technique, the loop bandwidth of the PLL is increased to reduce the effects of VCO phase noise. The VCO phase noise contribution is reduced because the transfer function from the VCO to the output has a high-pass characteristic for second-order phase locked loops [33]. By increasing the loop bandwidth the VCO phase noise contribution is reduced and the overall frequency synthesizer is phase noise is lowered.

One significant drawback of the wideband PLL approach is the inability of the frequency synthesizer to perform channel tuning because of the high reference frequency that is needed [34]. As a result the wideband PLL approach is best used in a frequency synthesizer that does not perform channel tuning. This raises the logical question as to whether the wideband approach should be used with the IF or RF frequency synthesizer?

To answer the question as to which frequency synthesizer should use the wideband approach, recall that the total effect of frequency synthesizer phase noise in the HRT is the sum of the phase noise from the two frequency synthesizers. Therefore, to minimize the total phase noise, the wideband PLL technique should be applied to the frequency synthesizer that has inherently higher phase noise and the channel tuning should be performed by the frequency synthesizer with inherently lower phase noise. This choice is made easier with the understanding that the phase noise of the IF synthesizer is inherently lower than the RF synthesizer because of the lower operating

frequency [35]. Therefore, the IF synthesizer should be used for channel tuning and the RF synthesizer should utilize the wideband PLL approach.

The wideband PLL approach facilitates the integration of the frequency synthesizers in CMOS by suppressing the phase noise contribution of the VCO and thus allowing low phase noise with relatively low-Q passive components. Although the IF frequency synthesizer does not use the same approach, its phase noise is inherently lower due to the lower frequency of operation.

5.5 Summary

This chapter presented the Harmonic Rejection Transmitter architecture. This architecture uses the HRM to facilitate integration while achieving high performance. Furthermore, the architecture is amenable to multi-standard operation because of the inherently flexible mixer based approach. It was also shown that the performance of the transmitter with respect to image rejection was substantially improved compared with a single quadrature modulator, which determines the image rejection in direct conversion or conventional heterodyne transmitter. A tuning method was presented which allows for even higher image rejection performance while only requiring low-frequency feedback. Finally, the advantages of the HRT with respect to synthesizer integration were discussed.

Chapter 6

Transmitter Prototype

6.1 Introduction

The previous chapters have focused on the techniques and ideas that might be involved in designing a high performance CMOS transmitter with the potential for multi-standard capability. This chapter shifts the focus to a prototype transmitter that was implemented to test some of the ideas previously presented. The transmitter was designed to meet the specification of DCS1800, which is an upbanded version of GSM. This cellular telephony standard was chosen because it requires high performance in terms of modulation accuracy and very low spectral emissions, particularly in the receive band. Furthermore, the narrow channel spacing requires an agile frequency synthesizer for

channel tuning purposes. Therefore, the potential for multi-standard operation exists because most other RF standard use wider bandwidth channels. DCS1800 was also chosen because it employs constant envelope modulation making well suited for use with and highly non-linear integrated power amplifier.

A block diagram of the HRT prototype is illustrated in Figure 6.1. The prototype integrated the entire transmit path from the digital-to-analog converters through the power amplifier. This IC contains four HRMs to perform the up-conversion to IF and two current-commutating RF mixers. In addition to the DACs, the baseband also consists of two low-pass reconstruction filters. Both frequency synthesizers were also integrated on the same test chip as well as a divide-by-four circuit to generate the phases needed by the HRMs and an RF quadrature generation circuit to generate the quadrature LO signals needed by the RF image-reject mixers. The transmitter includes a class-C PA and an RF testing buffer. The prototype IC was fabricated in a 0.35- μm double-poly, five-layer metal CMOS process.

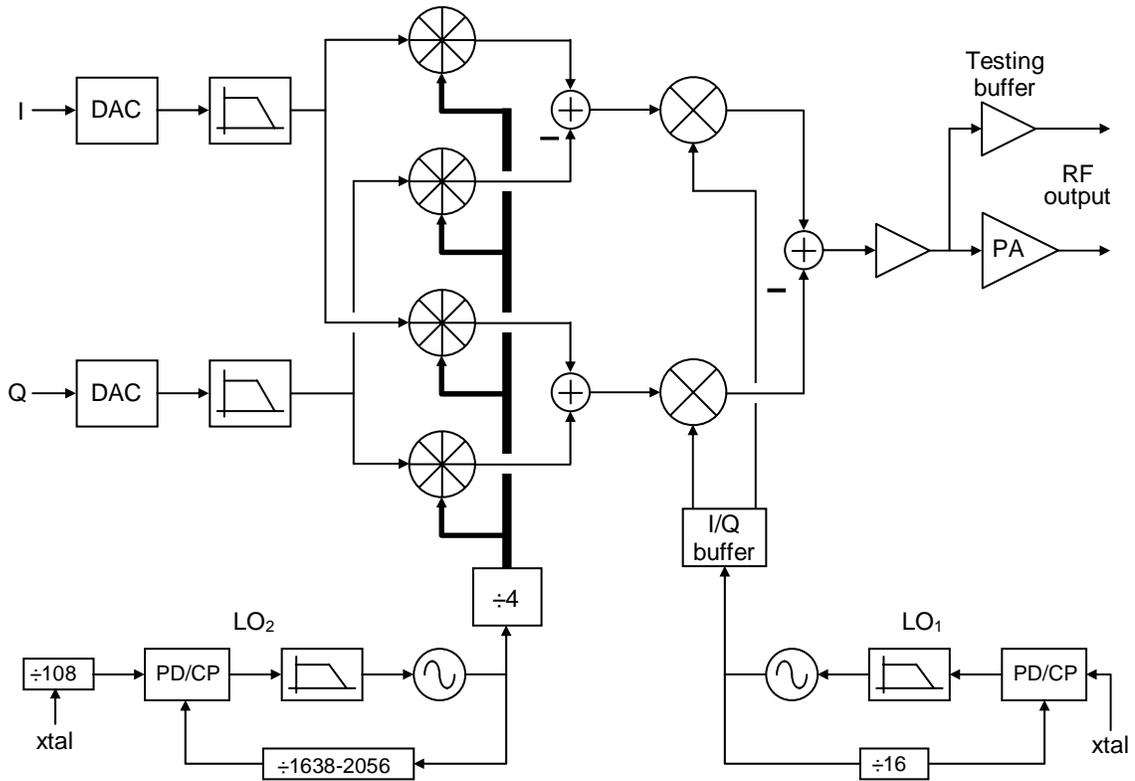


Figure 6.1 HRT prototype block diagram.

The following section will discuss the image rejection and carrier feedthrough requirements for DCS1800 and Section 6.3 will focus on the HRMs and the RF mixers. Section 6.4 will discuss some of the other circuits implemented in this prototype and Section 6.5 will discuss steps that were taken to reduce on-chip signal coupling. Section 6.6 will discuss the layout of the actual IC and Section 6.7 will summarize the chapter.

6.2 Image Rejection and Carrier Feedthrough

The performance of the quadrature modulator directly affects the modulation accuracy of the transmitted signal. For the DCS1800 standard modulation accuracy is defined by

RMS and peak phase error because DCS1800 employs GMSK modulation, which is a form of phase modulation. The DCS1800 standard requires that the RMS phase error is less than five degrees while the peak phase error below 20 degrees. Therefore, in the design of the transmitter the keeping each source of phase error below 0.5 degrees of RMS phase error was the goal.

A critical test which reveals the performance of the quadrature modulator is the single-sideband test in which the baseband signal is a tone as opposed to a modulated signal. This topic was introduced in Chapter 2 but the exact performance specifications were not addressed. In a single-sideband plot, the most important results are the levels of the unwanted sideband, the carrier feedthrough and the third order distortion. The third order distortion is typically dominated by the baseband input of the mixers and this will be discussed in the next section.

The unwanted sideband is function of gain error and phase mismatch in the quadrature modulator. Although phase errors in the quadrature modulator contribute to the unwanted sideband in a single-sideband test, in the case of GMSK they have little effect on the RMS phase error. This is evident by inspection of Figure 2.10 where it is shown that a phase error tends to move the constellation points either in or out. Although this movement will affect the error vector magnitude (EVM), it will have little affect on the phase error. Furthermore, it was shown in Section 5.3 that the image rejection of the HRT is a much stronger function of gain mismatch than phase error. Therefore, to determine the specifications for the image rejection in the HRT, it will be assumed that the image is purely caused by gain mismatch and no phase error. In practice a portion of the image will be caused by phase error and therefore determining the image rejection

requirement based only on gain mismatch represents a worst case scenario. A plot of the RMS phase error as function of the image rejection in a single sideband test is shown in Figure 6.2. In this case, the image is caused purely by a gain mismatch. To keep the RMS caused under 0.5 degrees, the image rejection needs to be larger than approximately 42 dB. In designing the transmitter, this level of image rejection provides a bound for tolerable gain error. Since the gain error is primarily caused by device mismatch, the matching requirements are also determined by the tolerable image rejection.

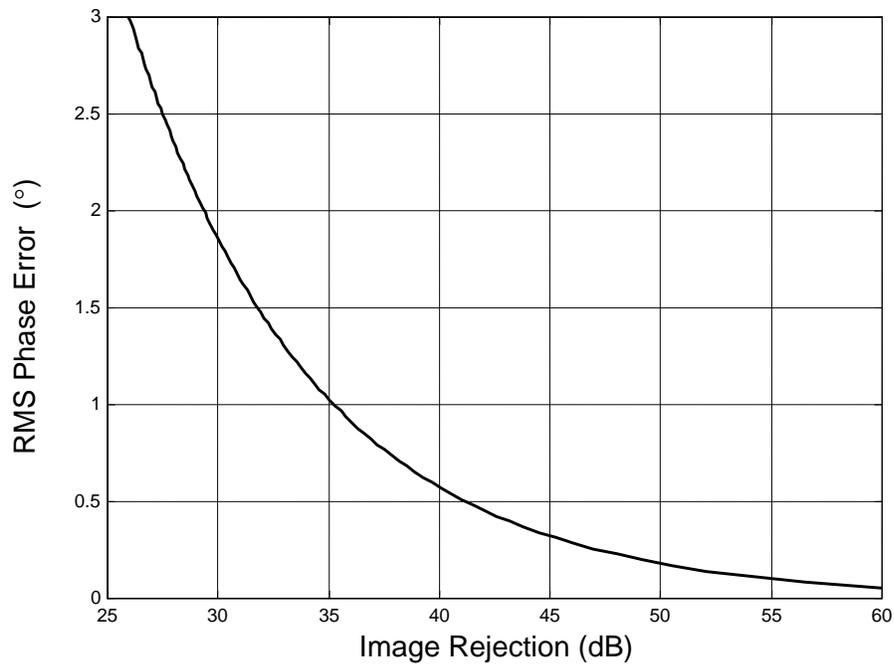


Figure 6.2 Simulated RMS phase error as a function of image rejection from only gain mismatch in a GMSK signal.

The carrier feedthrough is typically dominated by DC offsets in the baseband and LO feedthrough in the mixers. A plot of the simulated RMS phase error of GMSK modulated signal versus carrier feedthrough is shown in Figure 6.3. From this plot it is

evident that to keep the phase error contribution of the carrier feedthrough below 0.5 degrees, the carrier feedthrough needs to be less than approximately -38 dB.

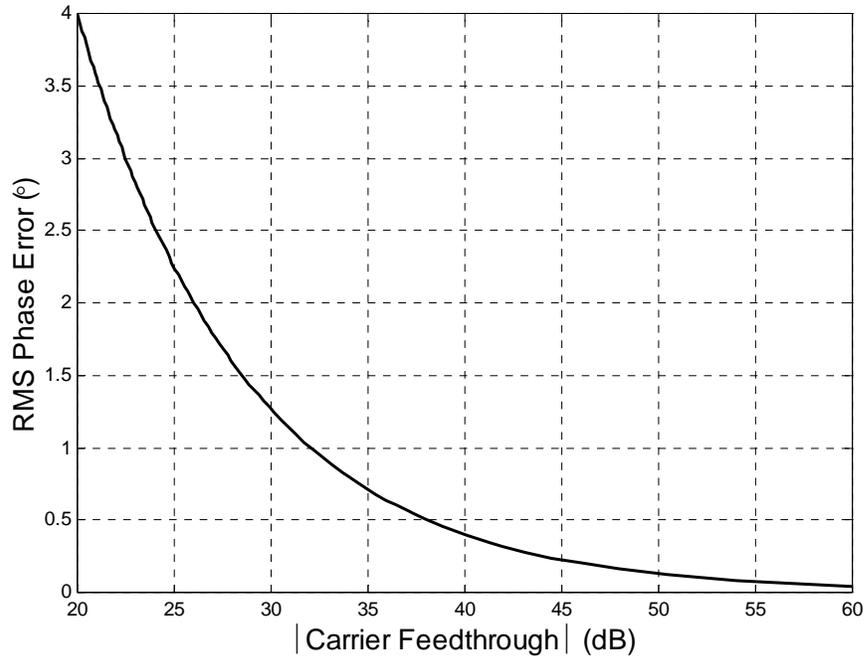


Figure 6.3 Simulated RMS phase error as a function of carrier feedthrough for GMSK signal.

6.3 Mixers

A radio transmitter performs three critical functions: modulation, up-conversion, and amplification. In the HRT mixers play an important role in all three of these functions. Mixers are used in the quadrature modulators to both modulate and up-convert the signal to IF. Mixer are also used to the frequency translate the IF signal to RF where it can be amplified with the PA. The performance of the mixers will affect the constraints on the PA. In a constant envelope modulation scheme AM resulting from the non-ideal

behavior of the mixers and other circuits will affect the linearity requirements of the PA and therefore its efficiency. Clearly mixers play a very important role in the HRT. This section will focus on some of the design and implementation issues of the HRMs and the mixers used for up-conversion to IF.

6.3.1 HRM Circuit

Although Chapter 4 included a detailed discussion of the harmonic-rejection mixers, the actual implementation is the focus of this section. Because each HRM consist of three Gilbert cell mixers in parallel, many of the design issues can be addressed with well known design trade-offs of Gilbert cell mixers. A circuit diagram of the HRM is illustrated in Figure 6.4 as well as a table of transistor sizes. The remainder of this section will examine some of the design issues of the HRT.

distortion in the baseband. This plot indicates that to keep the contribution of RMS phase error under 0.5 degrees, the third harmonic distortion should be close to -40 dB.

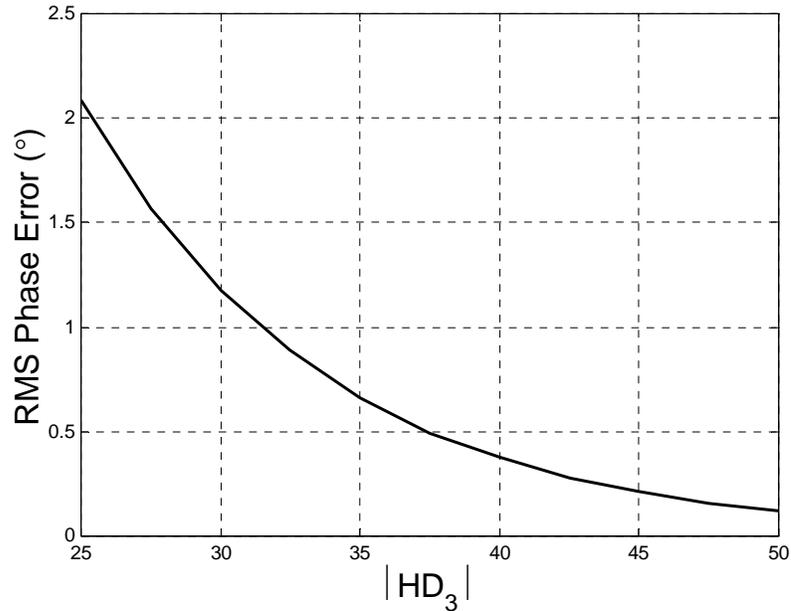


Figure 6.5 Simulated RMS phase error as a function of $|HD_3|$ for a GMSK signal.

The phase errors are not the only negative affect of the distortion. Baseband non-linearity can also cause spectral regrowth. To confirm that an HD_3 level of -40 dB is sufficient, Figure 6.6 shows the GMSK spectrum of an ideal signal and a signal with an HD_3 level of -40 dB. It is clear from the figure that for this level of distortion the affect on the spectral regrowth is minimal and is therefore acceptable for the baseband of the transmitter.

