RADIO-FREQUENCY CONVERSION AND SYNTHESIS (FOR A 115 MILLIWATT GPS RECEIVER)

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Arvin R. Shahani

March 1999

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

T.H. Lee (Principal Adviser)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

B.A. Wooley

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and in quality, as a dissertation for the degree of Doctor of Philosophy.

D.C. Cox

Approved for the University Committee on Graduate Studies:

Abstract

Recently, there has been a proliferation of wireless devices and services. These services have evolved from simple paging functions to voice applications (i.e., cellular telephony). One system that is poised to make an impact in the wireless arena is the Global Positioning System (GPS). In addition to position information, GPS also provides an accurate time reference. Both types of information are valuable not only to a mobile user, but also in the design and operation of other wireless systems.

The work in this dissertation is part of a complete front-end for a GPS receiver that was fabricated in a 0.5μm CMOS process. The focus of this project is on CMOS hardware techniques that are low power and suitable for radio-frequency integrated circuits. Thus, not only are the architectures that we present well-suited for a GPS receiver, they are also applicable to other systems that require low-power, radio-frequency circuits. The two front-end functions highlighted in this work are frequency conversion and synthesis.

The CMOS voltage mixer takes advantage of good voltage switches in the form of CMOS transistors, as well as the mixer's integrated environment. This voltage-domain passive mixer operates on a negligible amount of power, while achieving noise figures that are lower than any alternative CMOS mixer architecture. It additionally possesses excellent linearity.

The new phase-locked loop (PLL) architecture synthesizes a gigahertz local oscillator (LO) from a low-frequency reference without using a divide-by-N block in the PLL's feedback loop. In doing so, the power consumption of the loop is substantially lower than that of conventional techniques. The power consumption can be reduced even further if a power-efficient voltage-controlled oscillator (VCO) is used. A further benefit of removing the divider is the reduction of on-chip interference caused

by the divider's large high-speed transitions. The tradeoff with this architecture is greater acquisition complexity in exchange for the lower power consumption and reduced interference in lock.

Acknowledgments

It all began in the fall of 1993; that's when I had EE214 which was being taught by Tom Lee. In class, Tom was a cross between David Letterman[1] and Phineas J. Whoopee[2]. Needless to say, Tom was inspiring.

Tom continued to be an inspiration throughout my research. He did this by working harder than anyone in our group, by amusing us with anecdotes of his hobbyist activities (e.g., baking pennies[3] in his home oven to make diodes, and then using them in radio circuits), by his enthusiasm for our area of research, and by his broad knowledge of not only engineering, but also music, language, and great books (Accidental Empires, Journey Through Genius, etc.). I can truly say that during my PhD years at Stanford, Tom was a major contributor to my development, both as a researcher and as a person, and I am glad I had the opportunity to know him.

While Tom played the role of inspiring mentor, Derek Shaeffer set the standard for me among my peers. I recognized, as did many people, that Derek was truly gifted with circuits. I feel privileged to have had the opportunity to work with him throughout my time in the PhD program. In addition to being the main person I looked to for technical advice and collaboration, he was also a close, personal friend. Derek, I thank you for all your encouragement and support.

The year before I joined the PhD program (1994-1995), I was very fortunate to be Professors Saraswat and Wooley's teaching assistant. My talks with them convinced me to stay on for my PhD at Stanford. Also, Professor Wooley was kind enough to help me make the transition to research by inviting me to his group meetings, since Tom's group was in the process of being formed. I am grateful to Professor

[1]The Late Show.

[2]The man who knew everything in the Tennessee Tuxedo an Chumley cartoon.

[3]Pennies after 1983 are preferred.

Wooley for being my associate advisor and for lending his keen editorial pen to my thesis. I would also like to thank Professor Cox for doing double duty by chairing my orals committee and reading my thesis. I wish to also express my appreciation to Professors Saraswat and Wong for being on my orals committee.

I wish to thank all the people, from the various research groups, that made CIS a great place to work. I thank every member of SMIrC, Tamara Ahrens, Rafael Betancourt, David Colleran, Ramin Farjad-Rad, Joel Dawson, Ali Hajimiri, Mar Hershenson, S.S. Mohan, Hamid Rategh, Hirad Samavati, Derek Shaeffer, and Kevin Yu, for the insightful and educational discussions I shared with them. What also made SMIrC special was the collaborative projects we pursued and I would like to express my thanks to those teams I was a part of: Fastlane (Derek and Mohan), Waldo PLL (Hamid and Hirad), and Waldo (Derek, Mohan, Mar, Min, Patrick, and Dan). Outside of my own research group, there were many people who enriched my experience at CIS. I thank Katayoun Falakshahi, Joe Ingino, Adrian Ong, Gu-Yeon Wei, and Patrick Yue as professional colleagues and friends.

Ann Guerra has been especially helpful to me during my years in CIS. Ann has always said that she is just doing her job, but I'd like to thank her for the way she does her job. She removed the stress associated with administrative paper work, reimbursements, and mailing conference and journal submissions, so that I could focus on research. I also owe Pauline Prather many thanks for the nights she spent bonding dies in packages for me. Her services were invaluable during the testing phase of my projects.

One of the great benefits Stanford had to offer was the involvement with industry that one could have. I was fortunate to have had this opportunity on many occasions. For most of my PhD work, I was supported by an IBM fellowship, for which I thank IBM and especially Dr. Mehmet Soyuer, Dr. Modest Oprysko, and Dr. John Ewen. Some of the projects I mentioned could not have been done without the help of industry partners and key individuals such as Dan Dobberpuhl at DEC, Norm Hendrickson at Vitesse, Drs. Chris Hull, Paramjit Singh, and Julian Tham and Steve Lloyd at Rockwell. I would also like to thank Dr. Gunnar Björklund at Ericsson and Dr. Herbert Eichfeld at Siemens for providing me the opportunity to

share my research with their organizations through the SPIE program and for being excellent hosts.

Fortunately not all my time during my PhD program was spent doing research. I am sincerely grateful to my friends outside of Stanford, who allowed me to escape and recover from my thesis work. It turned out to be a blessing that most of them had no clue what I was researching at Stanford, because it forced me to take a break; so, I wish to thank Ravi, Rajas, Vic, Roy, Priya, Smruti, Anjali, Marsha, Steve, Mark, Heidi, Annie, Debbie, and Michael. To those I forgot, you get a beer.

Finally, I want to thank my family for their unconditional support. They were always willing to listen to me and encourage me. I am most proud to be able to share this thesis with them and so I dedicate it to my parents, Ramesh and Gulshan, and sister, Annik.

Contents

List Of Abbreviations

AGC	automatic gain control
APD	aperture phase detector
API	analog phase interpolation
BER	bit error rate
CP	charge pump
DSP	digital signal processing
DUT	device under test
ENR	excess noise ratio
FAA	frequency acquisition aid
FFT	fast Fourier transform
FM	frequency modulation
GPS	global positioning system
IC	integrated circuit
IF	intermediate frequency
IFA	intermediate-frequency amplifier
IP3	third-order intermodulation product
IIP3	input-referred third-order intermodulation product
LAN	local-area network
LNA	low-noise amplifier
LTI	linear time-invariant
LO	local oscillator
MSB	multi sideband
PA	power amplifier
PFD	phase/frequency detector
PHS	personal handyphone system
PLL	phase-locked loop
PM	phase modulation
PSD	power spectral density
RF	radio frequency
SAW	surface acoustic wave
SNR	signal-to-noise ratio

SSB	single sideband
VCO	voltage-controlled oscillator
WAN	wide-area network

List Of Tables

List Of Figures

Chapter 1

Introduction

IN the last 15 years, there has been renewed interest in wireless applications [1]. The popularity of cellular telephones, pagers, and cordless telephones has sparked manufacturers to enter additional wireless markets. Currently, wireless local-area networks (LANs) for data communications are being deployed. Furthermore, new spectrum is being allocated around 2.5GHz and between 5–6GHz for wireless systems, and many new applications are being proposed (e.g., wireless computer peripherals). But to accomplish these objectives of wireless connectivity, wireless transceivers are needed.

Let us take a step back from all the activity just described, and investigate the frequencies suitable for wireless transceivers. When we account for antenna physics and the radio propagation environment a working frequency range spanning 400MHz to 10GHz is found, with a broad optimum in the range of 500MHz to 1GHz [2]. Notice a correlation between new frequency allocations, given in the preceding paragraph, and the suitable frequencies identified in this paragraph.

Frequencies in the gigahertz range were thought to be the stronghold of GaAs technologies, so people believed that realization of the wireless transceiver's radio-frequency (RF) circuits in an integrated circuit (IC) required using these same GaAs technologies [3]. Recently however, less expensive SiGe, silicon bipolar, and BiCMOS technologies have been challenging this belief [4], especially for 900MHz applications. Because we recognize this trend, one of our goals in this dissertation is to demonstrate

CMOS technology's viability for RFICs. The benefits of using CMOS are in the areas of cost and integration, which are discussed more in Chapter 2.

Because one goal of this work is to demonstrate CMOS's viability for RFICs and hence for single-chip radios, it is important to show that high-performance RF circuits can be designed in CMOS. These circuits include low-noise amplifiers (LNAs), power amplifiers (PAs), mixers, and voltage-controlled oscillators (VCOs). Once these blocks have been produced, the wireless transceiver's complete analog circuitry can be designed in CMOS to produce a single-chip radio.

This work presents a CMOS mixer and frequency synthesizer. These designs were combined with other blocks, as part of a joint project, to produce the radio for a Global Positioning System (GPS) receiver [5]. The radio demonstrates a low-power, highly-integrated CMOS solution, and since GPS operates near 1.6GHz, it lends credibility for choosing CMOS for other applications in this frequency range.

The architectures for both the mixer and synthesizer are new and exploit CMOS technology, as well as the integrated environment. The primary benefit of these architectures is reduced power consumption, which is critical in portable electronics where battery life is of fundamental concern. While the architectures are used here specifically for GPS, they are applicable in general.

1.1 Organization

Chapter 2 discusses radio integration. Its purpose is to show that CMOS technology is suitable for the implementation of a single-chip radio. A specific system is selected as a test vehicle: a GPS coarse-acquisition (C/A) code radio. Therefore, Chapter 3 provides background on GPS, and elucidates important GPS receiver requirements. Chapter 3 concludes with the radio's architecture and highlights the portions of that architecture contained in this thesis.