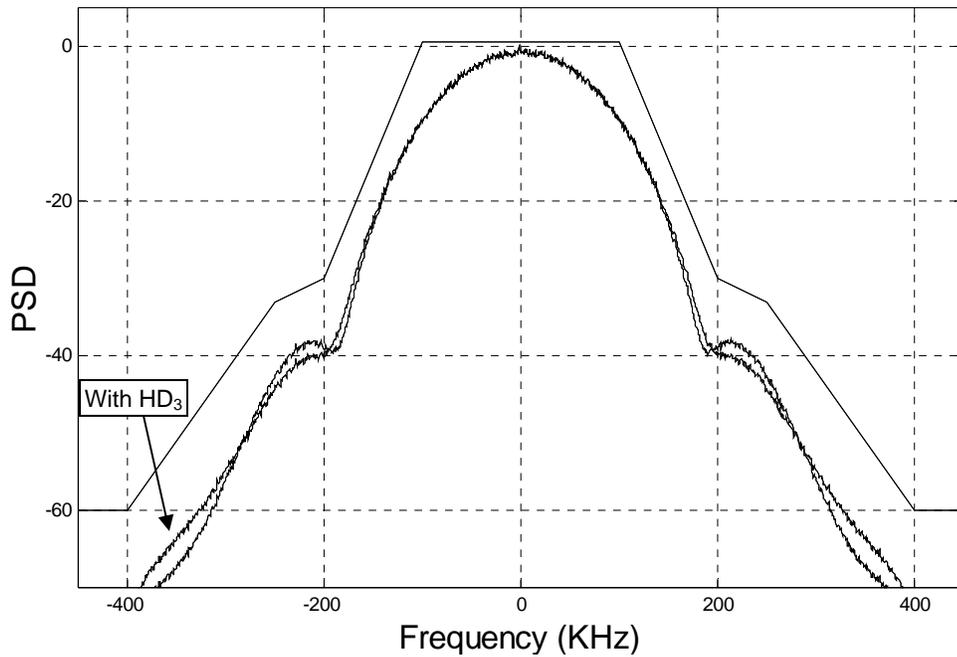


Figure 6.6 GSM spectrum with HD_3 level at -40 dB.

The linearity of the baseband section, which includes a DAC, a filter and the HRM, is typically dominated by the linearity of mixer input devices. Therefore it is critical that the input section is designed correctly. The linearity of the input differential pair was given in Equation (4.39). This equation shows that the third order non-linearity is a function of the input signal amplitude and the value of $(V_{GS}-V_t)$ for the input devices. Applying Equation (4.39) along with an HD_3 level of -40 dB results in the following relationship between the input signal amplitude and the value of $(V_{GS}-V_t)$ of the input devices:

$$V_{GS} - V_t = 1.77V_{in}. \quad (6.1)$$

For example, a 0.5 volt input signal would require that $(V_{GS}-V_t)$ to be 0.9 volts. This is quite large and thus the linearity of the input can limit the amplitude of the baseband signal.

Clearly a large input voltage amplitude is desired so that less gain is needed later in the transmitter, which potentially reduces the overall power consumption. However, this requires a large value of $(V_{GS}-V_t)$ to achieve the desired linearity. One method to improve linearity is through the use source degeneration resistors. While the linearity does improve, the use of source degeneration resistors has some drawbacks with respect to matching. Good matching between the sub-mixers of the HRM is critical for good performance with respect to the rejection of the third and fifth LO harmonics. Source degeneration degrades matching because it lowers the value of $(V_{GS}-V_t)$ and threshold voltage mismatch is more significant. Furthermore, the mismatch between the degeneration resistors also degrades the matching between the sub-mixers.

Given that a large $(V_{GS}-V_t)$ is needed, this can be obtained by either raising the current or using smaller (W/L) ratios for the input devices. More current is clearly beneficial to performance but minimizing the current is always recommended. Therefore, the use of smaller ratios is an appropriate method to increase the gate-source bias voltage. However, smaller device sizes increases the mismatch due to device geometry mismatch. Therefore, in the implementation of the HRM the (W/L) ratio of the input devices was lowered by increasing the length of the input devices. This allowed for a large $(V_{GS}-V_t)$ and also kept the device dimensions large for improved matching. Using a longer channel length is detrimental with regard to the speed of the transistors but in this case the input transistors are only operating with relatively low-frequency baseband signals.

To reduce the noise of the input devices and to increase the conversion gain of the mixer, it is beneficial to increase the g_m of the input devices. A large g_m requires a large current which in turn reduces the size, and consequently the noise contribution, of the load resistor. Consequently, the desired noise performance and conversion gain will determine the bias current of the mixer.

While the input devices needed a large $(V_{GS}-V_t)$, the requirements on the switching transistors are somewhat different. The gain of a Gilbert cell mixer was given in Equation (4.25). From this equation it is clear that raising the LO voltage relative to the value $(V_{GS}-V_t)$ in the switches will increase the gain. Furthermore, increasing the ratio of $V_{LO}/(V_{GS}-V_t)$ reduces the noise contribution from the switching transistors because they switch faster and thus are in the noise contributing transition zone for a shorter time. Increasing the ratio of $V_{LO}/(V_{GS}-V_t)$ can be accomplished by lower the value of $(V_{GS}-V_t)$ for the switching transistors. To lower the $(V_{GS}-V_t)$ of the switching transistors for a given current, large (W/L) ratios are needed. However, larger devices add load capacitance to the divide-by-four circuit which in turn increases their power consumption. Furthermore, lowering the value $(V_{GS}-V_t)$ too much and thus further raising the ratio of $V_{LO}/(V_{GS}-V_t)$ can degrade the performance with respect to linearity [21]. Therefore, the size of the switching transistors is dependent on the LO swing and the bias current in the mixer and the transistors should be sized such the ratio of $V_{LO}/(V_{GS}-V_t)$ is neither too large nor too small.

In the implementation of the HRM, the (W/L) ratio of the switches is considerably larger than the input devices thereby lowering the $(V_{GS}-V_t)$ of these devices. Furthermore, minimal channel lengths were used to minimize the capacitive loading on the divide-by-

four circuit. Device matching in the switches is less critical than the input device because the conversion gain of the mixer is a relatively weak function of the switch size.

The value of the load transistor was largely determined by the bias current. A larger load resistor would clearly increase the conversion gain of the mixer. However, the size of the resistor is limited by the DC voltage drop across the resistor caused by the bias current. The larger the resistor, the lower the bias point at the output and as a result, the output signal may begin to distort when clipping occurs or devices leave saturation. Consequently, assuming the noise contribution from the resistor is manageable, the largest resistor possible is desired provided that bias point can accommodate the output swing. In the implementation of the HRM, three resistors connected in parallel were used for each load. This was done to improve the matching between the sub-mixers and thus improve the rejection of the third and fifth harmonics.

6.3.2 SNR in HRMs

The signal to noise ratio at the output of the HRM as compared to a standard mixer is important when designing a transmitter based on the HRM. Noise is important primarily because it can cause spectral mask violations. To compare the noise of an HRM with a standard mixer, it is critical to make a fair comparison. The HRM consists of three sub-mixers which are driven with identical input signals but the phases of the LO signals are not identical.

To make the comparison the first thing to understand is the effect of multiple parallel mixers with respect to noise. Illustrated in Figure 6.7a is a single mixer with

device sizes (W/L) and current I and ideal switches. Figure 6.7b shows two mixers in parallel where the device size and the current are halved compared to the single mixer. By scaling both the current and the device sizes, the value of $(V_{GS}-V_t)$ is unchanged and therefore the g_m of all of the devices is reduced by a factor of two. To simplify the analysis it is assumed that the switches are ideal and contribute no noise and furthermore only thermal noise is considered.

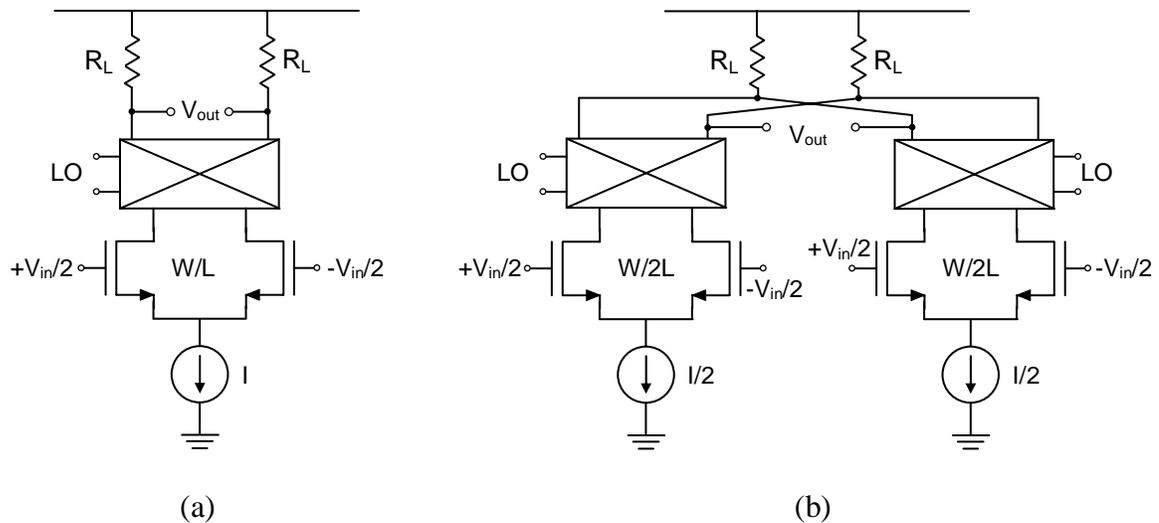


Figure 6.7 Mixers used for comparison. (a) Single mixer with ideal switches. (b) Parallel mixers with ideal switches.

The desired signal and the mean square noise at the output of the two mixer configurations will be compared. The conversion gain of the first mixer is given in Equation (4.24) and is equal to $g_m R_L / \pi$. Therefore the output of the single mixer is given by

$$V_{out} = V_{in} g_m R_L \frac{4}{\pi} \cos(\omega_{LO} t). \quad (6.2)$$

The conversion gain of each of the mixers in parallel is half the gain of the single mixer because the g_m of the input device was reduced by a factor of two. Therefore the signal output of the parallel mixers is identical to the output from the single mixer.

Considering only thermal noise, the output mean-square noise current from the input devices in the single mixer case is given by

$$\overline{i_o^2} / (\text{Hz}) = 4kt\gamma g_m \kappa . \quad (6.3)$$

where the factor κ is included to model the effects of noise folding caused by the harmonics of the LO. By inspection of Equation (6.3) it is clear that the mean square noise from each mixer in the parallel case is also reduced by a factor of two because of the g_m reduction. Furthermore, the noise from each mixer is uncorrelated and therefore the total noise can be determined by summing the mean-square noise from each mixer. Therefore, when the outputs of the two mixers are summed in the resistor, the resulting noise voltage is identical to the noise voltage produced by the single mixer because both mixers have the same load. The noise from the resistive loads is also identical because the loads are equal.

These results are not unexpected given the current is equal and the total device size is also equal. The first single mixer was essentially cut in half to form the two parallel mixers. These results do not change when the switch noise is included provided the size of the switching transistors is also scaled.

These results, indicating that parallel mixers have the same noise performance as a single mixer, can now be carried over to the HRM. Illustrated in Figure 6.8 are the circuits used for comparing a HRM with a single mixer. As a point of reference the

output noise and the output signal of HRM will be compared to a single mixer in which both the total current and the total device sizes are equal. Figure 6.8a shows a single mixer with current and device sizes that are scaled by a factor of $(2 + \sqrt{2})$. Figure 6.8b shows a HRM in which the total current and device sizes are equal to that of the single mixer. However, one important difference between the HRM and the parallel mixer case presented in Figure 6.7 is the phase of the LO signals.

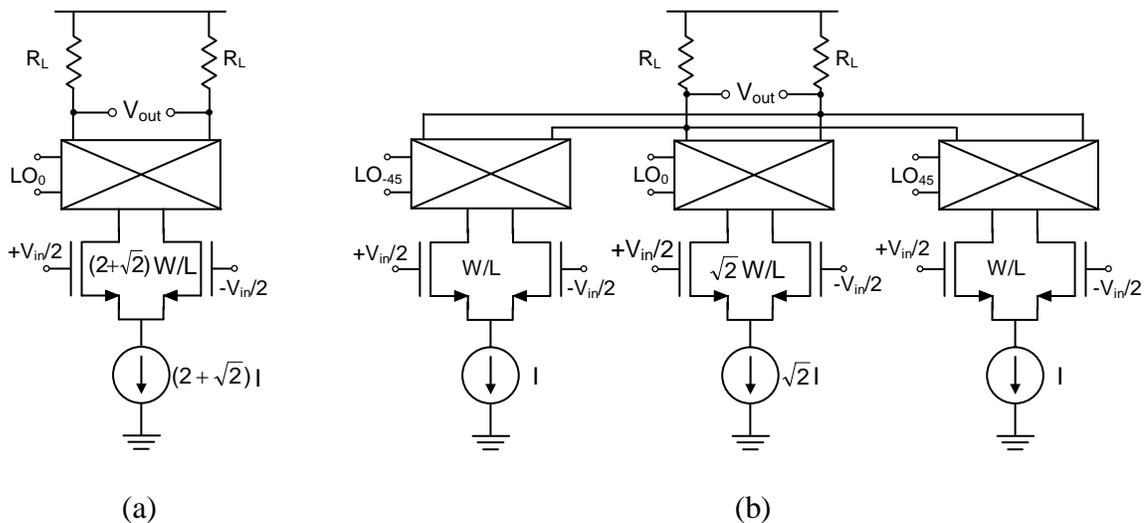


Figure 6.8 Mixers used for comparison. (a) Single mixer. (b) HRM.

The noise at the output of the HRM consists of noise from three primary sources: the input devices, the switching devices and the load. The phase of the LO has no effect on the noise contribution from any of these sources. Therefore the total noise in the HRM is equal to the noise in the comparison mixer. However, the amplitude of the output signal is not equal in these two cases.

The output signal for the single mixer was given by Equation (4.25) where the harmonics created in the mixing process have been ignored. This equation can be rewritten as follows

$$V_{os}(t) = V_{in}(t)A_s \cos(\omega_{LO}t) \quad (6.4)$$

where A_s is given by

$$A_s = \frac{g_m R_L \sqrt{2} V_{LO}}{\pi(V_{GS} - V_t)_{sw}} \sin\left(\frac{\sqrt{2}(V_{GS} - V_t)_{sw}}{V_{LO}}\right). \quad (6.5)$$

The device sizes and the current source in the single mixer in Figure 6.8a have been scaled by a factor of $(2 + \sqrt{2})$ and therefore, the values of $(V_{GS} - V_t)$ are unchanged. As a result, A_s is only affected by changes in g_m , which also scales by $(2 + \sqrt{2})$. Therefore the output signal of the single mixer in Figure 6.8a is given by

$$V_{os}(t) = V_{in}(t)(2 + \sqrt{2})A_s \cos(\omega_{LO}t). \quad (6.6)$$

The output signal of the HRM can also be determined by Equation (6.4) except now the phase of the LO must be accounted for. The phase of the middle sub-mixer is identical to the phase of the single mixer. Therefore the output signal created by the middle sub-mixer is given by

$$V_{om}(t) = V_{in}(t)\sqrt{2}A_s \cos(\omega_{LO}t). \quad (6.7)$$

The effect of the phase difference in the LO signal in sub-mixers that are driven by LO_{45} and LO_{45} is to reduce the gain contribution by a factor of $\sqrt{2}$. This is evident by

inspection of Equations (4.17) and (4.18). Therefore the output signal from each of these sub-mixers is given by

$$V_{o45}(t) = V_{in}(t)A_s \frac{1}{\sqrt{2}} \cos(\omega_{LO}t). \quad (6.8)$$

The total output signal of the HRM can now be determined by summing the output contribution from each of the three sub-mixers. The output signal of the HRM is given by

$$\begin{aligned} V_{oHRM}(t) &= V_{in}(t)A_s \cos(\omega_{LO}t) \left(\frac{1}{\sqrt{2}} + \sqrt{2} + \frac{1}{\sqrt{2}} \right) \\ &= V_{in}(t)2\sqrt{2}A_s \cos(\omega_{LO}t). \end{aligned} \quad (6.9)$$

Comparing Equations (6.9) and (6.6) it is evident that the gain of the HRM is smaller than the gain of the single comparison mixer by a factor β , which is given by

$$\beta = \frac{2\sqrt{2}}{2 + \sqrt{2}} = 0.83 = -1.64 \text{ dB}. \quad (6.10)$$

Because the noise of the single mixer and the HRM was unchanged, the gain reduction by β , represents a reduction in the SNR. Therefore, when designing a transmitter with the HRM, this SNR reduction needs to be accounted for.