Chapter 4 treats CMOS mixers. It begins by presenting the performance metrics used to evaluate mixers, and then uses the metrics to compare recently published implementations. From this comparison, the CMOS voltage mixer is determined to be the right mixer for the GPS radio application at hand. From this point, Chapter 4

focuses exclusively on the voltage mixer, examining the topics of local oscillator (LO) drive, conversion gain, noise figure, and linearity. Chapter 5 complements Chapter 4 by documenting the design of two experimental CMOS voltage mixers, including the design for the radio's mixer. Chapter 5 also provides experimental data on the voltage mixer. Relevant calculations and simulations are included.

Chapter 6 presents a method for frequency synthesis that is phase-locked loop (PLL) based, except that a divider is not necessary in the loop's feedback path. Dividerless frequency synthesis is accomplished with a new phase detector, called the aperture phase detector (APD). The APD's function in a PLL is mathematically described and its ideal input/output characteristic is found. Behavior specific to an APD PLL is also presented. Similar to the relationship between Chapter 4 and Chapter 5, Chapter 7 complements Chapter 6 by documenting the design of an experimental APD PLL. Relevant simulations and loop modeling are included. Chapter 7 also contains a record of the measured data taken from tests on the APD PLL. Verification of the theory presented in Chapter 6 is provided by the measured data in Chapter 7.

Chapter 2

Radio Integration

THE holy grail for CMOS RF circuit designers is a single-chip CMOS transceiver. A modern transceiver is composed of two parts: the radio and the baseband processor, most often a digital signal processing (DSP) engine. CMOS is the only technology that can entertain this goal, because it is the undisputed choice for DSP. But before a single-chip transceiver is made, we consider an intermediate goal first: a single-chip CMOS radio. Once this challenge is achieved, the mixed-signal interference issue between digital circuits and sensitive RF circuits can be addressed to integrate radio and digital processing functions in a single IC.

2.1 A Personal Handyphone System (PHS) Radio

Figure 2.1 [6] is a simplified block diagram of a PHS radio. It is typical of a modern gigahertz radio. In addition to identifying the radio's blocks, Figure 2.1 also indicates the technologies used to fabricate the blocks[1]. In this case, the design leverages the benefits of three technologies: GaAs MESFET, bipolar, and BiCMOS technologies. GaAs MESFET technology is used for high-frequency blocks near the antenna (such as the LNA and PA), bipolar technology is used for the frequency conversion and frequency synthesis portions of the radio, and BiCMOS is used in subsequent

[1]Figure 2.1 also shows some discrete components (i.e., filters) in the radio. There is always an interest in subsuming these discretes onto an IC.

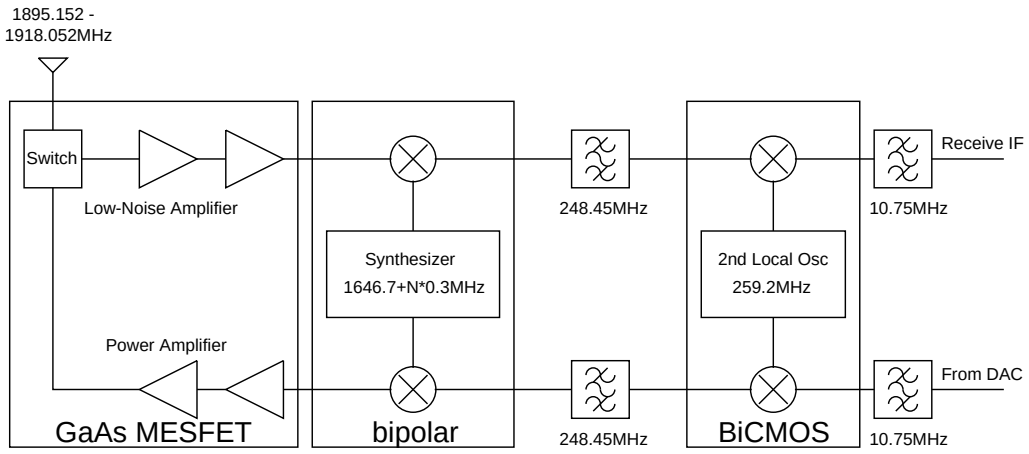


Figure 2.1: Simplified block diagram of PHS radio

intermediate frequency (IF) stages [7]. It is easy from technical arguments to justify this technology mix. GaAs MESFETs currently have the lowest device noise figures at radio frequencies, hence LNAs made with them will have the lowest noise figures, assuming all other parameters are the same. Bipolar transistors have much lower $\frac{1}{f}$ noise corners than their rival FET cousins, and because it has been widely believed that a device's $\frac{1}{f}$ noise corner will dictate the $\frac{1}{f^3}$ phase noise corner of a VCO built with that device [8], bipolar transistors are chosen for VCOs with stringent phase noise requirements. Rounding out the myriad technologies, BiCMOS proves its usefulness at low frequencies where CMOS transistors can be exploited, complemented by bipolars for potential further downconversion.

Recent research in the field of CMOS RF circuits, coupled with technology scaling trends, challenges the technology partition just discussed, and motivates a single-chip CMOS radio. Figure 2.2 [9] clearly shows that CMOS's performance is advancing at a faster rate than any competing technology, and will soon surpass the other technologies. Furthermore, [10] provides a method for a power-constrained LNA design to achieve optimum LNA noise figure, and scaling provides monotonically decreasing LNA noise figures. There have also been fundamental advances in the understanding of oscillator phase noise. The specific mechanism by which device $\frac{1}{f}$ noise turns into phase noise has finally been identified [11]. It is now known that the phase noise due to device $\frac{1}{f}$ noise (the $\frac{1}{f^3}$ portion) can be completely eliminated

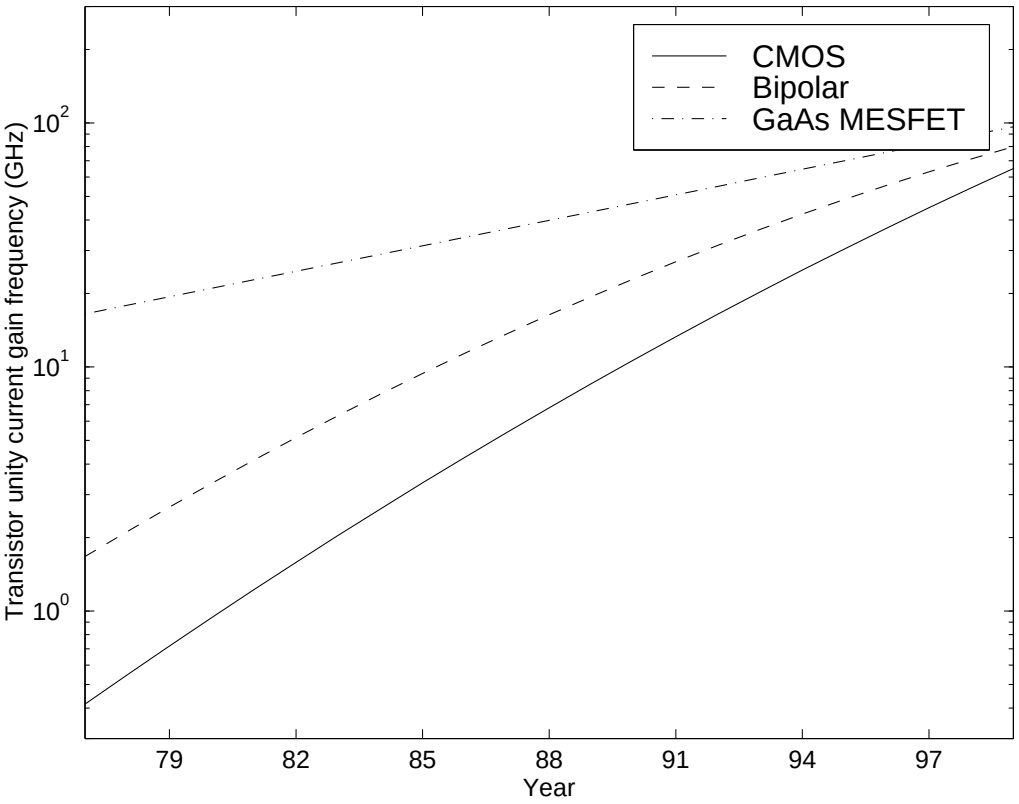


Figure 2.2: Technology scaling

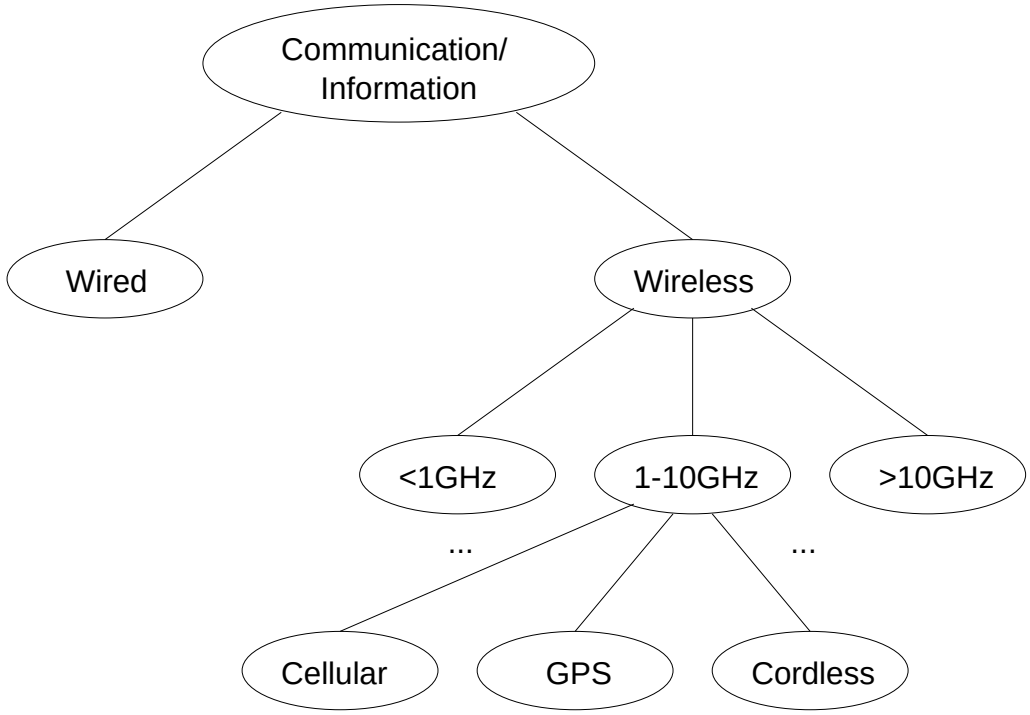


Figure 2.3: Communication and information systems

by satisfying certain symmetry properties in the oscillator. Thus, the time appears right to attempt a single-chip CMOS radio.