Almost all of the analysis and design to this point has been done assuming the goal is to reject only the third and fifth harmonics, which requires eight phases of the LO signal. If this goal changes and the rejection of higher harmonics is desired, more LO phases are needed and therefore more sub-mixers would be needed. This impacts the SNR because the sub-mixers would contribute to both the noise and the desired signal.

However, the noise contribution would be proportionally larger and the value of β would get lower thereby further reducing the SNR. As the number of sub-mixers is increased, the resulting waveform looks more and more sinusoidal and in the limit the waveform is sinusoidal. It was shown previously that having mixers in parallel did not affect the total noise but it does affect the amplitude of the desired output signal. In the limit the amplitude of the fundamental is equal to the output amplitude because the output is sinusoidal. In the case of a single mixer with a square wave output of unity amplitude, the amplitude of the fundamental is $4/\pi$. Therefore, in the limit, the value of β approaches $\pi/4$, which is equal to 0.79 or -2.1 dB. Fortunately the SNR reduction from the HRM is not severe even as the number of sub-mixers is increased but it still should be accounted for in the transmitter design.

6.3.3 I and Q HRMs

The HRT requires that I and Q HRMs to perform the up-conversion to IF. It was previously shown how a single HRM uses three LO phases to reject the third and fifth harmonics but the discussion did not include how the multiple LO signals are combined for in-phase and quadrature-phase HRMs. Four of the eight phases generated by the divide-by-four circuit are illustrated in Figure 6.1. From these four differential LO signals it is possible to implement I and Q HRMs. An in-phase HRM will use the following combination of signals while

$$I_{HRM} = \sqrt{2}D_1 + D_2 - D_4 \quad (6.11)$$

and the quadrature phase HRM will use

$$Q_{HRM} = D_2 + \sqrt{2}D_3 + D_4. \quad (6.12)$$

Notice that signals D_2 and D_4 are used in both the I and Q HRMs and therefore the loading presented to the divide-by-four circuit is not equal.

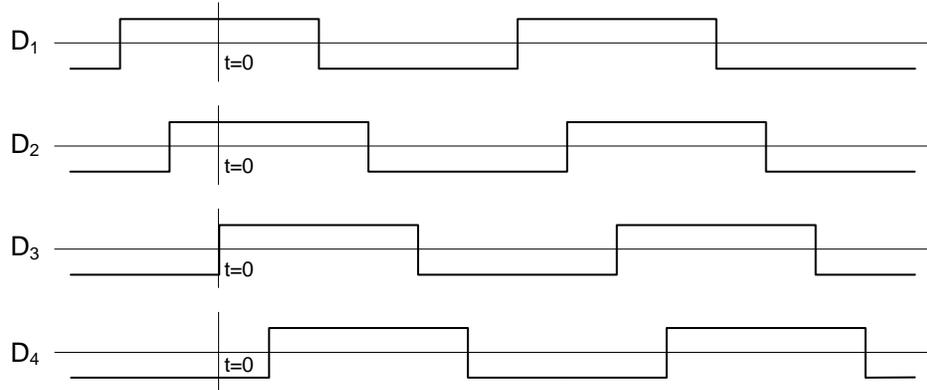


Figure 6.9 Output of divide-by-four showing multiple phases of the LO signal.

The capacitive loading presented to the flip-flops in the divide-by-four circuit is not equal. Assume that the gate capacitance presented by a sub-mixer with current I_o is given by C_o . Since the gate capacitance is generally proportional to the device size, the total gate capacitance of a sub-mixer with current $\sqrt{2}I_o$ is $\sqrt{2}C_o$. The HRT uses two I HRMs and two Q HRMs. Therefore, the total capacitive loading on signals D_2 and D_4 is $4C_o$ and the total capacitive loading on signals D_1 and D_3 is $2\sqrt{2}C_o$. This uneven capacitive loading on the flip-flops can lead to phase errors, which then degrades the harmonic rejection. To correct for the capacitive mismatch, dummy sub-mixers were added to the layout of both the I and Q HRMs. The switches of the dummy sub-mixers in the I and Q HRMs were connected D_3 to D_1 respectively. They were sized such that the total gate capacitance in the switches of one dummy sub-mixer was $(1 - 1/\sqrt{2})C_o$. In this

way the capacitive loading on each flip-flop in the divide-by-four circuit would be equal. To minimize excess power dissipation, the bias current in the dummy sub-mixers was lowered. The lower current in the dummy sub-mixers will change the transient behavior of these switches and because the value of $(V_{GS}-V_t)$ in these transistors will be larger than the switching transistors in the other sub-mixers. Therefore, the gate capacitance during switching will not be equal and the matching will not be exact. Despite this relatively minor drawback the use of dummy mixers is still an effective method of matching the capacitive load.

The layout of a HRM is shown in Figure 6.10. The layout shows four sub-mixers. The dummy sub-mixer is on the far left and the sub-mixer second from the left is scaled by a factor of $\sqrt{2}$. The two sub-mixers on the right complete the HRM.

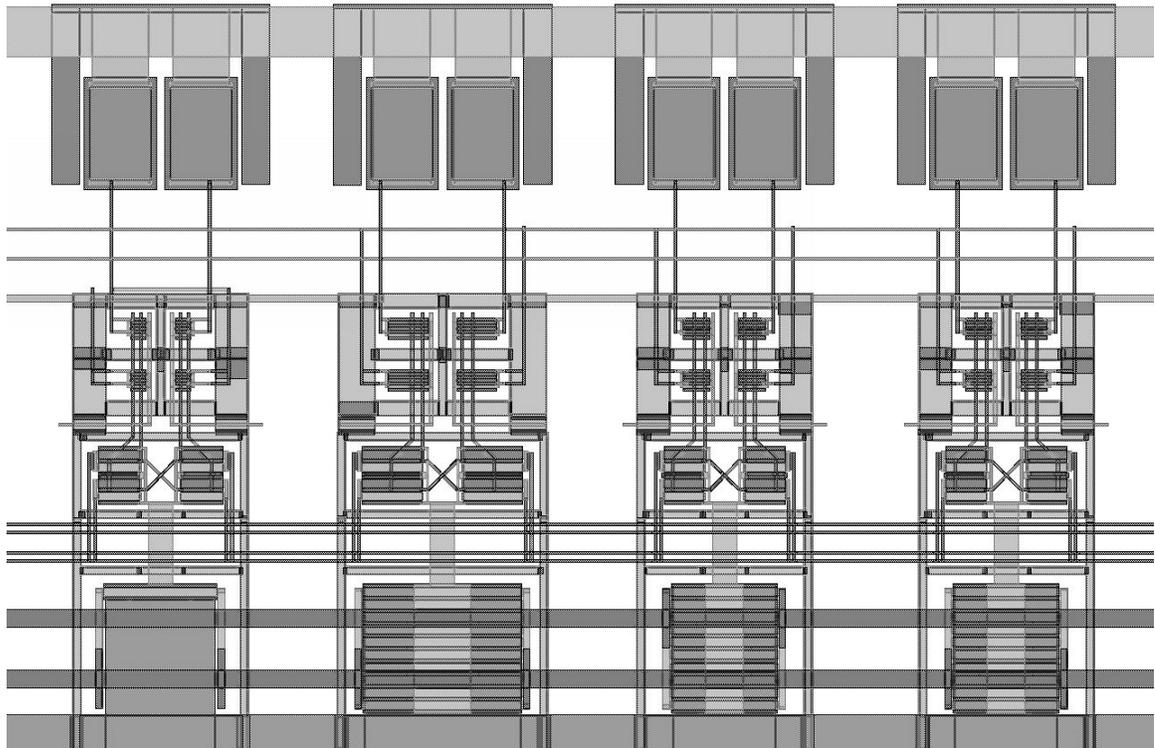


Figure 6.10 HRM layout.

6.3.4 RF Mixers

The mixers used to perform the up-conversion from IF to RF were inductively loaded double-balanced current commutating mixers. A circuit diagram is illustrated in Figure 6.11 showing the device sizes and the load inductor value. The inductor value was chosen such that the resonant frequency of the inductor with the parasitic capacitance at the output node would equal the output frequency. A resistive load was considered but it was determined that the impedance presented by the tuned load was higher than a resistive load at the current levels that were used in the mixer, despite the relatively low

Q inductors that are available in standard CMOS process. The tuned load also improves performance because it acts as a band-pass filter thus attenuating unwanted out-of-band signals and noise performance is improved because ideal inductors do not contribute any noise.

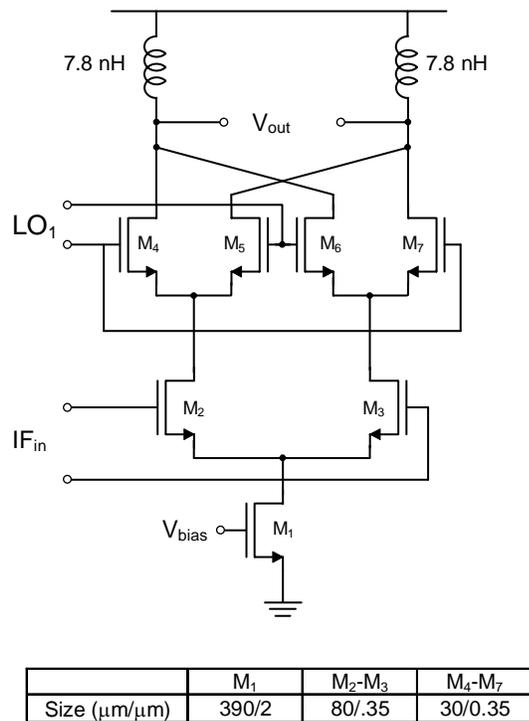


Figure 6.11 Inductively loaded Gilbert cell mixer used for up-conversion to RF.

Another advantage of the inductive load is related to the signal swing at the output. A resistive load biases the output at $V_{DD} - I_{bias}R_L$. Therefore the signal swing at the output becomes limited by the voltage drop in the resistor from the bias current. In contrast, the inductive load, biases the output signal at V_{DD} therefore allowing for a larger signal at the output. This is important in a transmitter because a large output swing will ease the amplification requirements after the mixer, which can reduce the overall power consumption.

The (W/L) ratio of the input devices of the RF mixers is considerably larger than that of the input devices in the sub-mixers of the HRM. This was done because the linearity of these devices is not critical because the IF signal is constant envelope. Conversion gain and noise performance are the most important aspects of this mixer. The size of the switching transistors is determined by a trade-off between conversion gain and capacitive loading, and minimum channel lengths were used to minimize capacitive loading.

The layout of the RF mixers is shown in Figure 6.12. This layout shows both the I and the Q RF mixers and the common inductive load. The inductors are realized with spiral inductors.

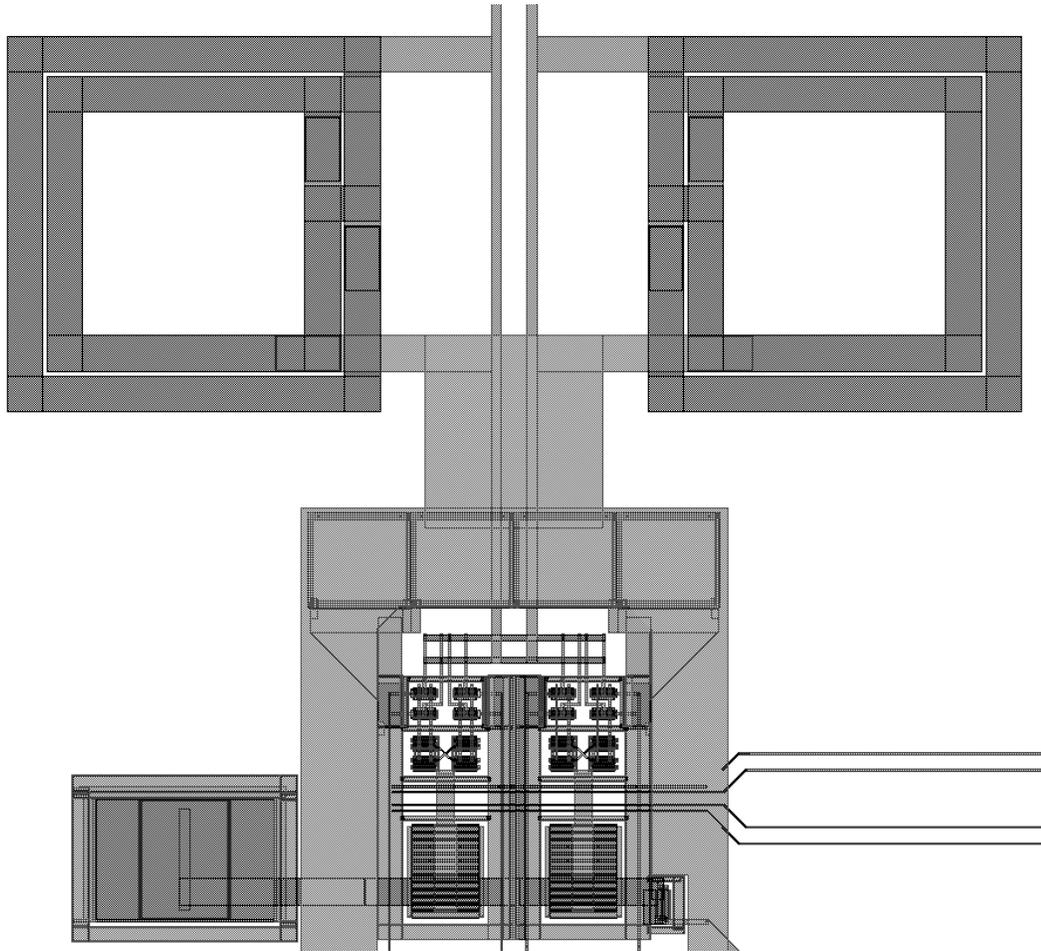


Figure 6.12 Layout of RF mixers.

6.4 Additional Circuits

A number of other circuits were included in the prototype transmitter. A brief description of some of these circuits will follow.

6.4.1 Digital to Analog Converter

The digital-to-analog converters which were used in the HRT prototype were previously described in [35] and the results will be summarized here. The DACs which were used in the baseband section of the HRT were implemented as 10-bit dual resistor string DACs with a maximum sampling frequency of 20 MHz. Although the channel bandwidth of a DCS1800 signal is only 200 kHz a higher sampling frequency was chosen to enable multi-standard operation with potentially higher bandwidth signals. Furthermore, the higher sampling rate eases the filtering requirements after the DAC. The conversion to digital is performed with a coarse 5-bit resistor string and a fine 5-bit resistor string. A simplified circuit diagram of the DAC is illustrated in Figure 6.13.

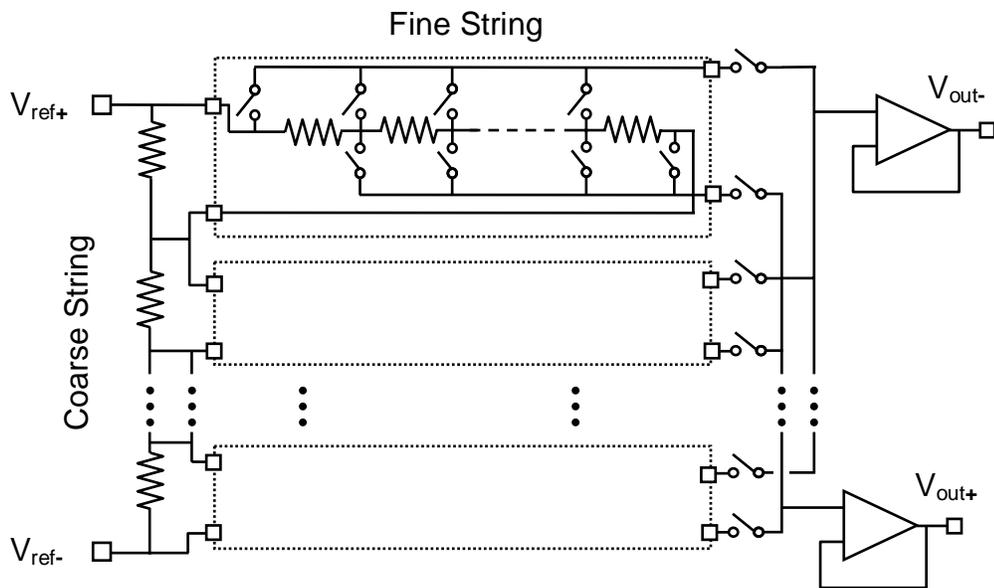


Figure 6.13 Dual resistor string DAC.