2.2 Systems

There are countless systems today for conveying voice, data, and video. Figure 2.3 categorizes these systems. We are primarily interested in systems operating in the 1-10GHz frequency range for demonstrating CMOS's feasibility. Examples of these types of systems are cellular telephony, cordless telephony, paging systems, and wireless data systems (e.g., wide-area networks (WANs), LANs, and GPS). These systems can be classified into high-tier and low-tier systems [12] based on system performance and the corresponding technical challenges to meet system guidelines. For example, cellular telephony is considered a high-tier system, whereas cordless telephony is a low-tier system. GPS is a low-tier system, but it is the most attractive

from a research point-of-view, since it only involves a receiver (CMOS PAs have still not been proven feasible for large transmit powers) and because it allows tremendous flexibility in architectural choices as well as individual block performance.

Chapter 3

The Global Positioning System

THE Global Positioning System is a satellite network whose purpose is to allow the determination of position and time anywhere in the world. The network consists of 24 satellites, 4 in each of 6 orbital planes, and became fully operational at the end of 1994 [13]. Each satellite broadcasts its estimated position and corresponding time, which allows a receiver to determine its range from the satellite. If the receiver is able to view four satellites, then the receiver can determine its position (latitude, longitude, and altitude) and time, the latter to atomic-clock precision. If the receiver happens to have a built-in altimeter so that altitude is known, then only three satellites need to be used to find the three remaining unknowns.

3.1 Signal Structure

There are two 20MHz wide frequency bands reserved for GPS satellite broadcasts: the L1 band, centered at 1.57542GHz, and the L2 band, centered at 1.2276GHz. Each of the 24 satellites transmits across the entire range of both bands. This broadcast scheme is possible because the signals are direct-sequence spread-spectrum signals. Direct-sequence spread spectrum is accomplished by multiplying a data stream (sequence of bits) with a code stream. The code stream uniquely identifies each satellites data stream making the shared use of the same spectrum possible. For GPS, a single data bit spans 20,460 code bits.

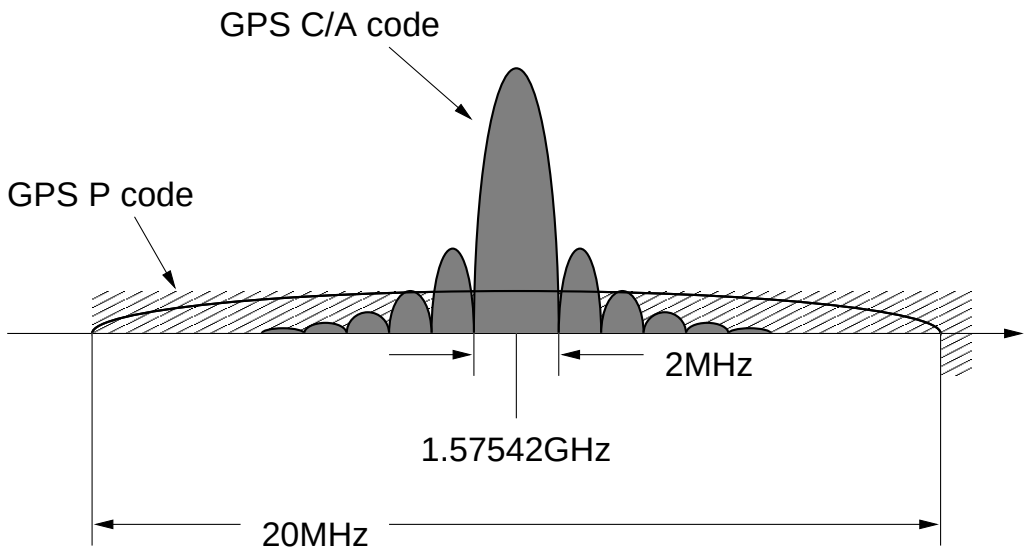


Figure 3.1: Transmitted GPS signal's PSD in the L1 band

Each satellite sends two different direct-sequence spread-spectrum signals. One signal is intended for military use only and the other is provided for civilian use. The military signal is constructed by multiplying the GPS navigation data by the precision code, or P code for short. The civilian signal multiplies by the coarse-acquisition code, or C/A code. The GPS navigation data is 50b/s, the P code is 10.23Mb/s, and the C/A code is 1.023Mb/s. These rates account for the power spectral density (PSD) shapes of the P-code and C/A-code signals shown in Figure 3.1. Notice that each code is transmitted in the L1 band. In the L2 band, only the military signal is transmitted. For this reason, the 3-dimensional rms position accuracy of the P-code signal is 16m, whereas it is only 22m for the C/A-code signal [13]. Furthermore, the C/A-code signal's accuracy can degrade to 72m at the discretion of the system's military operators. [1] Regardless, we are only interested in the C/A-code signal as it is the one intended for civilian use. Thus, we orient ourselves to receiving and processing a signal at 1.57542GHz.

[1]The purposeful degradation of the C/A-code, referred to as selective availability, will be abandoned in the future to assist commercial GPS applications.

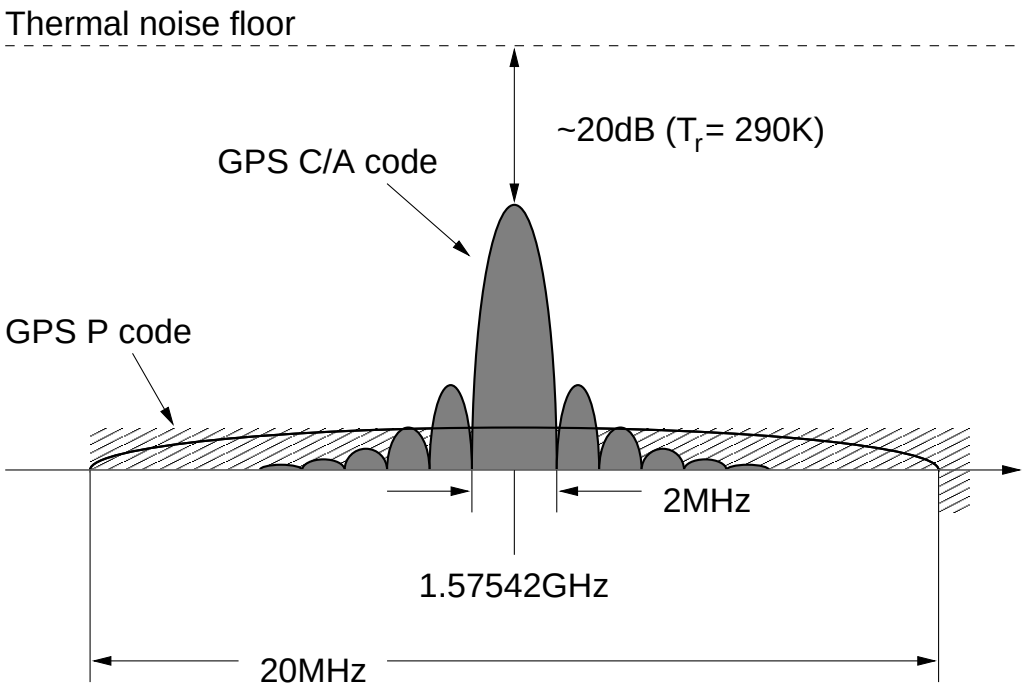


Figure 3.2: PSD of the received GPS signal and noise at the antenna

Figure 3.2 pictures the received L1 band signal at the antenna. The C/A-code signal's received signal power is typically -130dBm, while the noise power over the 20MHz L1 band is ≈ -101dBm. But from Figure 3.2, it is clear that most of the C/A-code signal's power lies in a 2MHz bandwidth occupied by its mainlobe. By simply restricting our bandwidth of interest to this 2MHz band (a factor of 10 narrower bandwidth), there is a 10dB improvement in noise power (≈ -111dBm). The bit error rate (BER) of the demodulated 50b/s GPS navigation data is insignificantly degraded by processing only the C/A code's mainlobe [14]. But even though reducing the bandwidth of interest from 20MHz to 2MHz decreases the amount of noise power, we are still faced with a raw signal-to-noise ratio (SNR) of ≈ -19dB at the antenna.

A negative SNR is still tolerable here because correlation of the received signal with the right satellite codes causes an increase in signal power over the noise power. The correlation process allows a GPS navigation bit's decision to be made after viewing $\frac{T_d}{T_c}$ code periods, where T_d is the navigation data bit period and T_c is the code bit period. Since the data bit is the same in each code interval, this process is equivalent to increasing the signal power in one code period by $\frac{T_d}{T_c}$. In contrast, viewing the noise in $\frac{T_d}{T_c}$ code periods does not reinforce the noise power, since noise is random. Thus, correlation gives rise to a processing gain,

$$G_P = 10\log_{10}\left(\frac{T_d}{T_c}\right) \approx 10\log_{10}\left(\frac{20ms}{1\mu s}\right) = 43dB, \tag{3.1}$$

which is approximately 43dB for the C/A code. So, if we have an antenna with an effective temperature of 290K and a noiseless receiver, then the SNR after the correlation process will be 24dB.

Unfortunately, of course, our receiver will not be noiseless. Figure 3.3 illustrates the radio's contribution to the noise level by showing the GPS signal at the IF after downconversion and IF processing. The amount of noise introduced by the radio, quantified by the radio's noise figure, will subtract from the correlator's processing gain. For this reason, it is imperative to keep the radio's noise figure low.

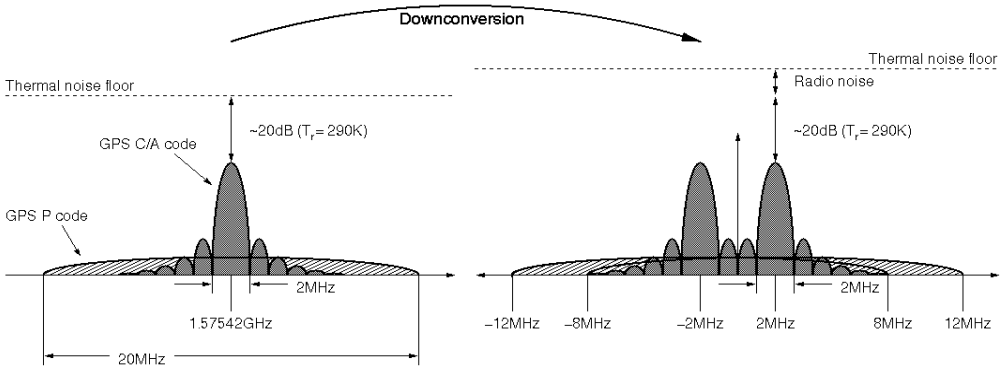


Figure 3.3: Effect of GPS radio

The wireless channel's disparate nature provides further justification for a low-noise radio. The radio's noise figure once the radio is designed is constant during operation, but the signal strength at the antenna can vary significantly due to fading, urban and real canyons, and other effects unique to wireless channels. Clearly this spread in received signal power affects the received SNR, so designing a low-noise radio that minimizes the radio's constant noise contribution is needed to accommodate wireless transmission.

3.2 GPS Receivers

In addition to a low noise figure, we also desire low power consumption. Because the purpose of GPS is to provide position information, a user's GPS receiver will most likely be portable. In any portable device battery life is of paramount concern, demanding low-power electronics. Thus the design of the GPS radio, containing the architectures presented in this thesis, focuses on realizing a low-power, low-noise GPS radio.

Figure 3.4 presents the implemented GPS radio architecture. To address the need for a low noise figure, the first block in the signal path is a low-noise amplifier (LNA). This LNA relaxes the noise figure requirements of subsequent blocks in the chain. In addition, development of new architectures for standard radio functions

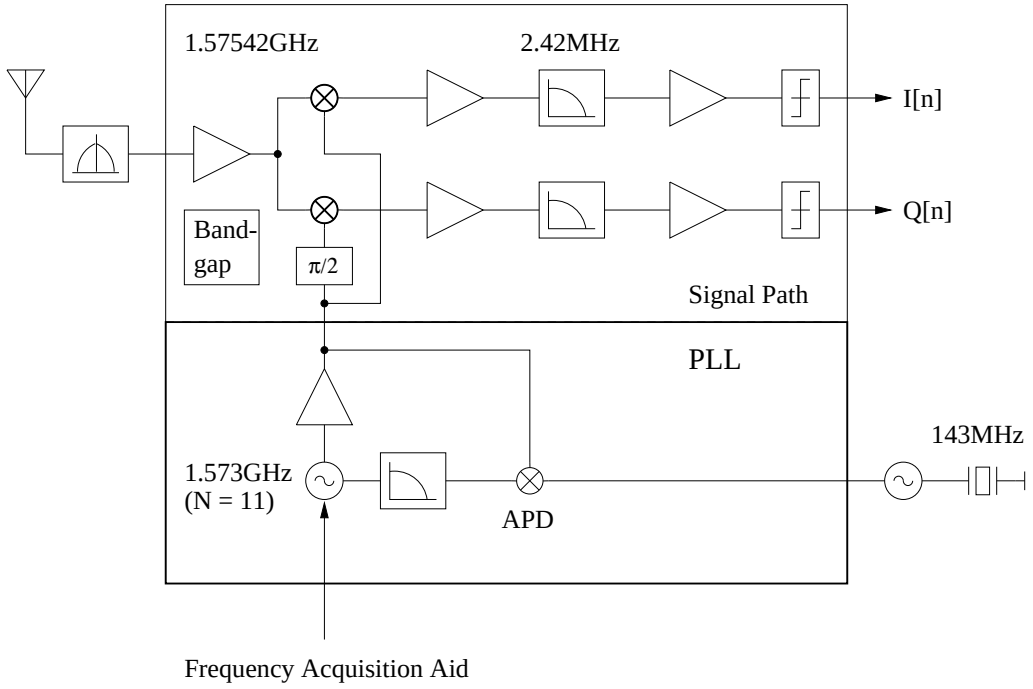


Figure 3.4: GPS radio architecture

addresses the low-power issue. In this case, those functions are frequency conversion and synthesis.

The corresponding blocks in Figure 3.4 highlight the two architectures discussed in this thesis. The architecture proposed for the two mixers, which immediately follow the LNA, has excellent power consumption with an acceptable noise figure. Each mixer converts its input signal at 1.57542GHz to a low IF of 2.42MHz (for a more detailed discussion of the low-IF architectural choice refer to [15], which is complementary to this thesis). This frequency plan requires the LO for each mixer to be 1.573GHz. Synthesis of this LO has power implications for the radio due to the high frequency involved. The radio's power consumption is mitigated by the architecture developed for the frequency synthesizer, which eliminates a power-consuming block used in conventional architectures.

Chapter 4

CMOS Mixers

IN any heterodyne radio there must be a mixer in the signal path in order to perform frequency conversion to an IF. An LNA, in comparison, is not a fundamental component of the signal path, but becomes necessary in most modern radios to meet noise-figure performance requirements. In fact, because of the dominance of Armstrong's superheterodyne architecture since its inception in 1918 [16], one is almost guaranteed to find a mixer in any high-performance communications receiver. So, clearly, mixer circuits are of interest to radio engineers. Thus, from discrete to integrated radios, people have investigated a variety of mixers. Naturally, when a paradigm shift occurs, innovation in circuit architectures tends to lag behind. Familiar architectures are mimicked in early work, but then the attributes of the new medium are exploited, leading to new implementations.

To begin this chapter, we first discuss how to characterize a mixer. This discussion establishes the basis to make comparisons among different architectures. Then, with the performance metrics defined, we conduct an examination of recently published CMOS mixers. The survey ends with this work's development, the double-balanced voltage mixer, which takes advantage of the similarity between a CMOS transistor and a switch, as well as the IC environment by using a capacitive load. This chapter concludes with a quantitative analysis of the voltage mixer's conversion gain for different LO drives, and addresses the voltage mixer's noise figure and linearity.

4.1 Performance Metrics

In this section, we enumerate and define specifications to assess a mixer's quality. This list has the following specifications: power consumption, conversion gain, noise figure, and linearity.

4.1.1 Power Consumption

Power is explicitly listed to emphasize that it should be kept *as small as possible.* We will consider the mixer's power consumption to include only the power used inside the mixer. The reason for this choice is simple: the reported power consumption for published mixers accounts only for the power dissipated in the mixer itself. Ideally, for a more accurate comparison, the mixer's power consumption would also include the portion of the total LO power spent in driving the mixer's LO port. The power used in the circuits that generate the LO signal is not included, because the generation of this signal can be accomplished in a number of ways, and therefore the accounting should keep it separate.

4.1.2 Conversion Gain

There are two types of conversion gain that can be reported: power conversion gain and voltage conversion gain. Power conversion gain may be used if the mixer's IF port is terminated in a real impedance. In that case, the power conversion gain is defined to be the power delivered to the load divided by the power available from the source:

$$G_P = \frac{power\ delivered\ to\ the\ load}{power\ available\ from\ the\ source}. \tag{4.1}$$

If the IF termination is not a resistance, then one is forced to use voltage conversion gain, because the average power delivered to the reactive load is zero. It is defined as the voltage amplitude across the IF port divided by the source voltage amplitude[1]:

$$G_V = \frac{\textit{voltage amplitude across the IF port}}{\textit{source voltage amplitude}}. \tag{4.2}$$

The voltage conversion gain can be found from the power conversion gain, in the case where the IF termination is resistive, through the following exercise. Let A_{RF} be the amplitude of a sinusoidal RF source and let A_{IF} be the signal amplitude across the load. Furthermore, let R_S be the source resistance and let R_L be the resistance at the IF port. Then,

$$G_P(dB) = 10\log_{10}\frac{4A_{IF}^2 R_S}{A_{RF}^2 R_L} \tag{4.3}$$

and

$$G_V(dB) = 20\log_{10}\frac{A_{IF}}{A_{RF}}, \tag{4.4}$$

by using the definitions of (4.1) and (4.2). Manipulating (4.3) allows us to express the power conversion gain in terms of the voltage conversion gain:

$$\begin{aligned} G_P(dB) &= 6 + 10\log_{10}\frac{A_{IF}^2}{A_{RF}^2} + 10\log_{10}\frac{R_S}{R_L} \\ &= 6 + G_V(dB) + 10\log_{10}\frac{R_S}{R_L}. \end{aligned} \tag{4.5}$$

Separating the voltage conversion gain from the other terms yields

$$G_V(dB) = G_P(dB) - 6 - 10\log_{10}\frac{R_S}{R_L}. \tag{4.6}$$

Large conversion gain is desirable to minimize the noise impact of subsequent stages, since a larger conversion gain means that the signal entering the subsequent

[1] Open circuit.

stages is stronger compared to the noise levels generated in these later blocks. However, excessive gain may not be desirable from a linearity viewpoint.

4.1.3 Noise Figure

A block's noise figure indicates its sensitivity, or the input signal power that makes the output SNR equal to its minimum acceptable value, SNR_{min}. A block can be an amplifier, mixer, or filter as in the partitioning of the signal path shown in Figure 3.4, or it can be increased in scope and taken as the entire signal path. To see that sensitivity can indeed be found from noise figure, we start with a definition of noise figure:

$$NF(dB) = SNR_i(dB) - SNR_o(dB), \tag{4.7}$$

where the i subscript denotes input and the o denotes output. Setting $SNR_o = SNR_{min}$ gives

$$NF(dB) = S_i(dBm) - N_i(dBm) - SNR_{min}. \tag{4.8}$$

where S_i in (4.8) is the input signal power that causes $SNR_o = SNR_{min}$, or the block's sensitivity, S, and N_i is the input noise power. Solving for sensitivity results in

$$S(dBm) = N_i(dBm) + SNR_{min} + NF(dB). \tag{4.9}$$

From (4.7) it is also clear that noise figure describes the amount of noise introduced by a block.

Quantifying and accounting for a mixer's noise contribution, as opposed to an amplifier or filter, is more complicated, because its output response includes noise from multiple input frequencies [17]. Typically the signal of interest is contained in a single sideband, so we are usually interested in the single sideband (SSB) noise figure. Furthermore, for most superheterodyne radios, the radio design causes the

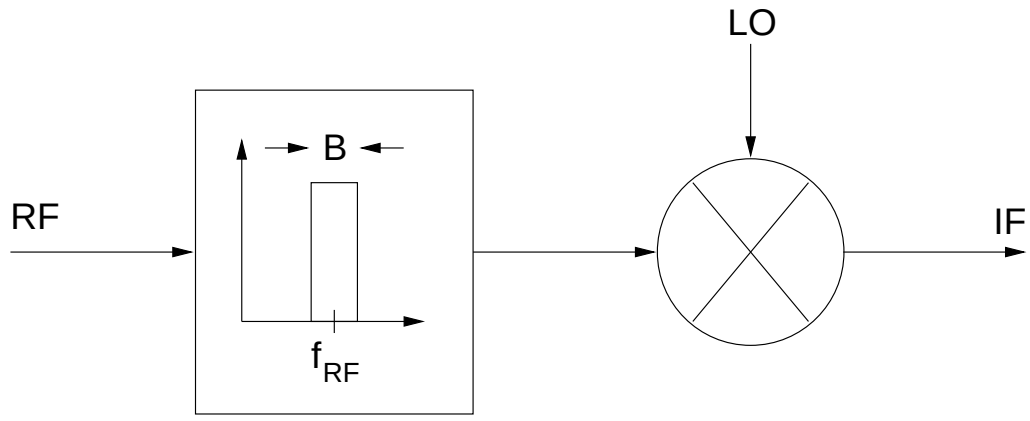


Figure 4.1: Single-response mixer

mixer to function as a single-response transducer. That is, the IF output is mainly due to the RF input. The following mental image for the mixer can be used when the previous two assumptions hold, and is shown in Figure 4.1. The mixer consists of an ideal brickwall filter with bandwidth, B, centered at the radio frequency, f_{RF}, followed by an ideal multiplication whose other input is driven with any periodic function, in general.