One of the design issues with resistor string DACs is the trade-off between power consumption and speed. To decrease the power larger resistors are used. However, in doing so, the settling time is increased and therefore the speed is reduced. To counteract this effect, complementary switches were added to the resistor string as shown in Figure 6.14. Furthermore, a second load capacitor is added to the complementary output V_{OUTC} . Without the complementary switches, the voltage at node A, termed V_A , is plotted as a function of time when a switching event occurs. The complementary switches create an equal and opposite voltage, V_{AC} , at node A which keeps the voltage at node A constant. In doing so the overall setting time is reduced as can be seen in the second plot which shows the output as a function of time with and without the complementary switches. In the actual implementation the complementary output was used as a differential output and therefore the load capacitance on the complementary output node was provided by the output buffer. Therefore, the complementary switches served two purposes by reducing the settling time and therefore the power, and creating a differential output.

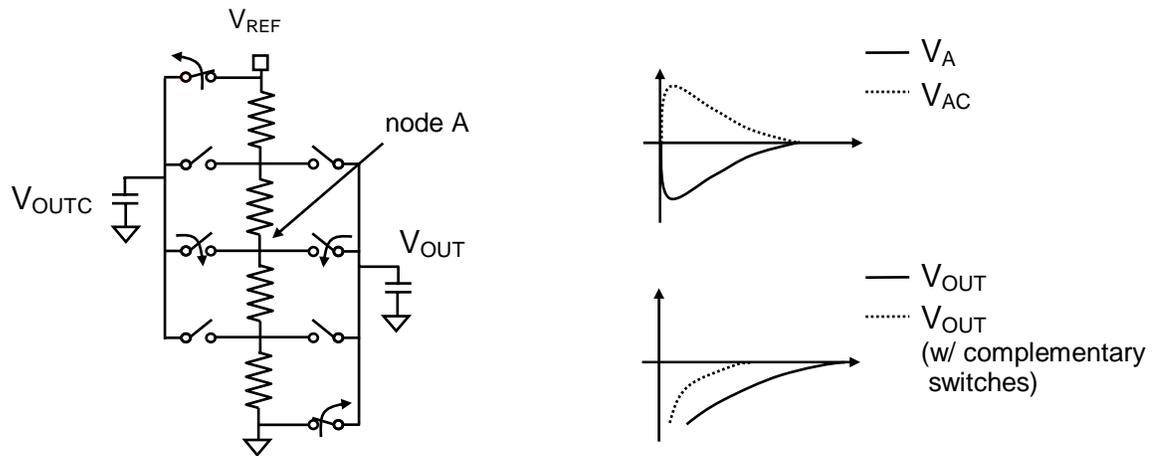


Figure 6.14 Complementary switches in dual resistor string DAC.

6.4.2 Baseband Filter

The main purpose of the baseband filter is to attenuate the aliased signals which are caused by the sampling in the DAC. It was shown in section 2.5.6 that a Butterworth filter represents a reasonable trade-off between passband distortion and stop-band attenuation. Clearly a lower -3 dB frequency is desirable to maximize the stop-band attenuation and a high -3 dB frequency is best for minimal distortion of the in-band signal. For the case of a baseband DCS1800 signal, the -3 dB frequency of the filter is determined by the amount of phase error that is tolerable. Illustrated in Figure 6.15 is a plot of RMS phase error of a modulated GMSK signal as a function of the -3 dB frequency for four Butterworth filters of order 2, 3, 4 and 5. To keep the phase error caused by the filter to less than 0.5 degrees, it is clear from this plot that the -3 dB bandwidth of the filter must be greater than 200 kHz, regardless of the order of the filter. As expected, higher order filters cause more phase error for equal corner frequencies. However, higher order filters also provide more stop-band rejection but they also generally require higher power consumption. All of these factors should be considered when choosing the appropriate filter order and bandwidth.

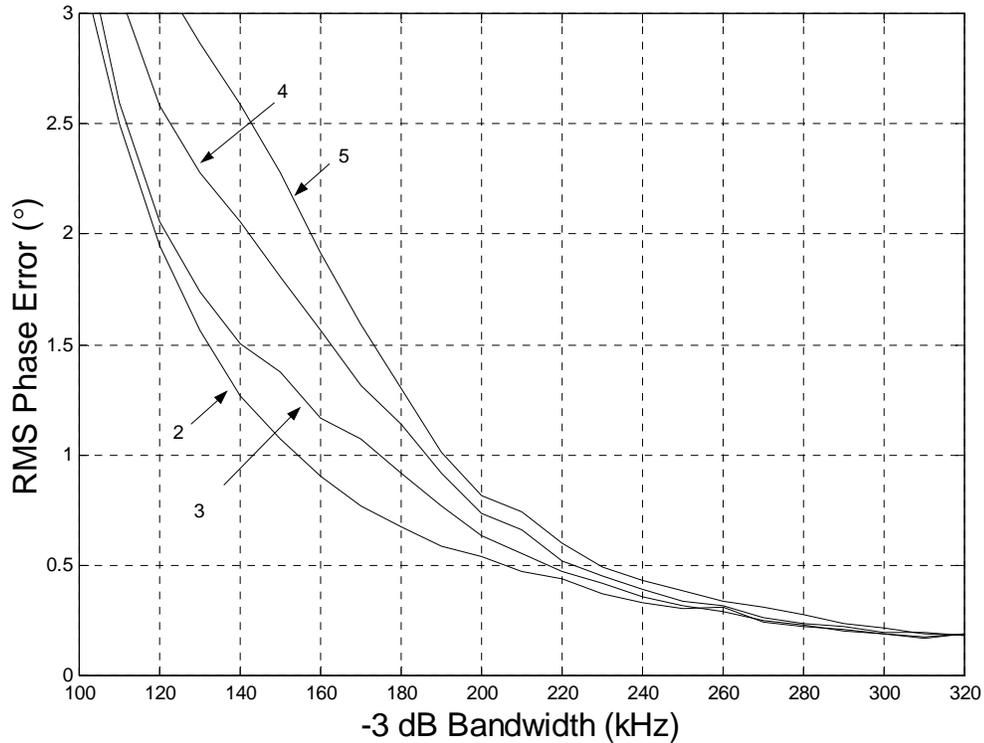


Figure 6.15 Plot of simulated RMS phase error of a GMSK modulated signal as a function of -3 dB frequency for four Butterworth filters of order 2, 3, 4 and 5.

The filtering is typically some type of continuous time filter which uses passive components such as resistors and capacitors. However, in a standard CMOS process, the tolerance of these components is often quite large and therefore the corner frequency of the filters must be increased to accommodate the variation. For example, if a third order Butterworth filter were chosen, from Figure it can be determined that the -3 dB bandwidth should be greater than 220 kHz. Assuming the -3 dB frequency of the filter is determined by an RC product and the tolerance of each component is 30 percent, then in the implementation, the filter would need to be designed with a corner frequency of 450 kHz. The higher corner frequency in turn potentially reduces the stop-band attenuation.

The RF frequency synthesizer was implemented as a wide loop bandwidth PLL. The loop bandwidth was 8 MHz and the PLL generated a fixed frequency of 1.3824 GHz. The implementation was fully differential including the control voltages to the VCO.

The IF frequency synthesizer was implemented by a PLL with a loop bandwidth of 40 kHz. The frequency synthesizer generated an output from 1.3104 GHz to 1.4704 GHz in steps of 800 kHz. This output was then lowered in frequency by the divide-by-four circuit which ultimately generated an output that ranged from 327.6 MHz to 367.6 MHz in steps of 200 kHz. By dividing the output of the frequency synthesizer the phase noise by a factor of four the phase noise was reduced by a factor of 16 [35].

6.4.4 Eight Phase Generation

Section 4.8 presented two methods to generate the multiple LO phases that are needed in the HRM: a divide method and passive RC method. In the implementation of the HRT, the divider method was used and the divide-by-four circuit was originally presented in [17].

The divide-by-four circuit was implemented with two master-slave D flip-flops similar to the configuration in Figure 4.21. Each flip-flop was implemented with source-coupled logic and the outputs from each flip-flop were buffered to effectively drive the switching transistors of the HRM.

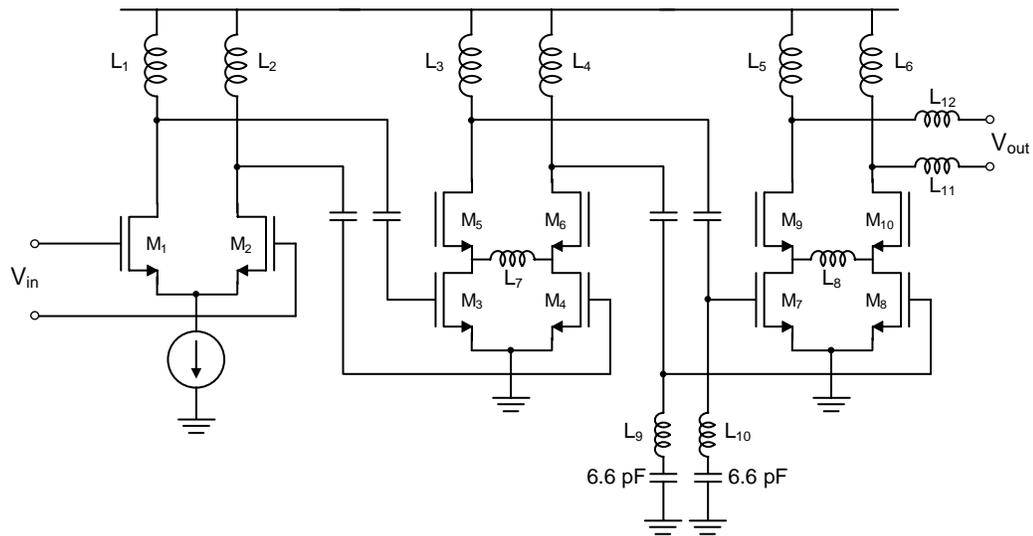
6.4.5 Power Amplifier

An excellent and detailed description of the power amplifier used in the HRT can be found in [36]. A brief overview of the implementation of the PA is provided in this section. Some of the basic design ideas are presented in this section because the PA was an important part of the complete transmitter.

The HRT was designed to operate with constant envelope modulation. This fact, combined with the use of HRMs allowed the use of a highly non-linear PA. Among non-linear PAs, switching PAs such as class D, class E or class F PAs can achieve reasonably high efficiency but controlling the output power is difficult. The output power of these types of PAs is generally controlled by regulating the supply voltage. Therefore, a DC-DC converter is typically necessary to effectively control the output power. However, integrating a high-performance DC-DC converter in a standard CMOS process is difficult and therefore a different class of PAs was explored.

The PA which was included in the HRT was biased for class C operation, which is similar to of class A and class B amplifiers except the conduction angle is kept below 180 degrees. This is done by keeping the bias voltage below the threshold voltage of the input device. As a result, the circuit is potentially more efficient because it is not on for the entire period for entire period. However, in doing so, the nonlinearity of the PA is increased. Furthermore, with the lower conduction angle it is difficult to achieve high output power because the device is not on for the entire period. Therefore, in choosing the PA type, tradeoffs between efficiency, maximum output power, and linearity must be evaluated.

The PA was designed for 1.5 watts of peak power and achieving this power requires a combination of voltage swing and current draw. The peak voltage on chip is limited by the relatively low oxide breakdown voltage of V_{ox} in a standard CMOS process and as a result, large currents are necessary to achieve high power. Due to the low transconductance of CMOS devices, these large currents require large device sizes. A schematic diagram of the fully differential class-C PA is illustrated in Figure 6.17 showing the large device sizes. The PA was composed of three stages. The first stage was biased for class A operation while the second and third stages were biased for class C operation. The series LC prior to the third stage was added to allow for the use of a larger load inductor on the second stage. Due to the large gate capacitance of the third stage, the load inductor of the second stage would have been very small and therefore difficult to implement without the series LC.



	M ₁ -M ₂	M ₃ -M ₄	M ₅ -M ₆	M ₇ -M ₈	M ₉ -M ₁₀
Size (μm/μm)	300/.35	2500/.35	2400/.35	18300/.35	9240/0.35

	L ₁ -L ₂	L ₃ -L ₄	L ₅ -L ₆	L ₇	L ₈	L ₉ -L ₁₀	L ₁₁ -L ₁₂
Value (nH)	1.03	0.42	0.55	2.6	0.54	1.49	1.59

Figure 6.17 Schematic of class-C PA.

6.5 On-chip Signal Coupling

Given the high level of integration, it was anticipated that coupling between circuits could be a potentially very serious problem which would either degrade the performance or possibly prevent operation. This problem seemed particularly important given that the chip included a PA that was designed to output 32 dBm and the digital circuits in the DAC. The steps that were taken to reduce the effects of signal coupling fall into two broad categories: steps to prevent coupling from occurring in the first place and steps to mitigate the effects of unwanted signals that have coupled into a circuit.

To prevent unwanted signal coupling, two steps were taken. First, each circuit block had a unique supply and ground on chip. This was done in an attempt to isolate the supply and ground for each circuit block and to therefore prevent signal coupling from occurring. In DACs, which contain both analog and digital signals, digital and analog supplies were also separated.

Second, in addition to the independent ground connections, multiple substrate contacts were placed around the transistors. The multiple substrate contacts provide a potentially lower impedance path to ground for the current that is injected into the substrate. Therefore, the current is more likely to flow to local ground and then off-chip via the bond wire connections to off-chip ground than through the substrate and into another circuit. In addition to the substrate contacts close to the individual transistors, a guard ring of substrate contacts surrounded the PA.

Although the aforementioned techniques attempt to prevent signal coupling, some level of coupling is inevitable. To minimize the impact of this coupling a number of steps were taken. First, the entire signal path was fully differential including the control voltage on the VCOs in both frequency synthesizers. This does not prevent signal coupling but it minimizes the effects because the coupled signals will most likely appear as a common mode signal.

Second, in addition to the separate supplies, each supply was heavily bypassed on chip. This was done so that the supply and ground for a given circuit block would move up and down in unison when unwanted signals coupled to either the local supply or ground. In this way there would be minimal impact on the signal path in the event of unwanted coupling. The previously mentioned substrate contacts might also provide

some benefit along these lines. The substrate contacts would hopefully move the body connection in unison with the local supply and ground and thus further mitigate the effect of unwanted coupled signals.

Finally, large capacitors were added between critical bias lines and ground in an attempt to provide some filtering for signals that have coupled directly to the bias lines.

6.6 Transmitter Test Chip

A prototype transmitter was fabricated in 0.35- μm , double-poly, five-layer metal, CMOS process. A chip micrograph is shown in Figure 6.18. The odd shape of the chip stems from the fact that the prototype transmitter was part of a transceiver and only the transmitter portion of the chip is shown here. The two DACs are located in the upper right corner and directly to the left of them are the I and Q baseband filters. The signal path heads down leading to the harmonic-rejection mixers, the RF mixers and the PA located at the bottom of the chip. To the left of the PA is IF frequency synthesizer and above the RF mixes is the RF frequency synthesizer. The dimensions of the die shown in Figure 6.18 are approximately 9.9 mm on the right edge and 5.9 mm on the bottom edge. The layout was not particularly compact and the total active die area was approximately 24 mm².