At this point, it is informative to work out the SSB noise figure for Figure 4.1's mixer. If we drive the mixer by a voltage source with source resistance, R_S, the signal and noise densities at the source will be as shown in Figure 4.2. The input SNR is calculated over the bandwidth, B, to be

$$SNR_i = \frac{a\alpha B}{4kT_oR_SB} = \frac{a\alpha}{4kT_oR_S}, \tag{4.10}$$

where a is a factor between 0 and 1 representing the deviation from the maximum possible SNR in the bandwidth B, α is the peak signal density, k is Boltzmann's constant, and T_o is 290K. Employing the mixer's voltage conversion gain, G_V, allows us to infer the signal and noise densities at the IF, f_{IF}, due to the input, as shown in Figure 4.3. Additionally, the mixer contributes to the output noise density. The output SNR is calculated over the same bandwidth, B, but now at the IF:

$$SNR_o = \frac{G_V^2a\alpha B}{(G_V^2 4kT_oR_S + \beta)B} = \frac{G_V^2a\alpha}{G_V^2 4kT_oR_S + \beta}, \tag{4.11}$$

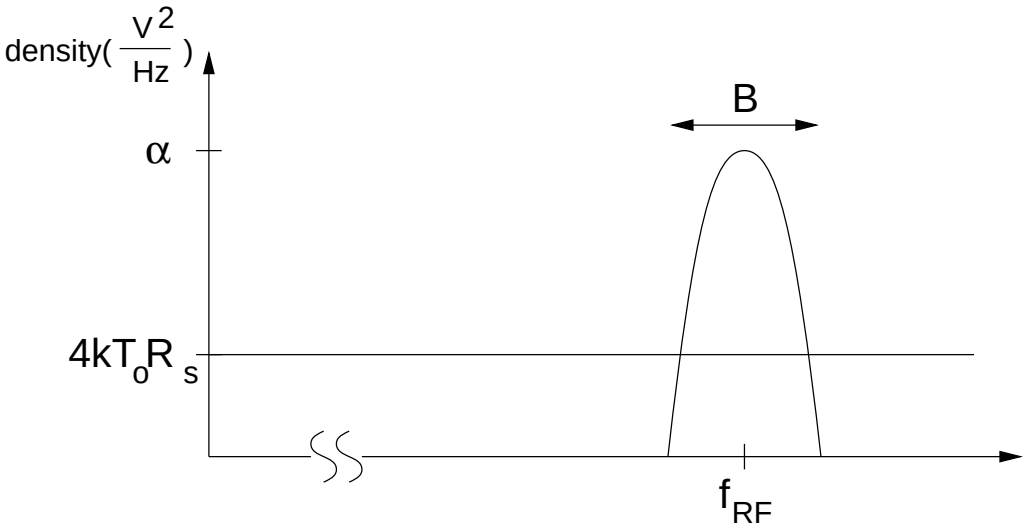


Figure 4.2: RF signal and noise densities

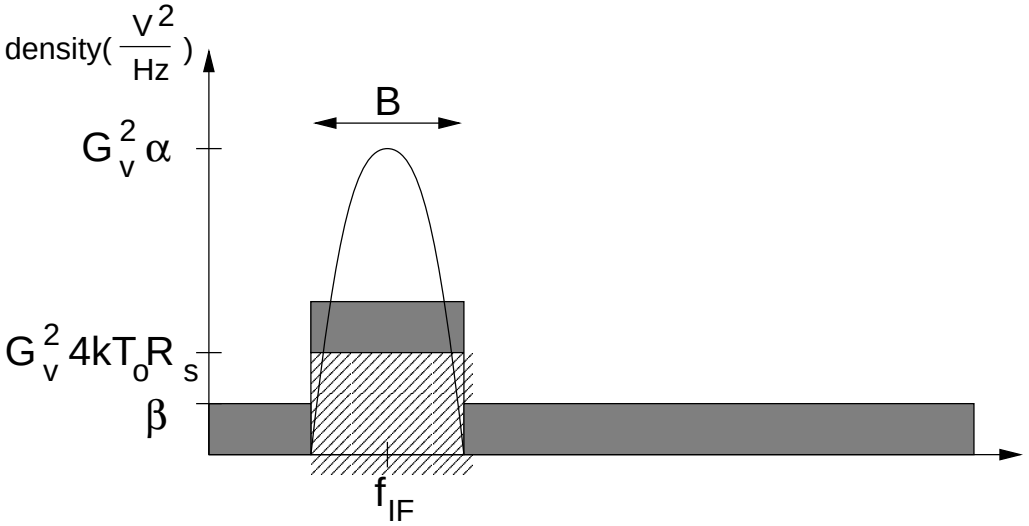


Figure 4.3: IF signal and noise densities

where β is the noise density introduced by the mixer referred to its output. Note that Figure 4.1's ideal brickwall filter bandlimits the input noise to the signal bandwidth. Finally, the ratio of SNR_i to SNR_o is the SSB noise figure for Figure 4.1's mixer:

$$F_{SSB} = \frac{a\alpha}{4kT_oR_S}\frac{G_V^2 4kT_oR_S + \beta}{G_V^2 a\alpha} = 1 + \frac{\beta}{G_V^2 4kT_oR_S}. \tag{4.12}$$

Now that we have an analytic expression for F_{SSB}, we address how to measure F_{SSB} with a noise figure meter. To answer this question, it is first necessary to explain how such a meter operates. A noise figure meter makes two measurements, one with its noise source off and one with its noise source on [18]. The ratio of the source power density in those two measurements is quantified by the excess noise ratio (ENR), which is generally a function of frequency. By collecting the received noise power from the two cases, combined with knowledge of the ENR, the noise figure meter can determine the device under test's (DUT's) noise figure.

The meter's received noise power when the source is off is

$$N_o(off) = FkT_oB\sum_{r=1}^{R} G_r, \tag{4.13}$$

where the G_rs are the insertion gains from each input frequency that produces a response at the IF output frequency. When the source is on, the meter's received noise power is

$$\begin{aligned} N_o(on) &= N_o(off) + kT_oB\sum_{r=1}^{R}(ENR_r - 1)G_r \\ &= kT_oB\sum_{r=1}^{R} ENR_r G_r + (F-1)kT_oB\sum_{r=1}^{R} G_r \\ &\approx kT_oB ENR_1\sum_{r=1}^{R} G_r + (F-1)kT_oB\sum_{r=1}^{R} G_r, \end{aligned} \tag{4.14}$$

where ENR_1 is the RF's ENR. Since the noise figure meter measures the noise from each input frequency that produces a response at the IF output frequency, the noise

figure meter actually makes a multi-sideband (MSB) noise figure measurement. By taking the ratio of (4.14) to (4.13), and rearranging terms to isolate F gives

$$F_{MSB} = \frac{ENR_1 - 1}{\frac{N_o(on)}{N_o(off)} - 1}. \tag{4.15}$$

For the mixer shown in Figure 4.1, there is only one response ($R = 1$), because of the input bandpass filter. So, the noise figure meter reports F_{SSB} in this case, since $F_{MSB} = F_{SSB}$ here.

4.1.4 Linearity

Linearity is important for elements in the signal path, since nonlinearity can produce distortion products that occupy the same spectrum as that of desired signals. For a mixer in a typical superheterodyne architecture, the nonlinearity of greatest concern is third-order. Let's consider a simple depressed cubic

$$y = ax + bx^3, \tag{4.16}$$

where a is the coefficient of the first-order term and b is the coefficient of the third-order term.

If the input to this nonlinearity is a sinusoid, i.e.,

$$x = A\sin(\omega t), \tag{4.17}$$

where A is the amplitude and ω is the angular frequency, then we can solve for the amplitude at which the output's linear component at ω and nonlinear component at ω are equal. First, we substitute (4.17) into (4.16) which gives

$$y = aA\sin(\omega t) + bA^3\sin^3(\omega t). \tag{4.18}$$

Next, we expand $\sin^3$ to obtain

$$y = aA\sin(\omega t) + bA^3(\frac{3}{4}\sin(\omega t) - \frac{1}{4}\sin(3\omega t)). \tag{4.19}$$

Grouping similar frequencies yields

$$y = (aA + \frac{3}{4}bA^3)\sin(\omega t) - \frac{1}{4}bA^3\sin(3\omega t). \tag{4.20}$$

Ignoring the term at 3ω, we see that the term at ω has a coefficient consisting of two terms: a linear replica of the input and a nonlinear distortion of the input. The magnitudes of these two terms are equal when

$$A = A^* = \sqrt{\frac{4|a|}{3|b|}}. \tag{4.21}$$

When we are given a block, we don't know what a and b are, but we can infer them from a measurement. The linearity measurement performed to determine A^* is called a third-order intermodulation product (IP3) measurement. A test is conducted in which two tones, closely spaced in frequency, are applied to the DUT. The tones have equal amplitudes so that

$$x = A\sin(\omega_1 t) + A\sin(\omega_2 t), \tag{4.22}$$

where A, the amplitude used in testing, is kept low for a reason that is explained later. The output contains many cross-products, but in particular, there are two cross-products that lie on either side of the primary output tones at ω_1 and ω_2. The spacing between neighboring tones in this group of four is uniform, as shown in Figure 4.4[2]. By measuring A_1 and A_3 on a spectrum analyzer, IP3 referred to the input can be calculated as

$$IIP3(dBm) = \frac{A_1(dBV) - A_3(dBV)}{2} + A(dBm), \tag{4.23}$$

where we have assumed that the DUT is driven from signal generators calibrated in available power. The additional "I" preceding IP3 indicates that the measurement is input referred. Equation (4.23) follows from the construction shown in Figure 4.5.

[2]If the IP3 measurement is performed on a mixer, the output products look the same as in Figure 4.4, but they are converted to the IF ($\omega_1 - \omega_{LO}$ and $\omega_2 - \omega_{LO}$ for a downconversion mixer or $\omega_1 + \omega_{LO}$ and $\omega_2 + \omega_{LO}$ for an upconversion mixer).

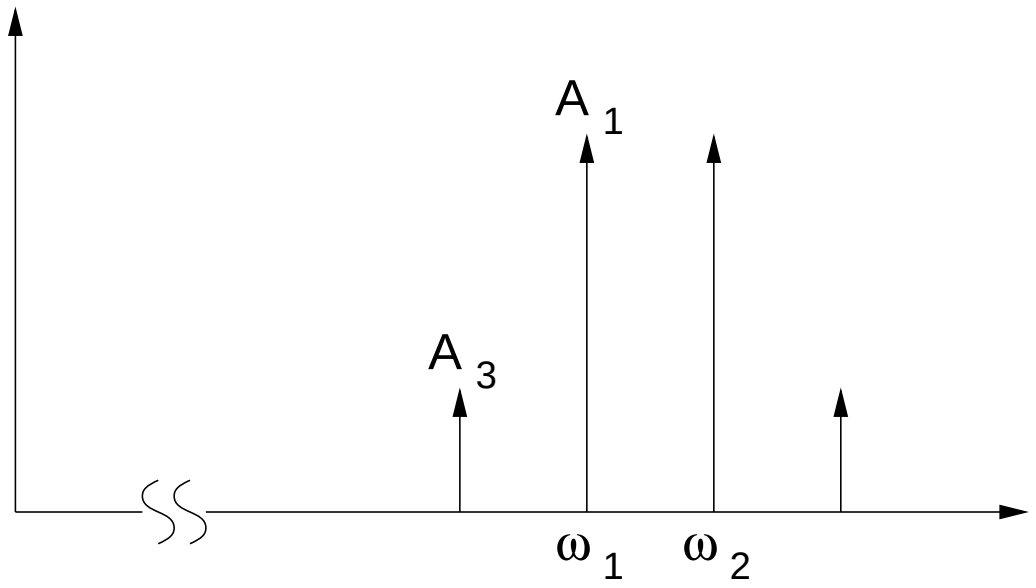


Figure 4.4: Spectrum analyzer display during an IP3 measurement

In this figure, a line with a slope of one has been drawn extending from the measurement point of A_1, and a line with a slope of three has been drawn extending from the measurement point of A_3. Again, the specific slope values will be explained shortly. The point of intersection corresponds to an equality of amplitudes of the extrapolated first-order and third-order products. As we will show next, expressing the corresponding input source power, $IIP3(dBm)$, as an amplitude gives A^*, that is,

$$10^{\frac{IIP3(dBm)-10}{20}} = \sqrt{\frac{4|a|}{3|b|}} = A^*. \tag{4.24}$$

In order to show that the point of intersection indeed corresponds to input amplitudes of A^*, we substitute (4.22) into (4.16) to yield

$$\begin{aligned} y = aA\sin(\omega_1 t) + aA\sin(\omega_2 t) + bA^3\sin^3(\omega_1 t) + 3bA^3\sin^2(\omega_1 t)\sin(\omega_2 t) + \\ 3bA^3\sin(\omega_1 t)\sin^2(\omega_2 t) + bA^3\sin^3(\omega_2 t). \end{aligned} \tag{4.25}$$

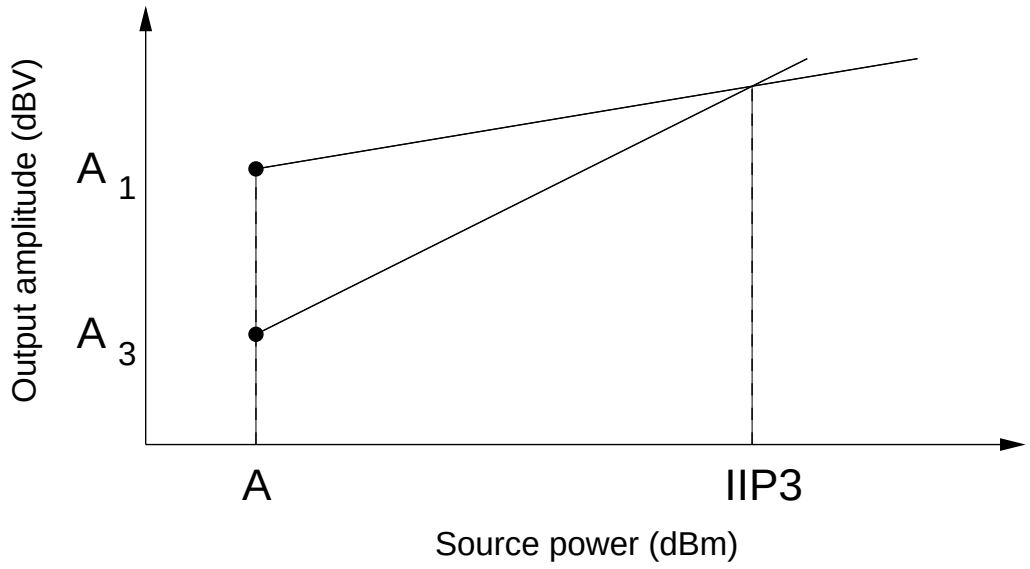


Figure 4.5: IIP3 construction from measured data

Next, expand $\sin^3$ and $\sin^2$ to get

$$\begin{aligned} y =& aA\sin(\omega_1 t) + aA\sin(\omega_2 t) + \\ & \frac{3}{4}bA^3\sin(\omega_1 t) - \frac{1}{4}bA^3\sin(3\omega_1 t) + \\ & \frac{3}{2}bA^3\sin(\omega_2 t) - \frac{3}{4}bA^3\sin((2\omega_1+\omega_2)t) + \frac{3}{4}bA^3\sin((2\omega_1-\omega_2)t) + \\ & \frac{3}{2}bA^3\sin(\omega_1 t) - \frac{3}{4}bA^3\sin((2\omega_2+\omega_1)t) + \frac{3}{4}bA^3\sin((2\omega_2-\omega_1)t) + \\ & \frac{3}{4}bA^3\sin(\omega_2 t) - \frac{1}{4}bA^3\sin(3\omega_2 t). \end{aligned} \tag{4.26}$$

Grouping similar frequencies yields

$$\begin{aligned} y =& \frac{3}{4}bA^3\sin((2\omega_1-\omega_2)t) + (aA+\frac{9}{4}bA^3)\sin(\omega_1 t) + \\ & (aA+\frac{9}{4}bA^3)\sin(\omega_2 t) + \frac{3}{4}bA^3\sin((2\omega_2-\omega_1)t) - \\ & \frac{1}{4}bA^3\sin(3\omega_1 t) - \frac{3}{4}bA^3\sin((2\omega_1+\omega_2)t) - \\ & \frac{3}{4}bA^3\sin((2\omega_2+\omega_1)t) - \frac{1}{4}bA^3\sin(3\omega_2 t). \end{aligned} \tag{4.27}$$

The coefficient of the term at ω_1 is $aA + \frac{9}{4}bA^3$, which can be approximated by aA if the amplitude is kept small. This simplification is one reason why the amplitude used in testing is kept low. The coefficient of the term at $2\omega_1 - \omega_2$ is $\frac{3}{4}bA^3$. Now, on a dB-dB scale where A is represented on the x-axis, terms proportional to A have a slope with magnitude one and terms proportional to A^3 have a slope with magnitude three. This axis selection is why the specific slope values for extrapolation were used. Since we are measuring aA and $\frac{3}{4}bA^3$, we can plot these quantities for all A. This plot gives a graphical solution for A^*, the point of intersection, which is expressed in dBm as $IIP3(dBm)$.

4.2 Various CMOS Mixer Architectures

In the last few years, a variety of CMOS mixer topologies have been proposed. Armed with a firm understanding of the different mixer performance metrics, we are in a position to evaluate competing architectures for use in radio systems. A review of the literature suggests the following four mixer architectures: subsampling, potentiometric, Gilbert-type, and voltage-mode.

4.2.1 Subsampling Mixer

The subsampling mixer performs frequency conversion by sampling the RF signal at a rate that satisfies the Nyquist rate, but allows a higher harmonic to act as the LO. Its circuit implementation is shown in Figure 4.6.

Table 4.1 provides a representative set of numbers for the subsampling mixer [19]. Notice that it consumes a large amount of power, and additionally it has a high noise figure. Analysis shows that even though the frequency of the subsampling clock can be very low compared to the RF carrier, the clock edge uncertainty has to be small compared to the RF period [20]. This stringent requirement on the sampling clock translates into more power spent on clock generation. Generation of the low-jitter sampling clock requires an additional 49mW for this particular example, exacerbating the total power consumption. The high noise figure accompanying

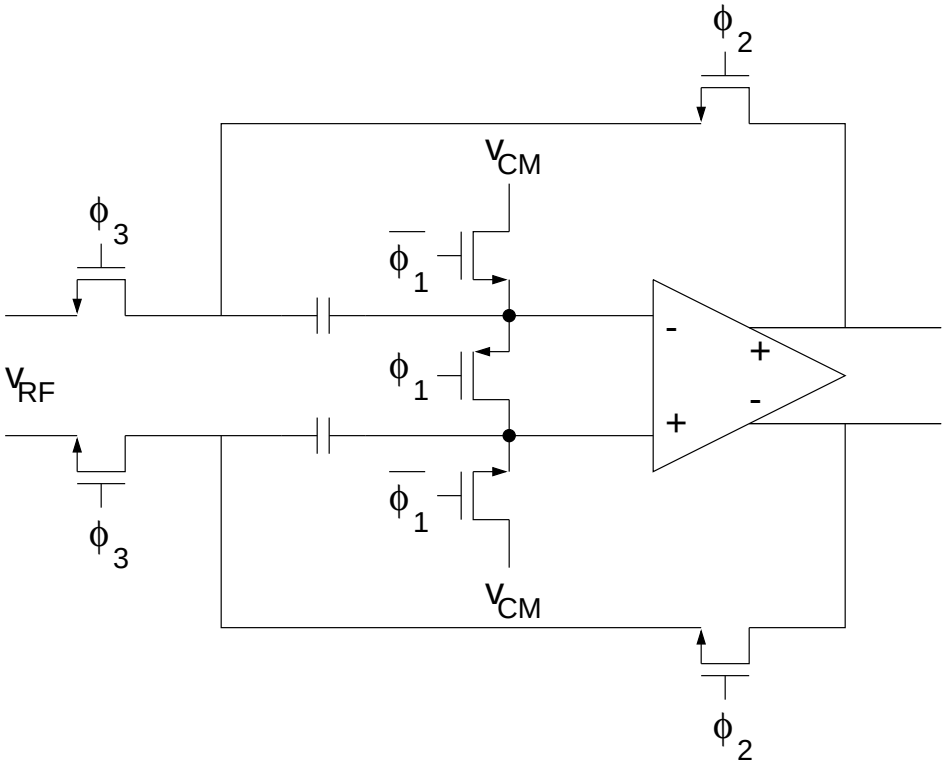


Figure 4.6: Subsampling-mixer circuit diagram

Table 4.1: Subsampling-mixer performance

Power *	41mW
Voltage Conversion Gain	36dB
SSB Noise Figure	47dB
IIP3	−16dBm
Technology	0.6μm BiCMOS
Die Area †	0.88mm^2

*3.3V supply.
†Not including the area of pads.