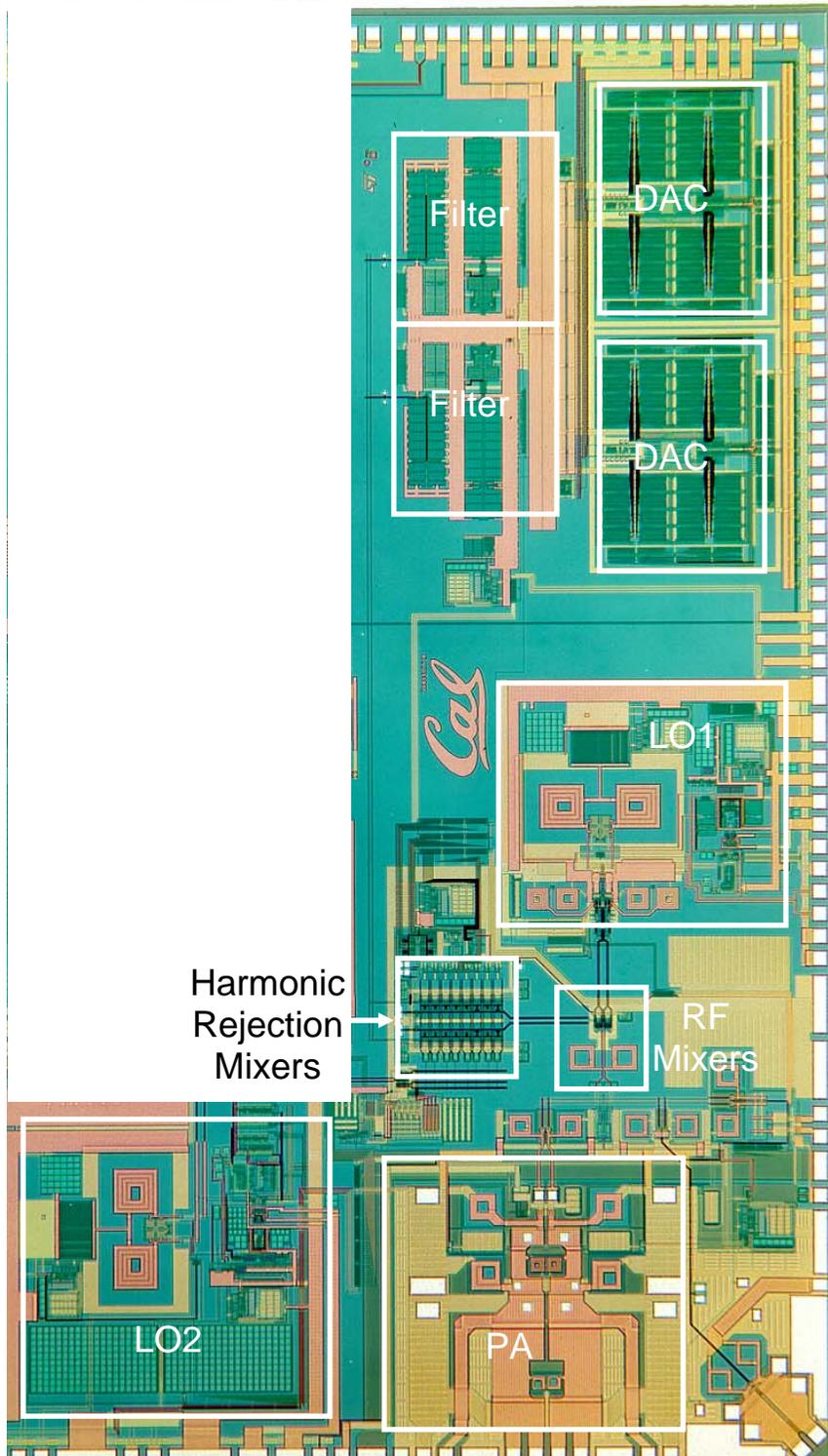


Figure 6.18 Transmitter test chip micrograph.

6.7 Summary

This section discussed the prototype transmitter that was built to evaluate the concepts presented in the earlier chapters. Furthermore, some of the design issues associated with the HRT and the HRMs were discussed such as the effect of the HRM on SNR, the tolerable levels of baseband non-linearity for the DCS1800 standard, and the generation of I and Q HRMs. This section also discussed some of the other circuits that were included in the HRT prototype and some of the requirements for these circuits, such as the effect of filter bandwidth on the RMS phase error of a DCS1800 GMSK modulated signal. The section ended with a discussion of strategies to reduce on-chip signal coupling and finally a micrograph of the fabricated chip was presented. The following chapter will discuss the measured results from this prototype.

Chapter 7

Measurement Results

7.1 Introduction

While the previous chapter focused on some of the design issues associated with harmonic-rejection mixer and the harmonic-rejection transmitter, this chapter present the measurement results from the HRT prototype. The HRMs were only tested in the context of a complete transmitter and were not tested as stand-alone circuit.

The HRT test chip included the entire signal path from the DACs to the PA and two fully integrated frequency synthesizers. The entire signal path was differential including the PA output and the testing buffer output. A diagram of the test setup is shown in Figure 7.1. The input to the baseband I and Q DACs was provided by an HP 16522 pattern generator. This was used to provide both a modulated input and a single

tone input. The differential output from either the testing buffer or the PA was first converted to single ended signal with a balun. The single-ended signal was then analyzed with an HP 8560 series spectrum analyzer to measure the transmitted spectrum. To measure the modulation accuracy, an HP 89441 vector signal analyzer was used.

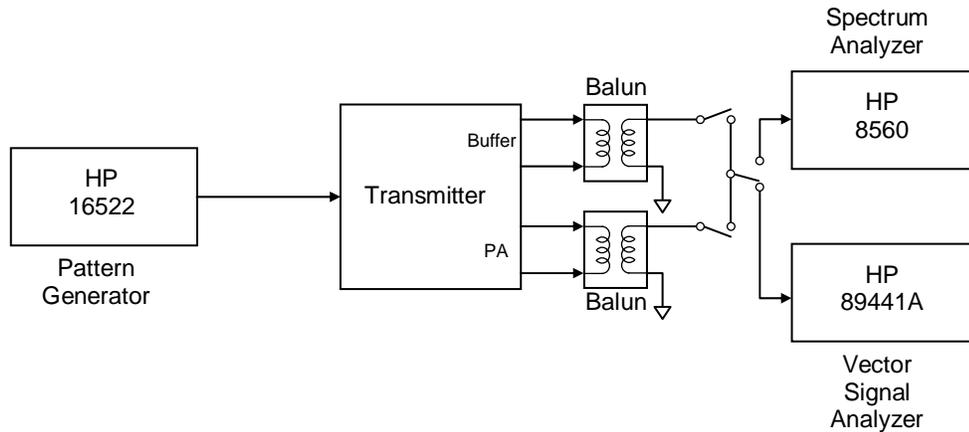


Figure 7.1 Transmitter test setup.

To test the transmitter, a technology referred to as chip-on-board was used. In this technique, the die is attached directly to a test board with conductive epoxy. Bond wires then connect the pads of the chip to landing zones on the board. The advantage of this technique is that the lead inductance associated with standard packaging can be reduced. However, since a die is attached directly to the test board, a separate board is needed for each die. A die can be removed from the board and another die bonded in its place, but this can only be done a limited number of times due to bonding limitations.

A photograph of the test board showing the area close to the attached die is shown in Figure 7.2. The transmitter occupies the lower half of the chip. The transmitter was fabricated as part of an entire transceiver and the remaining portion of the chip is comprised of circuits used by the receiver. The differential PA output is located on the

lower left portion of the chip and a microstrip balun, similar to the balun presented in [37], was used for the differential to single ended conversion.

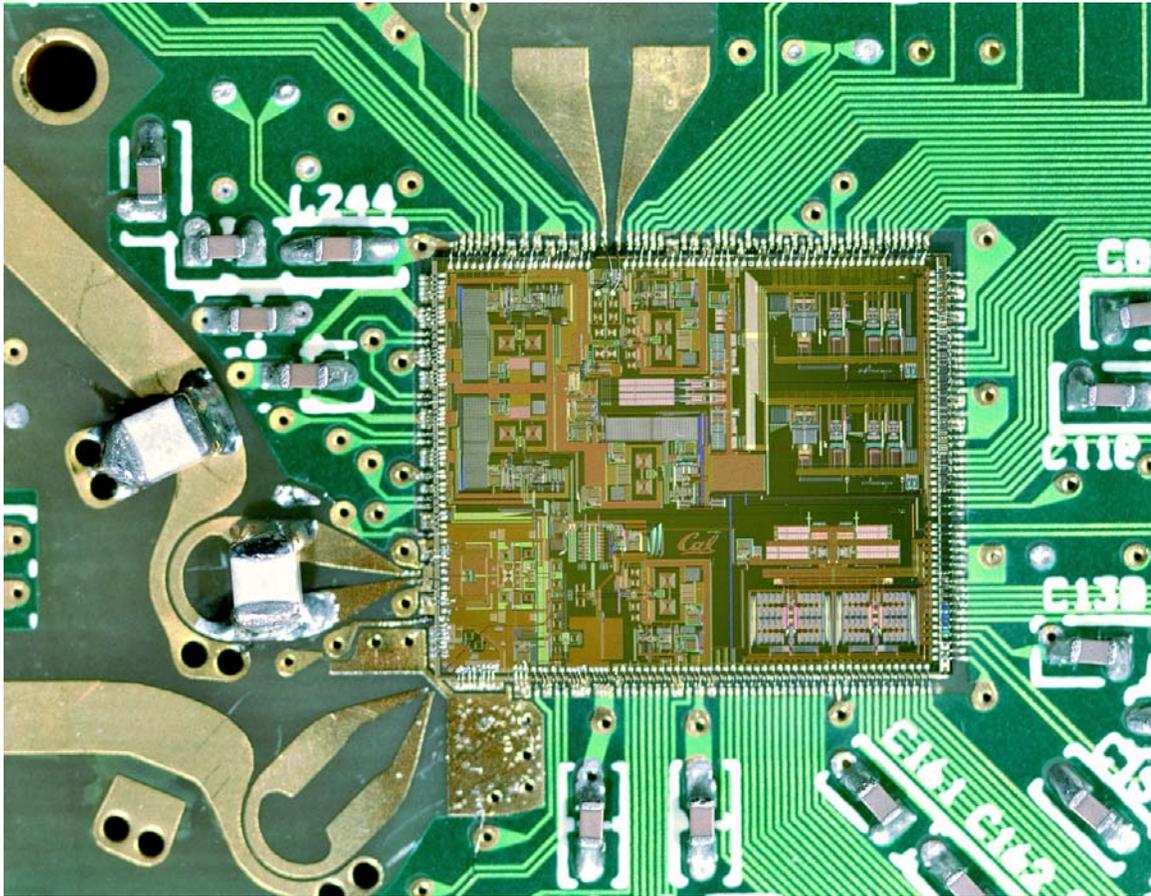


Figure 7.2 Photo of test board with chip-on-board technology.

One major concern regarding the performance of transmitter was the coupling between circuits, particularly with a fully integrated PA and the digital circuits in the DAC. Specifically, the large signals that are injected into the substrate can potentially interfere with sensitive analog signals in the frequency synthesizers and the mixers. This topic was discussed in the previous chapter and a number of steps were taken to prevent on-chip signal coupling and to reduce the effects of coupled signals. The supplies on

chip were heavily bypassed and this also was done on the test board. Furthermore, independent voltage regulators were used for the multiple supplies that were used in each circuit block on chip. Although separate ground connections were used on chip, these were connected to a common ground on the test board.

7.2 Measurement results

One of the primary goals in the design of the transmitter prototype was to demonstrate that a high performance transmitter could be fully integrated in to a standard CMOS process. To evaluate the performance, the transmitter was designed to meet the requirement of the DCS1800 cellular telephony standard. The measured results will be discussed with respect to modulation accuracy and unwanted spectral emissions. Furthermore, the effect of the HRMs will also be demonstrated.

One of the most critical tests for any transmitter is the single-sideband test in which a baseband tone is up-converted instead of a modulated signal. This test allows for the evaluation of quadrature gain and phase mismatch, LO feedthrough and DC offsets, and third order distortion, all of which will affect the modulation accuracy. To perform this test a 50 KHz baseband tone was used and the output from the testing buffer is shown in Figure 7.3. The largest peak in this plot is the upper sideband and the levels of the other peaks are measured relative to this. The lower sideband is a measure of the quadrature phase mismatch and the gain error. The measured value in this case is approximately -56 dB which is a low level given that no tuning was performed. From Figure 6.2 it can be determined this corresponds to approximately 0.1 degrees of RMS

phase error in the worst case. Although this low value is better than what is required by the DCS1800 standard, it is indicative of the high performance capability of the HRT. The result is not unexpected given the improved image-rejection that was predicted in Section 5.3.

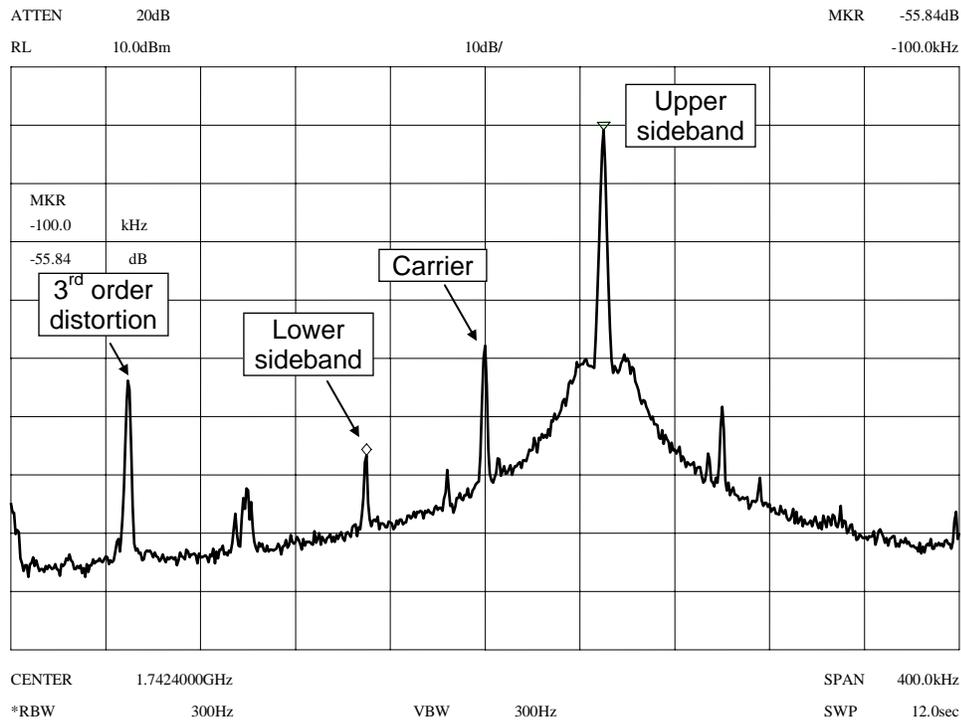


Figure 7.3 Single sideband output at testing buffer output.

The other important peaks in Figure 7.3 are the third order distortion, which is located 150 kHz below the carrier and the carrier feedthrough, which is located at the carrier. The third order distortion was measured at -44 dB which is below the target level of -40 dB. This distortion was dominated by the input to the HRMs and this low level confirms the rejection of the third IF harmonic by the HRMs. If the third IF harmonic were not rejected, the third order distortion from the RF mixers and the PA would have

increased this value above -44 dB. Figure 6.5 shows that this level of third order distortion corresponds to approximately 0.25 degrees of RMS phase error.

The peak at the carrier frequency is generally caused by baseband DC offsets or LO feedthrough in the mixers. The carrier feedthrough in this case was measured at -38 dB and from Figure 6.3 it is evident that this corresponds to approximately 0.5 degrees of phase error.

Although the single sideband test is a good indicator of modulation accuracy, ultimately a modulated output signal must be compared with an ideally modulated signal. The comparison of the two signals is done with a vector signal analyzer and the results of this test are shown in Figure 7.4. The figure shows the measured constellation, eye diagram and plot of the phase error at each symbol. The RMS phase error was 1.3 degrees and the peak phase error was 3.7 degrees. The DCS1800 requirement specifies that the RMS phase error is less than 5 degrees and the peak phase error is less than 20 degrees. Clearly the HRT prototype easily satisfies this requirement indicating very high performance with regard to modulation accuracy. This was evident in the single sideband test and confirmed with this measurement. The total RMS phase error predicted by the single-sideband test is approximately 0.85 degrees indicating that approximately 0.45 degrees of RMS phase error was caused by other sources such as the baseband filter, phase noise, and thermal noise.