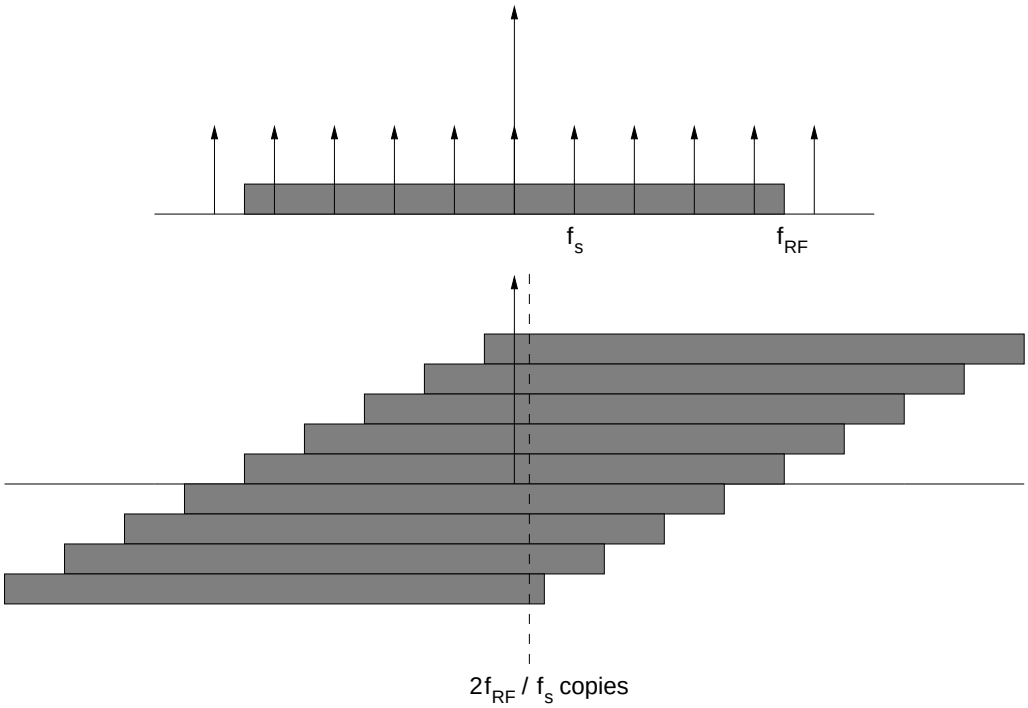


Figure 4.7: Illustration of noise folding

the subsampling mixer is a consequence of broadband switch noise being multiplied by twice the subsampling factor, as shown in Figure 4.7. Subsampling in the time domain is equivalent to convolution in the frequency domain by an impulse train [21], and therefore broadband noise is shifted in frequency by integer multiples of the subsampling frequency. These shifted copies add at the IF, increasing the spot noise[3]. In this picture, the minimum switch bandwidth is f_{RF}, the subsampling frequency is f_s, making

$$M = \frac{f_{RF}}{f_s}, \tag{4.28}$$

where M is the subsampling factor. Finally, we notice that $2M$ is the spot noise figure degradation factor.

[3]Spot noise measures the noise power in a 1 Hz bandwidth

4.2.2 Potentiometric Mixer

The potentiometric mixer, shown in Figure 4.8, produces a current that is proportional to the product of the RF and LO voltages. This current is then converted to a voltage by an op-amp feedback structure.

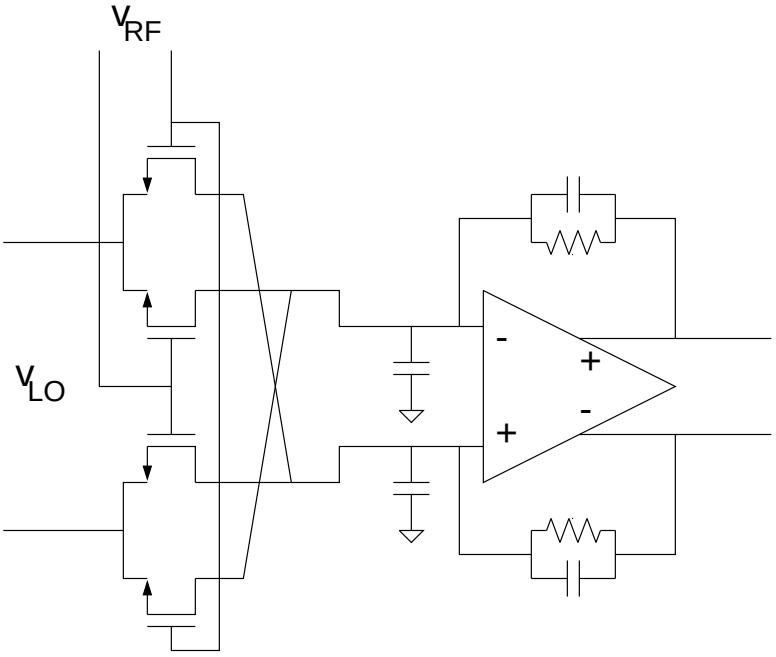


Figure 4.8: Potentiometric-mixer circuit diagram

The numbers for the potentiometric mixer, listed in Table 4.2, show an improvement in power consumption, but the noise figure is still too large [22]. The high noise figure is a consequence of using the weak RF input to modulate the transistor quad's channel conductance via the gates. This choice of connection does not allow the switch resistance to be very small, leading to large noise contributions from the transistor quad. Also, the authors of [22] attribute a substantial portion of the noise to the op-amp in the design. Thus, even though this architecture exhibits very good linearity, that linearity would have to be compromised to achieve a tolerable noise figure. For instance, if the potentiometric mixer were preceded by an LNA to improve noise figure (this LNA is now part of the mixer and is in addition to any LNA prior to the mixer), then there is an increase in power due to LNA power

Table 4.2: Potentiometric-mixer performance

Power *	1.3mW
Voltage Conversion Gain †	18dB
SSB Noise Figure	32dB
IIP3	45.2dBm
Technology	1.2μm CMOS
Die Area ‡	0.42mm^2

*5V supply.
†12dBm LO.
‡Not including the area of pads.

consumption, a decrease in linearity which now is LNA limited, and an increase in area because of the additional LNA.

4.2.3 Gilbert-Type Mixer

The Gilbert-type mixer performs a voltage-to-current conversion of the RF signal that can then be commutated by the LO, usually to a resistive load to produce a voltage at the output. The topology of this mixer is illustrated in Figure 4.9.

Table 4.3 contains results from the measurement of a CMOS Gilbert-type mixer [23]. The main advantage of this architecture over the voltage mixer is in conversion gain. Conversion gain can be helpful if the later stages have large noise figures and if it is difficult to obtain additional gain from the LNA block. Because the noise figure and linearity are limited by the transconductance stage, elimination of the voltage-to-current conversion can yield better performance, as evidenced in the voltage mixer.

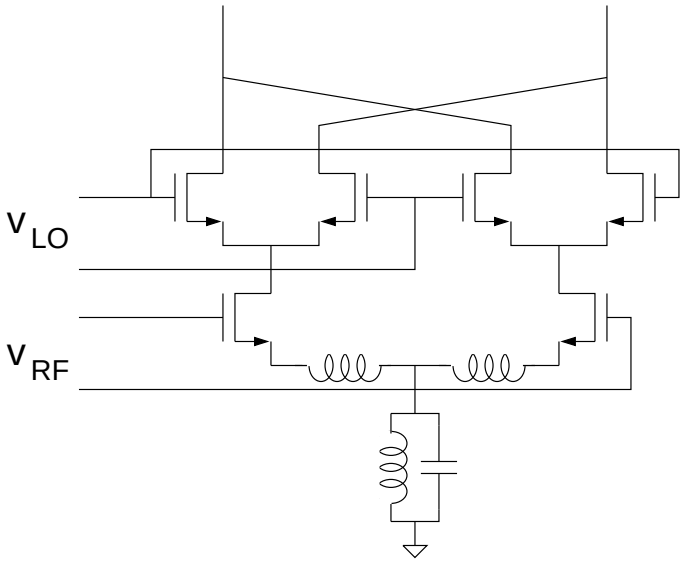


Figure 4.9: Gilbert-type mixer circuit diagram

Table 4.3: Gilbert-type-mixer performance

Power *	7mW
Power Conversion Gain	8.8dB
SSB Noise Figure	9.2dB
IIP3	−4.1dBm
Technology	0.5μm CMOS
Die Area †	$0.14mm^2$

*2.7V supply.
†Not including the area of pads.

4.2.4 Voltage Mixer

As just mentioned, the voltage mixer skips the voltage-to-current conversion of the RF signal, keeping everything in the voltage domain. This architecture, pictured in Figure 4.10, naturally accommodates a capacitive load, which is the dominant type of load impedance on-chip. A further benefit of a capacitive load is that it contributes no noise.

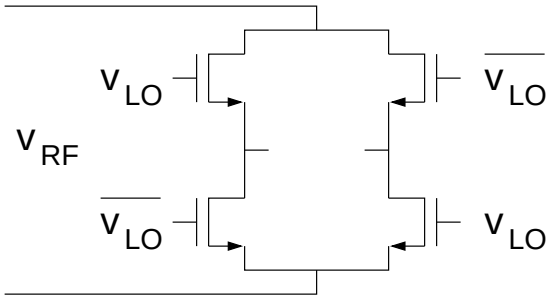


Figure 4.10: Voltage-mixer circuit diagram

It's clear from Table 4.4 that the main drawback with the voltage mixer is its low conversion gain, which is a disadvantage for the radio's noise figure but advantageous for the radio's linearity. However, in the context of the radio's receive path, where an LNA with 20-30dB of gain precedes the mixer, the voltage mixer's conversion loss is acceptable. In the GPS radio, the voltage mixer's approximate contribution to the radio's noise figure is a negligible 0.1dB. Subsequent stages contribute an additional 0.4dB.