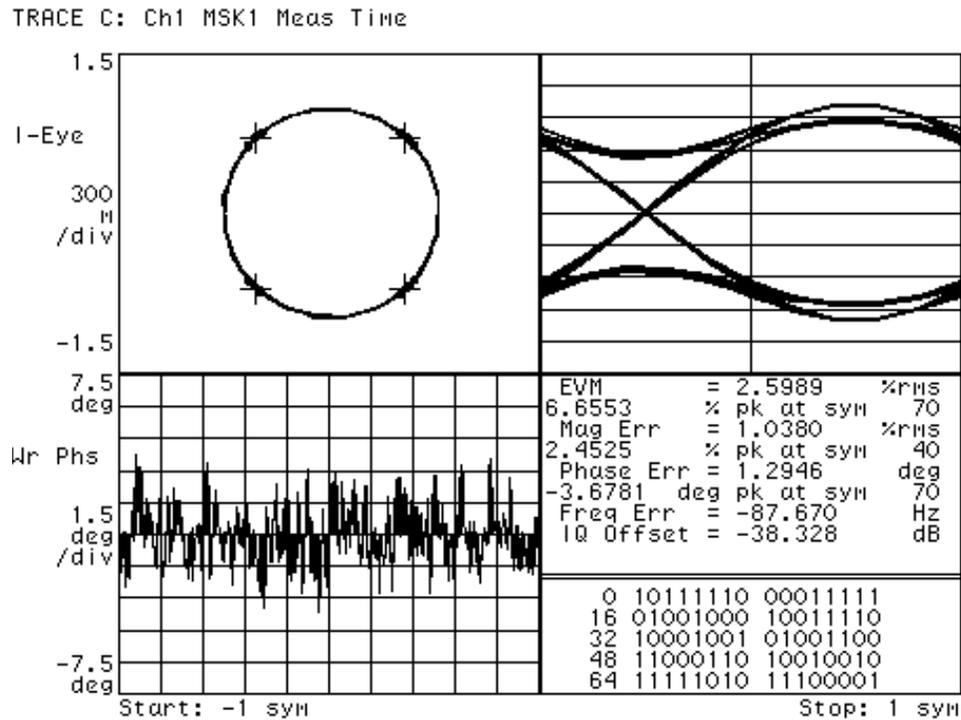


Figure 7.4 Vector signal analyzer output.

The signals that were analyzed in Figure 7.3 and 7.4 were driven off-chip by the testing buffer. To better understand the effects of the non-linear PA on the output signal, a single sideband test was also performed at the PA output. The resulting spectrum, shown in Figure 7.5, almost identical to the spectrum from the testing buffer indicating the highly non-linear PA had very little effect on the modulation accuracy. The third order distortion term increased to -43 dB which is still below the target of -40 dB.

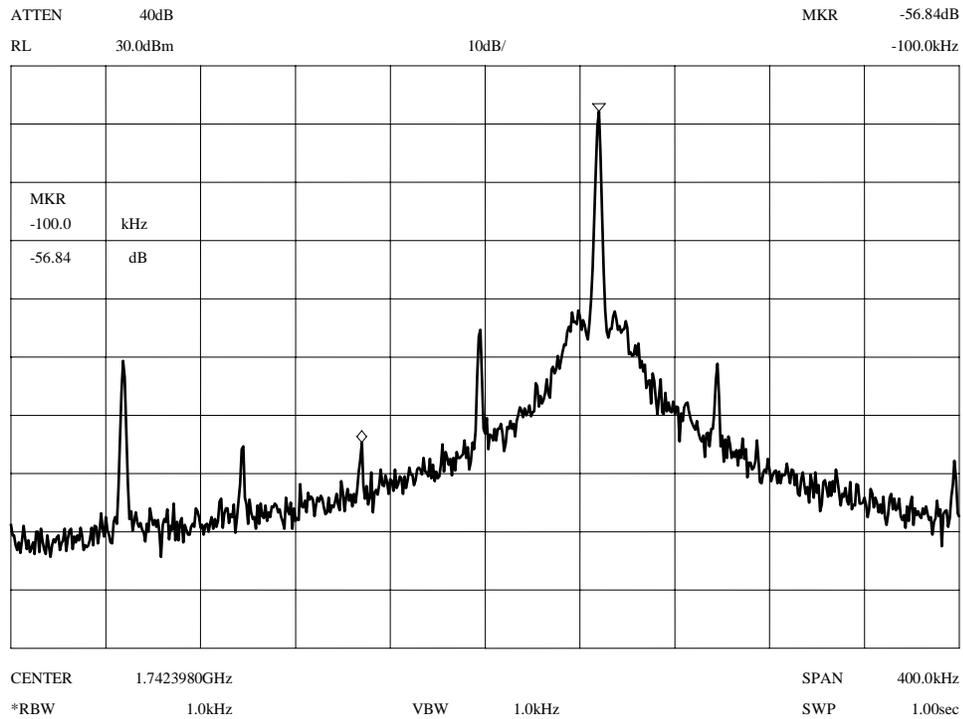


Figure 7.5 Single sideband spectrum at PA output.

Another critical test for the transmitter is the ability to produce a modulated spectrum which adheres to the spectral mask requirements. The modulated output spectrum of a DCS1800 GMSK signal is shown in Figure 7.6 along with the spectral mask requirement. To meet this requirement the modulated signal must fall below the spectral mask at all frequencies, which is clearly accomplished by this transmitter.

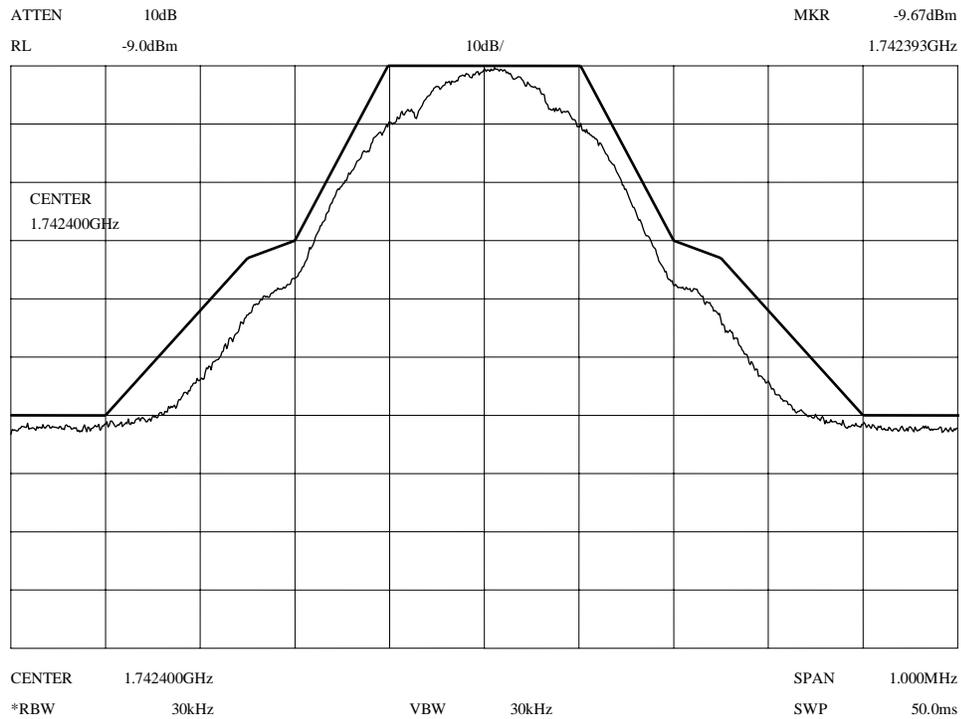


Figure 7.6 Modulated output spectrum with DCS1800 spectral mask.

To this point the performance of the HRMs has been inferred by the low third order distortion. To more accurately examine the performance of the HRMs, the wideband output spectrum is shown in Figure 7.7. The rejection of the third and fifth harmonic was measured at approximately -68 dB and -69 dB respectively. Although the rejection was aided by the bandpass nature of tuned loads, the relatively low-Q of the inductors limited this benefit. This plot confirms the functionality of the HRMs and it implies accurate gain and phase matching in the HRM. The other spurs seen in Figure 7.7 are either caused by spurs in the frequency synthesizers or higher order mixing terms.

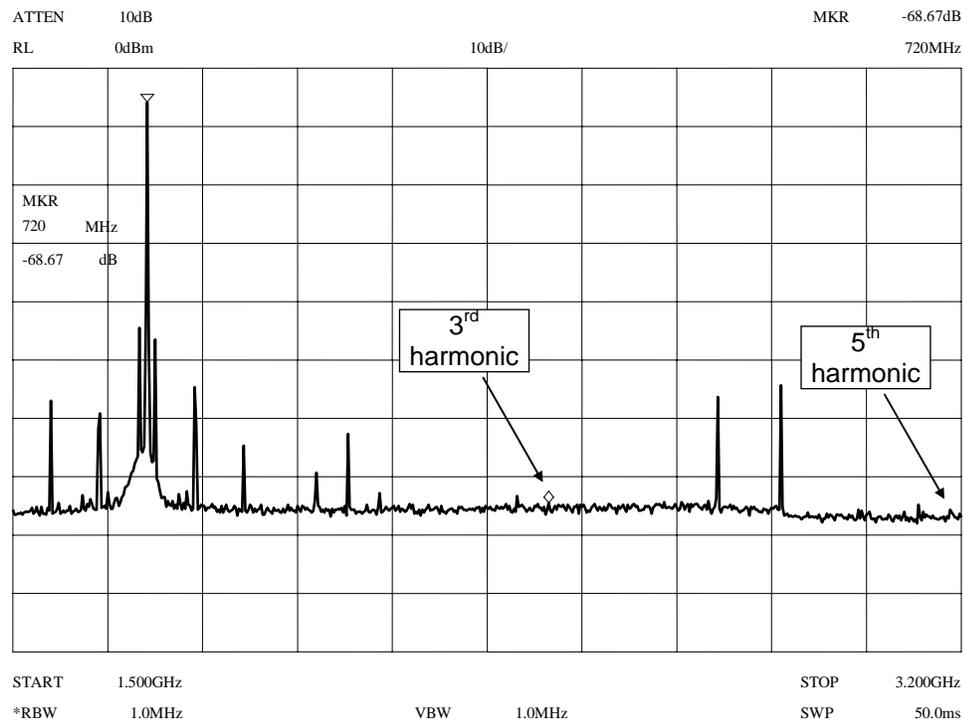


Figure 7.7 Wideband plot showing the rejection of 3rd and 5th IF harmonics.

The modulated spectrum shown in Figure 7.6 showed the spectral mask requirements close to the carrier. One of the most difficult requirement of the DCS1800 standard is the spectral mask requirement at 20 MHz offset from the carrier. The standard requires that the noise at this offset frequency is less than -151 dBc/Hz. This low value is required because 20 MHz separates the transmit and receive bands and therefore, a transmitter on the highest channel frequency could be interfering with a receiver that is 20 MHz away operating at the lowest channel frequency. Figure 7.8 shows a plot of the noise at the output of the testing buffer and the noise at 20 MHz offset and the measured value was -126 dBc/Hz. This noise was most likely dominated by the phase noise from the frequency synthesizers. When the output was taken from the PA this

level increased to -121 dBc/Hz. This implies that a filter after the PA with 30 dB of rejection would be needed to meet the requirement, which is possible with a single filter after the PA. Therefore, a filter is not needed before the PA and single chip implementation could meet the DCS1800 specification for noise in the receive band.

The large spurs in Figure 7.8 were generated in the frequency synthesizers.

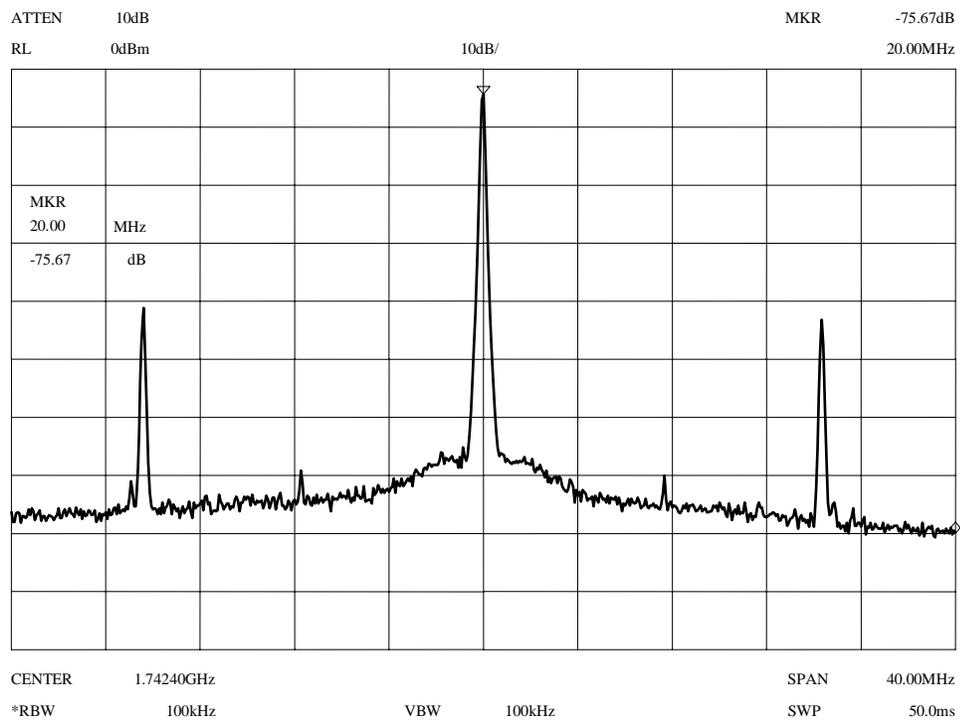


Figure 7.8 Plot showing wideband noise at 20 MHz offset.

A performance summary is given in Table 7.1. The IF image suppression is a measure of the IF image that is located at $f_{RF}-2f_{IF}$, which was caused by the up-conversion from IF to RF. The image suppression is a measure of the effectiveness of the image-reject mixer formed by the two RF mixers. The chip operated with a 3 volt power supply

and the maximum output power from the PA and the testing buffer was 25 dBm and 0 dBm, respectively.

	Testing Buffer	PA	DCS1800 Spec.
LO2 3rd Harmonic Spur	-68 dBc	-68 dBc	-
LO2 5th Harmonic Spur	-69 dBc	-69 dBc	-
Single Sideband Suppression	-56 dBc	-57 dBc	-
3rd Order Distortion	-44 dBc	-43 dBc	-
RF Carrier Feedthrough	-38 dBc	-38 dBc	-
RMS Phase Error	1.3°	N/A	5°
Peak Phase Error	3.7°	N/A	20°
20 MHz Carrier Offset Noise	-126 dBc/Hz	-121 dBc/Hz	-151 dBc/Hz
IF Image Suppression	-60 dBc	-60 dBc	-
Output Power	0 dBm	25 dBm	30 dBm
Supply Voltage	3 V	3 V	-
Active Chip Area	20 mm ²	24 mm ²	-
Technology	0.35 mm DP- 5M CMOS	0.35 mm DP-5M CMOS	-

Table 7.1 HRT performance summary.

The prototype transmitter consumed 151.3 mA from a 3 volt supply, excluding the PA and the testing buffer. breakdown of the current consumption for the HRT is shown in Figure 7.9. The Q of the inductors was believed to be substantially lower than what was predicted in simulation. As a result, to achieve the desired signal levels, the current in many of the circuit blocks was significantly increased. The lower than expected Qs also reduced the gain of PA and therefore the output power and efficiency of the PA were also lower than expected.

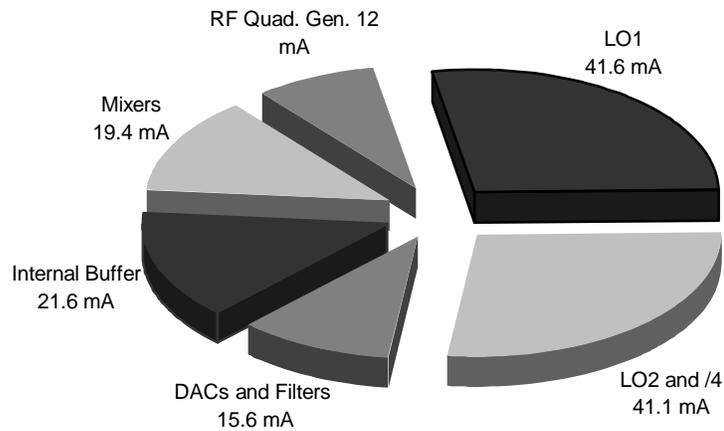


Figure 7.9 Breakdown of current consumption in HRT.

7.3 Summary

This chapter discussed the measurement results of the prototype HRT transmitter. The chip was designed and tested to meet the DCS1800 specification and the measured results indicate that the vast majority of the specifications were met. The modulation accuracy, which is defined as phase error for DCS1800, was easily within the specification and the modulated spectrum fell beneath the spectral mask requirements. The wideband noise was not below the specification but this could potentially be accomplished with a single filter after the PA. The results were not significantly different when the output was generated with an internal testing buffer or the integrated PA.

Chapter 8

Conclusions

8.1 Research Summary

The advancement of technology combined with the desire for portable communications has driven demand for wireless services and with this demand has come a proliferation of wireless standards. Central to these wireless services is the ability of the technology to provide high performance, low cost and small form factor solutions which potentially operate with multiple RF standards. Concurrently, CMOS technology has continued to improve raising the potential the ability to operate in the RF domain.

The primary contributions of this thesis were advancements at both the architectural and circuit level that allowed for high performance operation on a single CMOS chip while maintaining the potential for operation with multiple RF standards.

The first primary research contribution relates to the unwanted harmonics that are created by typical switching mixers. These harmonics can degrade the performance and often require the use of discrete filters. In this work a mixer was described that rejects the harmonics and thus eases the filtering requirements. This mixer, termed the harmonic-rejection mixer, was designed to eliminate the two closest harmonics, which are the third and fifth. The key aspects of the design of the HRM were included which included the rejection level as a function of mismatch and the scaling issues of the HRM. Furthermore, the HRM was analyzed with respect to gain, noise and distortion. In addition, it was shown that the rejection of the harmonics is not dependent on a specific LO waveform.

While the HRM was at the circuit level, the second primary research contribution was an architectural level advancement. By leveraging the abilities of the HRM to ease filtering requirements, a transmitter architecture was designed that simultaneously allowed for single chip integration and high performance. Furthermore, the architecture, termed the harmonic-rejection transmitter, was designed to facilitate operation with multiple RF standards. The HRT architecture was analyzed with respect to image-rejection and it was shown to have significantly improved image rejection performance as compared with a typical quadrature modulator. To even further improve the performance, a low frequency tuning scheme was proposed.

The third primary research contribution was the implementation of a prototype transmitter in a standard CMOS process. To assess the performance, the transmitter was designed to operate with the DCS1800 cellular telephony standard. The transmitter achieved satisfied most of the requirements for DCS1800. The prototype included the

DACs, baseband filters, HRMs, RF mixers, frequency synthesizers and a PA. In conclusion a high performance transmitter was demonstrated that was implemented in a single CMOS IC and facilitated operation with multiple RF standards.

8.2 Future Work

The transmitter that was designed in this work was tested with a constant envelope modulation scheme. Due to the inherent flexibility of quadrature modulators the architecture facilitated multi-standard operation. However, for more effective multi-standard operation, the linearity of the transmit path must be high enough to accommodate non-constant envelope modulation and this includes the power amplifier. Therefore, future integrated transmitters should demonstrate multi-standard operation with both types of modulation. However, linear PAs generally suffer from poor efficiency, particularly when implemented in a CMOS technology. One solution to this problem is to use linearization techniques so that non-linear PAs can be used with non-constant envelope modulation schemes. However, many of the architectures built around transmitter linearization have difficulties with wide channel bandwidths and are therefore limited to lower bandwidth RF standards. Therefore, future work into transmitter architectures in which enable multi-standard operation while using non-linear PAs is needed.

One of the key building blocks in most transmitters is the quadrature modulator. Although this circuit block is very useful for modulating and up-converting a baseband signal, it generally limits the wideband noise performance of a transmitter. Phase noise

from frequency synthesizers plays an important role as does the relatively high noise figure of switching mixers. Future research is needed to reduce the wideband noise of quadrature modulators so that mixer based transmitters may operate without the need for filtering after the PA. Currently, many PLL based transmitters do not need a filter after the PA because of the bandpass transfer function from the quadrature modulator to the output. However PLL based transmitters are limited with respect to multi-standard operation. Therefore, research into low-noise quadrature modulators might allow for the use of multi-standard, mixer based transmitters that require no filtering after the PA.

With regards to single chip integration, on-chip coupling issues pose a significant problem to the overall performance and this is only exacerbated by integrated PAs. Currently, the ability to predict the interactions between circuits is limited. Future research into better on-chip modeling would be beneficial at both the architectural and circuit level.

Appendix A

Image Rejection Derivation of HRT

The image rejection of the HRT, given Equation (5.3) in Section 5.3 was stated without a derivation. This section will derive this equation based on the model of phase error and gain mismatch in the HRT given in Figure A.1. This figure is a copy of Figure 5.3 and is repeated here for easier reference. The gain mismatch is modeled by α and the LO_1 and LO_2 phase errors are modeled by θ_1 and θ_2 , respectively.

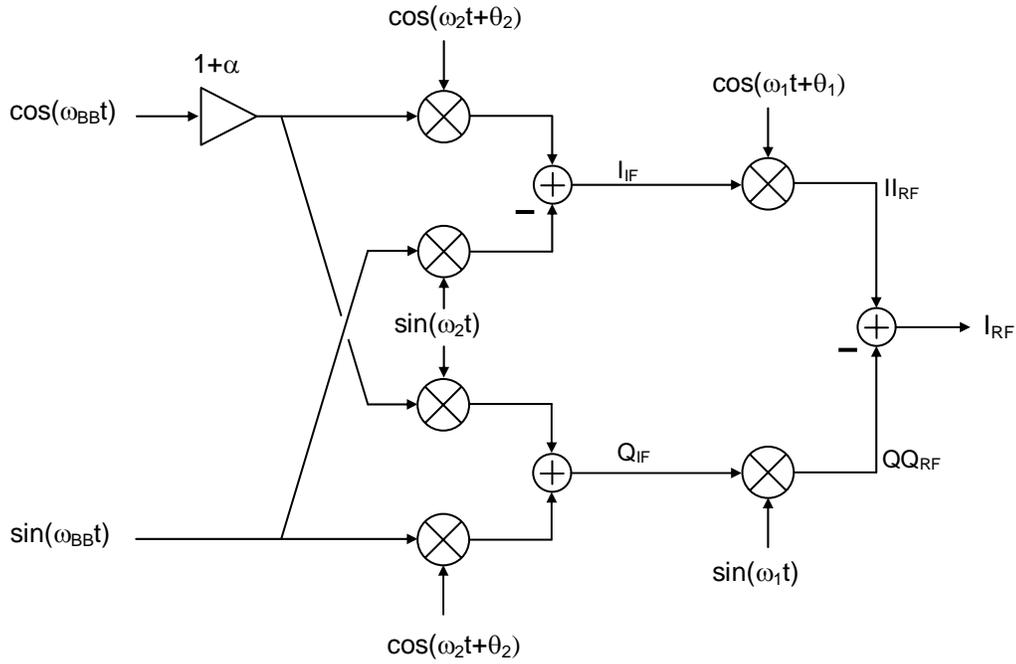


Figure A.1 Model used to evaluate image rejection in the HRT.

To begin the derivation, the signal I_{IF} is given by

$$I_{IF} = (1 + \alpha) \cos(\omega_{BB} t) \cos(\omega_2 t + \theta_2) - \sin(\omega_{BB} t) \sin(\omega_2 t) \quad (\text{A.1})$$

and after some manipulation I_{IF} can be re-written in the following form:

$$\begin{aligned} I_{IF} = & \frac{1}{2} \left((1 + \alpha) \cos(\theta_2) + 1 \right) \cos((\omega_2 + \omega_{BB})t) \\ & + \frac{1}{2} \left((1 + \alpha) \cos(\theta_2) - 1 \right) \cos((\omega_2 - \omega_{BB})t) \\ & - \frac{1}{2} \left((1 + \alpha) \sin(\theta_2) \right) \sin((\omega_2 + \omega_{BB})t) \\ & - \frac{1}{2} \left((1 + \alpha) \sin(\theta_2) \right) \sin((\omega_2 - \omega_{BB})t) . \end{aligned} \quad (\text{A.2})$$

To simplify the previous equation, three additional terms are defined as follows:

$$K_1 = \frac{1}{2}((1 + \alpha) \cos(\theta_2) + 1) \quad (\text{A.3})$$

$$K_2 = \frac{1}{2}((1 + \alpha) \cos(\theta_2) - 1) \quad (\text{A.4})$$

$$K_3 = \frac{1}{2}(1 + \alpha) \sin(\theta_2). \quad (\text{A.5})$$

Substituting Equations (A.3), (A.4), and (A.5) into Equation (A.2) results in the following expression for I_{IF} :

$$I_{IF} = K_1 \cos((\omega_2 + \omega_{BB})t) + K_2 \cos((\omega_2 - \omega_{BB})t) - K_3 \sin((\omega_2 + \omega_{BB})t) - K_3 \sin((\omega_2 - \omega_{BB})t). \quad (\text{A.6})$$

With this expression for I_{IF} it is now possible to continue down the signal path and generate an expression for I_{RF} , which is given by

$$\begin{aligned} I_{RF} &= I_{IF} \cos(\omega_1 t + \theta_1) \\ &= K_1 \cos((\omega_2 + \omega_{BB})t) \cos(\omega_1 t + \theta_1) + K_2 \cos((\omega_2 - \omega_{BB})t) \cos(\omega_1 t + \theta_1) \\ &\quad - K_3 \sin((\omega_2 + \omega_{BB})t) \cos(\omega_1 t + \theta_1) - K_3 \sin((\omega_2 - \omega_{BB})t) \cos(\omega_1 t + \theta_1) \\ &= \frac{1}{2} K_1 (\cos((\omega_1 + \omega_2 + \omega_{BB})t + \theta_1) + \cos((\omega_1 - \omega_2 - \omega_{BB})t + \theta_1)) \\ &\quad + \frac{1}{2} K_2 (\cos((\omega_1 + \omega_2 - \omega_{BB})t + \theta_1) + \cos((\omega_1 - \omega_2 + \omega_{BB})t + \theta_1)) \\ &\quad - \frac{1}{2} K_3 (\sin((\omega_1 + \omega_2 + \omega_{BB})t + \theta_1) - \sin((\omega_1 - \omega_2 - \omega_{BB})t + \theta_1)) \\ &\quad - \frac{1}{2} K_3 (\sin((\omega_1 + \omega_2 - \omega_{BB})t + \theta_1) - \sin((\omega_1 - \omega_2 + \omega_{BB})t + \theta_1)). \end{aligned} \quad (\text{A.7})$$

This equation shows that the signal I_{RF} has terms centered at both the sum and difference of the two LO frequencies. The only terms of interest in Equation (A.7) are the terms at the sum of the two LO frequencies because these represent the desired transmitted signal.

The other terms will get filtered by the RF image-reject mixer and the bandpass filter after the PA. Therefore, ignoring the difference terms, I_{RF} can be written as follows:

$$\begin{aligned} I_{RF} = & \frac{1}{2} K_1 (\cos((\omega_1 + \omega_2 + \omega_{BB})t + \theta_1)) + \frac{1}{2} K_2 (\cos((\omega_1 + \omega_2 - \omega_{BB})t + \theta_1)) \\ & - \frac{1}{2} K_3 (\sin((\omega_1 + \omega_2 + \omega_{BB})t + \theta_1)) - \frac{1}{2} K_3 (\sin((\omega_1 + \omega_2 - \omega_{BB})t + \theta_1)). \end{aligned} \quad (\text{A.8})$$

In Equation (A.8) it is clear that this signal contains an upper and a lower sideband which are centered around the sum of the LO frequencies. To simplify Equation (A.8) it is helpful to define the frequencies of the upper and lower sideband by the following expressions:

$$\omega_U = \omega_1 + \omega_2 + \omega_{BB} \quad (\text{A.9})$$

$$\omega_L = \omega_1 + \omega_2 - \omega_{BB}. \quad (\text{A.10})$$

Substituting Equations (A.8) and (A.9) into Equation (A.8) results in the following expression:

$$\begin{aligned} I_{RF} = & \frac{1}{2} K_1 (\cos(\omega_U t) \cos(\theta_1) - \sin(\omega_U t) \sin(\theta_1)) \\ & + \frac{1}{2} K_2 (\cos(\omega_L t) \cos(\theta_1) - \sin(\omega_L t) \sin(\theta_1)) \\ & - \frac{1}{2} K_3 (\sin(\omega_U t) \cos(\theta_1) + \cos(\omega_U t) \sin(\theta_1)) \\ & - \frac{1}{2} K_3 (\sin(\omega_L t) \cos(\theta_1) + \cos(\omega_L t) \sin(\theta_1)). \end{aligned} \quad (\text{A.11})$$

To this point the discussion has focused on the top half of Figure A.1, which has led to expressions for I_{IF} and I_{RF} . The focus now shifts to the lower portion in effort to

generate expressions for Q_{IF} and QQ_{RF} . The first step is to determine Q_{IF} , which is given by the following expression:

$$Q_{IF} = (1 + \alpha) \cos(\omega_{BB}t) \sin(\omega_2t + \theta_2) - \sin(\omega_{BB}t) \cos(\omega_2t + \theta_2) \quad (\text{A.12})$$

After some manipulation Q_{IF} can be re-written in the following form:

$$\begin{aligned} I_{IF} &= \frac{1}{2}(1 + \alpha + \cos(\theta_2)) \sin((\omega_2 + \omega_{BB})t) \\ &+ \frac{1}{2}(1 + \alpha - \cos(\theta_2)) \sin((\omega_2 - \omega_{BB})t) \\ &+ \frac{1}{2} \sin(\theta_2) \cos((\omega_2 + \omega_{BB})t) \\ &- \frac{1}{2} \sin(\theta_2) \cos((\omega_2 - \omega_{BB})t). \end{aligned} \quad (\text{A.13})$$

To simplify the previous expression it is useful to define three additional term as follows:

$$M_1 = \frac{1}{2}(1 + \alpha + \cos(\theta_2)) \quad (\text{A.14})$$

$$M_2 = \frac{1}{2}(1 + \alpha - \cos(\theta_2)) \quad (\text{A.15})$$

$$M_3 = \frac{1}{2} \sin(\theta_2). \quad (\text{A.16})$$

Substituting Equations (A.14), (A.15), and (A.16) into Equation (A.13) results in the following expression for Q_{IF} :

$$\begin{aligned} Q_{IF} &= M_1 \sin((\omega_2 + \omega_{BB})t) + M_2 \sin((\omega_2 - \omega_{BB})t) \\ &+ K_3 \cos((\omega_2 + \omega_{BB})t) - K_3 \cos((\omega_2 - \omega_{BB})t). \end{aligned} \quad (\text{A.17})$$

With this expression for Q_{IF} it is now possible to determine an expression for QQ_{RF} . Similar to Equation (A.8), only the terms centered at the sum of the LO frequencies will be included and this results in the following expression for QQ_{RF} :

$$\begin{aligned} QQ_{RF} = & -\frac{1}{2}M_1 \cos((\omega_1 + \omega_2 + \omega_{BB})t) - \frac{1}{2}M_2 \cos((\omega_1 + \omega_2 - \omega_{BB})t) \\ & + \frac{1}{2}M_3 \sin((\omega_1 + \omega_2 + \omega_{BB})t) - \frac{1}{2}M_3 \sin((\omega_1 + \omega_2 - \omega_{BB})t). \end{aligned} \quad (\text{A.18})$$

Substituting Equations (A.8) and (A.9) into Equation (A.18) results in the following expression for QQ_{RF} :

$$QQ_{RF} = -\frac{1}{2}M_1 \cos(\omega_U t) - \frac{1}{2}M_2 \cos(\omega_L t) + \frac{1}{2}M_3 \sin(\omega_U t) - \frac{1}{2}M_3 \sin(\omega_L t). \quad (\text{A.19})$$

With the expressions for I_{RF} and QQ_{RF} complete an expression for the output signal, I_{RF} , can now be determined and this is given by

$$\begin{aligned} I_{RF} = & I_{RF} - QQ_{RF} \\ = & \left(\frac{1}{2}K_1 \cos(\theta_1) - \frac{1}{2}K_3 \sin(\theta_1) + \frac{1}{2}M_1 \right) \cos(\omega_U t) \\ & - \left(\frac{1}{2}K_1 \sin(\theta_1) + \frac{1}{2}K_3 \cos(\theta_1) + \frac{1}{2}M_3 \right) \sin(\omega_U t) \\ & + \left(\frac{1}{2}K_2 \cos(\theta_1) - \frac{1}{2}K_3 \sin(\theta_1) + \frac{1}{2}M_2 \right) \cos(\omega_L t) \\ & - \left(\frac{1}{2}K_2 \sin(\theta_1) + \frac{1}{2}K_3 \cos(\theta_1) - \frac{1}{2}M_3 \right) \sin(\omega_L t) \\ = & P_1 \cos(\omega_U t) - P_2 \sin(\omega_U t) + P_3 \cos(\omega_L t) - P_4 \sin(\omega_L t) \end{aligned} \quad (\text{A.20})$$

where P_1 , P_2 , P_3 , and P_4 have been substituted for clarity. Each of these coefficients must now be expanded by substituting back in the expressions for K_1 , K_2 and K_3 and also M_1 , M_2 , and M_3 . The first coefficient, P_1 , is given by

$$\begin{aligned}
P_1 &= \frac{1}{2} K_1 \cos(\theta_1) - \frac{1}{2} K_3 \sin(\theta_1) + \frac{1}{2} M_1 \\
&= \frac{1}{4} ((1 + \alpha) \cos(\theta_2) + 1) \cos(\theta_1) \\
&\quad - \frac{1}{4} (1 + \alpha) \sin(\theta_2) \sin(\theta_1) + \frac{1}{4} (1 + \alpha + \cos(\theta_2)) \\
&= \frac{1}{4} [(1 + \alpha)(\cos(\theta_1 + \theta_2) + 1) + \cos(\theta_1) + \cos(\theta_2)].
\end{aligned} \tag{A.21}$$

The second coefficient, P_2 , is given by

$$\begin{aligned}
P_2 &= \frac{1}{2} K_1 \sin(\theta_1) + \frac{1}{2} K_3 \cos(\theta_1) + \frac{1}{2} M_3 \\
&= \frac{1}{4} ((1 + \alpha) \cos(\theta_2) + 1) \sin(\theta_1) + \frac{1}{4} (1 + \alpha) \sin(\theta_2) \cos(\theta_1) + \frac{1}{4} \sin(\theta_2) \\
&= \frac{1}{4} [(1 + \alpha)(\cos(\theta_1 + \theta_2) + 1) + \cos(\theta_1) + \cos(\theta_2)].
\end{aligned} \tag{A.22}$$

The third coefficient, P_3 , is given by

$$\begin{aligned}
P_3 &= \frac{1}{2} K_2 \cos(\theta_1) - \frac{1}{2} K_3 \sin(\theta_1) + \frac{1}{2} M_2 \\
&= \frac{1}{4} ((1 + \alpha) \cos(\theta_2) - 1) \cos(\theta_1) \\
&\quad - \frac{1}{4} (1 + \alpha) \sin(\theta_2) \sin(\theta_1) + \frac{1}{4} (1 + \alpha - \cos(\theta_2)) \\
&= \frac{1}{4} [(1 + \alpha)(\cos(\theta_1 + \theta_2) + 1) - \cos(\theta_1) - \cos(\theta_2)].
\end{aligned} \tag{A.23}$$

The fourth coefficient, P_4 , is given by

$$\begin{aligned}
P_4 &= \frac{1}{2} K_2 \sin(\theta_1) + \frac{1}{2} K_3 \cos(\theta_1) - \frac{1}{2} M_3 \\
&= \frac{1}{4} ((1 + \alpha) \cos(\theta_2) - 1) \sin(\theta_1) + \frac{1}{4} (1 + \alpha) \sin(\theta_2) \cos(\theta_1) - \frac{1}{4} \sin(\theta_2) \\
&= \frac{1}{4} [(1 + \alpha) \sin(\theta_1 + \theta_2) - \sin(\theta_1) - \sin(\theta_2)].
\end{aligned} \tag{A.24}$$

The image rejection of the HRT is determined the ratio of the power of signal at the desired sideband to the power of the unwanted sideband. In this case the upper sideband is the desired sideband and therefore the image rejection is given by

$$IR_{HRT} = \frac{P_1^2 + P_2^2}{P_3^2 + P_4^2}. \quad (\text{A.25})$$

To evaluate this equation it is helpful to evaluate the numerator and the denominator separately. The numerator is given by

$$\begin{aligned} P_1^2 + P_2^2 = & \frac{1}{16} \left[(1 + \alpha)^2 (\cos^2(\theta_1 + \theta_2) + 2 \cos(\theta_1 + \theta_2) + 1) \right. \\ & + 2(1 + \alpha)(\cos(\theta_1 + \theta_2) + 1)(\cos(\theta_1) + \cos(\theta_2)) \\ & \left. + \cos^2(\theta_1) + \cos^2(\theta_2) + 2 \cos(\theta_1) \cos(\theta_2) \right] \\ & + \frac{1}{16} \left[(1 + \alpha)^2 \sin^2(\theta_1 + \theta_2) \right. \\ & + 2(1 + \alpha) \sin(\theta_1 + \theta_2)(\sin(\theta_1) + \sin(\theta_2)) \\ & \left. + \sin^2(\theta_1) + \sin^2(\theta_2) + 2 \sin(\theta_1) \sin(\theta_2) \right]. \end{aligned} \quad (\text{A.26})$$

After some simplification and manipulation of the terms, Equation (A.26) can be rewritten as follows:

$$\begin{aligned} P_1^2 + P_2^2 = & \frac{1}{16} \left[1 + \cos(\theta_1 - \theta_2) + (1 + \alpha)^2 (1 + \cos(\theta_1 + \theta_2)) \right. \\ & \left. + 2(1 + \alpha)(\cos(\theta_1) + \cos(\theta_2)) \right]. \end{aligned} \quad (\text{A.27})$$

The next step in determining the image rejection is to evaluate the denominator of Equation (A.25), which is given by

$$\begin{aligned}
P_3^2 + P_4^2 = & \frac{1}{16} \left[(1 + \alpha)^2 (\cos^2(\theta_1 + \theta_2) + 2 \cos(\theta_1 + \theta_2) + 1) \right. \\
& - 2(1 + \alpha)(\cos(\theta_1 + \theta_2) + 1)(\cos(\theta_1) + \cos(\theta_2)) \\
& \left. + \cos^2(\theta_1) + \cos^2(\theta_2) + 2 \cos(\theta_1) \cos(\theta_2) \right] \\
& + \frac{1}{16} \left[(1 + \alpha)^2 \sin^2(\theta_1 + \theta_2) \right. \\
& - 2(1 + \alpha) \sin(\theta_1 + \theta_2)(\sin(\theta_1) + \sin(\theta_2)) \\
& \left. + \sin^2(\theta_1) + \sin^2(\theta_2) + 2 \sin(\theta_1) \sin(\theta_2) \right].
\end{aligned} \tag{A.28}$$

After some simplification and manipulation of the terms, Equation (A.28) can be rewritten as follows:

$$\begin{aligned}
P_3^2 + P_4^2 = & \frac{1}{16} \left[1 + \cos(\theta_1 - \theta_2) + (1 + \alpha)^2 (1 + \cos(\theta_1 + \theta_2)) \right. \\
& \left. - 2(1 + \alpha)(\cos(\theta_1) + \cos(\theta_2)) \right].
\end{aligned} \tag{A.29}$$

With the expression for the numerator and the denominator complete, an expression for the image rejection can now be determined. Substituting Equations (A.27) and (A.29) into Equation (A.25) yields the following expression for the image rejection of the harmonic-rejection transmitter:

$$IR_{HRT} = \frac{1 + \cos(\theta_1 - \theta_2) + (1 + \alpha)^2 (1 + \cos(\theta_1 + \theta_2)) + 2(1 + \alpha)(\cos(\theta_1) + \cos(\theta_2))}{1 + \cos(\theta_1 - \theta_2) + (1 + \alpha)^2 (1 + \cos(\theta_1 + \theta_2)) - 2(1 + \alpha)(\cos(\theta_1) + \cos(\theta_2))}. \tag{A.30}$$

References

- [1] P. R. Gray and R. G. Meyer, "Future Directions of Silicon IC's for RF Personal Communications," *Proceeding of the IEEE 1995 Custom Integrated Circuit Conference*, pp. 83-90, May 1995.
- [2] ETSI, "Global System for Mobile Communications", Digital cellular telecommunications system (Phase 2+); Radio transmission and reception, GSM 5.05, Version 5.2.0, Copyright 1996.
- [3] M. E. Valkenburg, *Analog Filter Design*. New York: Oxford University Press, 1982.
- [4] T. Yamawaki et al, "A 2.7-V GSM RF Transceiver IC," *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 2089-2096, Dec. 1997.
- [5] L. R. Kahn, "Single-sideband transmission by envelope elimination & restoration," *Proc. IRE*, July 1952, pp. 803-806.
- [6] W. Sander, S.Schell, an dB. Sander, "Polar modulator for multi-mode cell phones," *Proc. CICC*, Sept. 2003, pp. 439-445.
- [7] M. Elliot, T. Montalvo, B. Jeffries, F. Murden, J. Strange, A. Hill, S, Nandipaku, and J. Harrebek, "A polar modulator transmitter for GSM/EDGE," *IEEE J. Solid-State Circuits*, vol. 39, pp. 2190-2199, Dec. 2004.
- [8] P. Nagle, P. Burton, E. Heaney, and F. McGrath, "A wide-band linear amplitude modulator for polar transmitters based on the concept of interleaving delta modulation," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 1748-1756, Dec. 2002.
- [9] T. Sowlati, D. Rozenblit, R. Pallela, M. Damgaard, E. McCarthy, D. Koh, D. Ripley, F. Balteanu and I. Gheorghe, "Quad-band GSM/GPRS/EDGE polar loop transmitter," *IEEE Journal of Solid-State Circuits*, vol. 39, pp. 2179-2189, Dec. 2004.
- [10] B. Razavi, *RF Microelectronics*, Upper Saddle River, NJ: Prentice-Hall, 1998.

- [11] K. Negus, B. Koupal, J. Wholey, K. Carter, D. Millicker, C. Snapp, N. Marion; "Highly-integrated transmitter RFIC with monolithic narrowband tuning for digital cellular handsets," *IEEE International Solid-State Circuits Conference*, vol. 27, pp. 38 - 39, February 1994.
- [12] T. D. Stetzler, I. Post, J. Havens and M. Koyama, "A 2.7-4.5 V single chip GSM transceiver RF integrated circuit," *IEEE Journal of Solid-State Circuits*, vol. 30, pp. 1421-1429, Dec. 1995.
- [13] H. Darabi, S. Khorram, H. Chien, M. Pan, S. Wu, S. Moloudi, J. C. Leete, J. J. Rael, M. Syed, R. Lee, B. Ibrahim, M. Rofougaran, and A. Rofougaran, "A 2.4-GHz CMOS Transceiver for Bluetooth," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 2016-2024, Dec. 2001.
- [14] K. Lee, S. Lee, Y. Koo, H. Huh, H. Nam, J. Lee, J. Park, K. Lee, D. Jeong and W. Kim, "Full-CMOS 2-GHz WCDMA direct conversion transmitter and receiver," *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 43-53, Jan. 2003.
- [15] B. Gilbert, "A precise four-quadrant multiplier with subnanosecond response," *IEEE Journal of Solid-State Circuits*, vol. SC-3, pp. 365-373, Dec. 1968
- [16] A. V. Oppenheim, A. V. Willsky and I. T. Young, *Signals and Systems*. New Jersey: Prentice Hall, 1983.
- [17] J. C. Rudell, Frequency Translation Techniques for High-Integration High-Selectivity Multi-Standard Wireless Communication Systems, Ph.D. Dissertation, University of California, Berkeley, 2000.
- [18] P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*. New York: John Wiley, 1993.
- [19] R. G. Meyer, Class Notes, *EECS 242: Advanced Integrated Circuits for Communications*, University of California, Berkeley, Spring 1997.
- [20] M. T. Terrovitis, "Noise in Current-Commutating CMOS Mixers," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 1461-1473, Oct. 2000.
- [21] M. T. Terrovitis, "Intermodulation Distortion in Current-Commutating CMOS Mixers," *IEEE Journal of Solid-State Circuits*, vol. 35, pp. 772-783, June 1999.

- [22] J. W. Archer, J. Granlund, and R. E. Mauzy, "A broad-band VHF mixer exhibiting high image rejection over a multidecade baseband frequency range," *IEEE Journal of Solid-State Circuits*, vol. 16, pp. 385 - 392, August 1981.
- [23] D. Pache, J. M. Fournier, G. Billiot, and P. Senn, "An improved 3V 2GHz BiCMOS image reject mixer IC," *1995 IEEE Custom Integrated Circuits Conference*, pp. 95 - 98, May 1995.
- [24] M. D. McDonald, "A 2.5GHz BiCMOS image-reject front-end," *IEEE International Solid-State Circuits Conference*, vol. 26, pp. 144-145, February 1993.
- [25] M. Gingell, "Single Sideband Modulation using Sequence Asymmetric Polyphase Networks," *Electrical Communication*, pp. 21-25, 1973.
- [26] R. C. V. Macario and I. D. Mejallie, "The Phasing Method for Sideband Selection in Broadcast Receivers," *EBU Review (Technical Part)*, no. 181, pp.119-125, 1980.
- [27] F. Behbahani, Y. Kishigami, J. Leete, and A. A. Abidi, "CMOS Mixers and polyphase filters for large image rejection," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 873 - 887, June 2001.
- [28] J. Crols and M. Steyaert, *CMOS Wireless Transceiver Design*. Massachusetts: Kluwer Academic Publishers, 1997.
- [29] J. W. Archer, J. G. Granlund and R. E. Mauzy, "A Broad-Band UHF Mixer Exhibiting High Image Rejection over a Multidecade Baseband Frequency Range," *IEEE Journal of Solid-State Circuits*, Vol. SC-16, pp. 385-392, August 1981.
- [30] J. C. Rudell, J. -J. Ou, T. B. Cho, G. Chien, F. Brianti, J. A. Weldon, and P. R. Gray, "A 1.9 GHz Wide-Band IF Double Conversion CMOS Receiver for Cordless Telephone Applications," *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 2071-2088, December 1997.
- [31] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proceedings of the IEEE*, vol. 54, pp. 329-330, February 1996.
- [32] L. Lin, L. Tee, and P. Gray, "A 1.4 GHz Differential Low-Noise CMOS Frequency Synthesizer using a Wideband PLL Architecture," *Digest of Technical Papers, 2000 International Solid-State Circuits Conference*, vol. 43, pp. 204-205.

-
- [33] D. H. Wolaver, *Phase-Locked Loop Circuit Design*, New Jersey: Prentice Hall, 1991.
- [34] L. Lin, "Design Techniques for High-Performance Integrated Frequency Synthesizers for Multi-Standard Wireless Communication Applications," Ph.D. Dissertation, Electronics Research Laboratory, University of California, Berkeley, 2000.
- [35] J. A. Weldon, R. S. Narayanaswami, J. C. Rudell, L. Lin, M. Otsuka, S. Dedieu, L. Tee, K. Tsai, C. Lee, and P. R. Gray, "A 1.75-GHz Highly Integrated Narrow-band CMOS Transmitter with Harmonic-Rejection Mixers", *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 2003 - 2015, December 2001.
- [36] R. S. Narayanaswami, *RF CMOS Class C Power Amplifiers for Wireless Communications*, Ph.D. Dissertation, University of California, Berkeley, Electronics Research Laboratory, Memorandum No. UCB/ERL M01/37, 2001.
- [37] K. C. Tsai, P. R. Gray, "A 1.9 GHz CMOS Class E Power Amplifier for Wireless Communications," *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 962 - 970, July 1999.