A final note on the voltage mixer is that it benefits the most from technology scaling of the four CMOS mixer architectures presented. This mixer scales well due to the architecture's dependence on CMOS transistors behaving as switches. Furthermore, it can tolerate the accompanying supply-voltage reduction more gracefully than the other architectures, since it has no stacked transistors between the supply and ground and does not require large gate biases.

Table 4.4: Voltage-mixer performance

Power	$< 500\mu$W
Voltage Conversion Gain	-7.3dB
SSB Noise Figure	6dB
IIP3	4dBm
Technology	0.5μm CMOS
Die Area *	$0.0084mm^2$

*Not including the area of pads.

4.3 CMOS Voltage Mixer

We now proceed with a more detailed look at the double-balanced CMOS voltage mixer. Choosing the LO as the gate drive and the RF to be commutated by the switches is the natural selection, since the LO amplitude is much larger than the RF amplitude. This choice explains the voltage labels in Figure 4.10, but we must still explain the configuration and function of the figure's four transistors.

The mixer's four transistors are labeled M_1–M_4 as shown in the bottom of Figure 4.11. These four transistors are grouped together into two pairs. One pair is shown at the top left of Figure 4.11, and the other is shown at the top right of Figure 4.11. Transistors M_1 and M_4 work together and are controlled by the LO signal, while transistors M_2 and M_3 form a unit controlled by the LO's inverse. This drive selection is explained shortly. Each pair serves the function of connecting the mixer's IF port to the RF port. The difference between the two pairs is the polarity with which they perform this connection. M_1 and M_4 connect the load with a positive polarity for half a cycle and disconnect the load for the other half, but M_2 and M_3 connect the load with a negative polarity for half a cycle, disconnecting the load for the other half. Thus, by operating the two LOs with a 180 degree phase shift, the two individual circuits at the top of Figure 4.11 can be merged into the circuit shown at the bottom, achieving a double-balanced architecture.

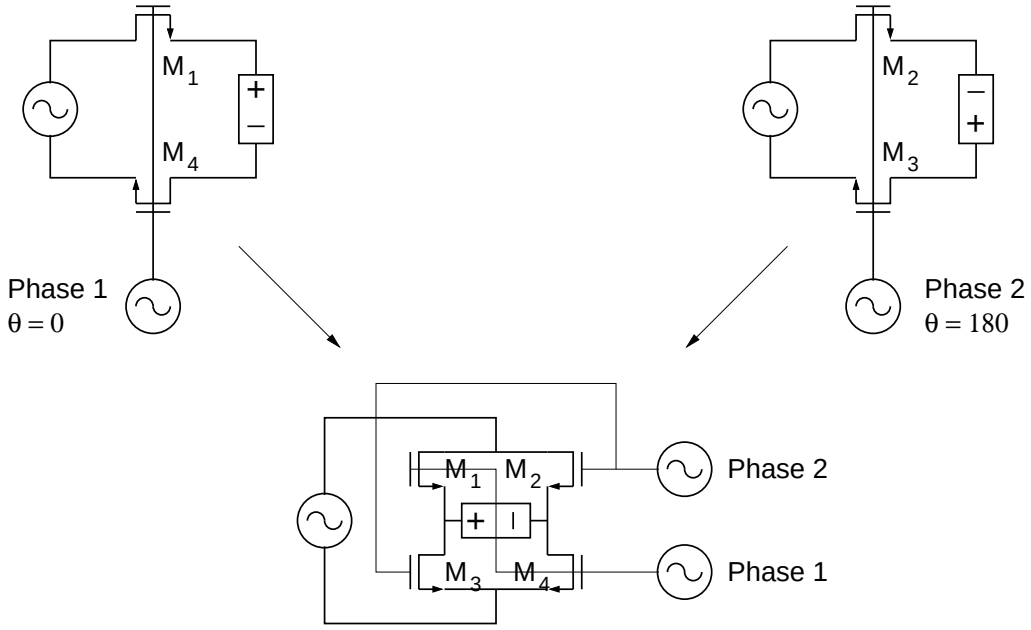


Figure 4.11: Double-balanced CMOS voltage mixer

To further our understanding of the preceding description, it is helpful to examine typical waveforms during operation. One particular set of time domain waveforms for this mixer's operation appear on the left of Figure 4.12, with their respective frequency domain pictures on the right. When the mixing function is 1 (i.e., M_1 and M_4 are on), the output is a replica of the input, but when the mixing function is -1 (i.e., M_2 and M_3 are on), the output is an inverted replica of the input. So, in the time domain, we see that the output is the product of the input and the mixing function. To see that this multiplication does indeed produce downconversion, it is instructive to look at what occurs in the frequency domain. The Fourier transform of the mixing function has spectral energy at odd harmonics of the LO. The Fourier transform of the input signal gets convolved with the Fourier transform of the mixing function to produce the output's Fourier transform. The convolution produces a copy of the RF signal at the IF, which can be filtered out.

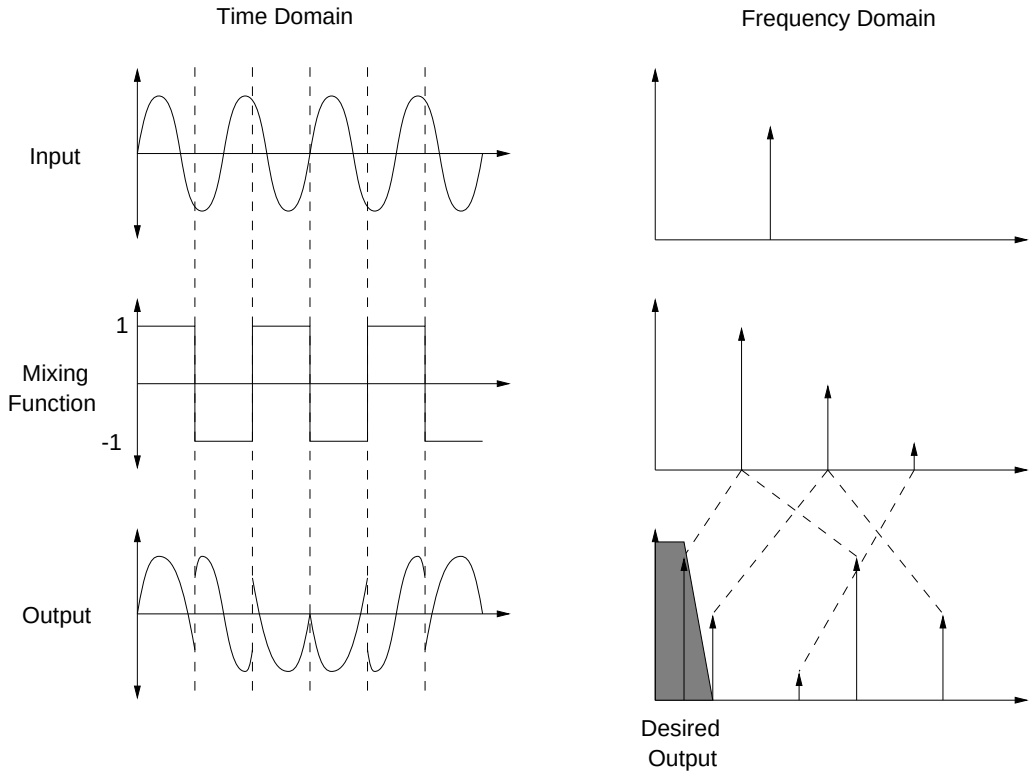


Figure 4.12: Principle of operation

Returning to the four core transistors, we see that M_1–M_4 present a primarily capacitive load to the circuits that generate the LO. M_1–M_4 are identically sized, and furthermore are all minimum length devices. These conditions cause each pair's gate capacitance, C, to be proportional to M_1–M_4's width, W,

$$C \propto W. \tag{4.29}$$

Since the mixer's LO port is primarily a capacitive load, it is natural to absorb this capacitance into the tuned circuit at the output of a narrowband amplifier, thereby reducing the LO power consumed in driving the mixer. This idea is illustrated in Figure 4.13. As a consequence, the LO drive will be nearly sinusoidal, in practice, as opposed to square in appearance.

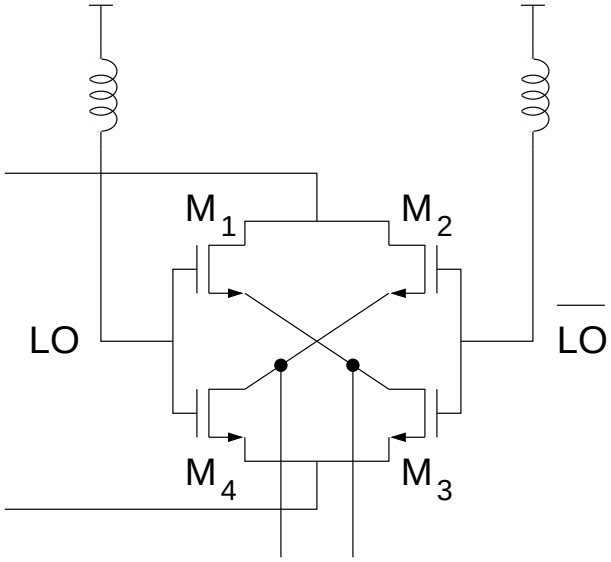


Figure 4.13: Resonating the gates

4.3.1 Voltage Conversion Gain

Since the LO drive will generally be sinusoidal, a study of various sinusoidal drives and their effect on conversion gain is therefore important. We study four specific, but representative, LO drives, but alternate drives can be investigated with the general technique developed here. This technique to examine conversion gain is based on modeling the mixer's four transistors as time-varying conductances. By doing so, voltage conversion gains for the different LO drives representing various operating conditions can be presented and tabulated.

4.3.1.1 Mixer Loading

The type of source, narrowband or broadband, will affect how the analysis proceeds. A broadband source can be included in the mixer's transformation to a Thévenin equivalent circuit, which is the subject of §4.3.1.3. But for RF downconversion, the source is typically narrowband (e.g., an LNA with a tuned output immediately preceding the mixer), so it must be treated separately since it has different impedances at the radio and intermediate frequencies.

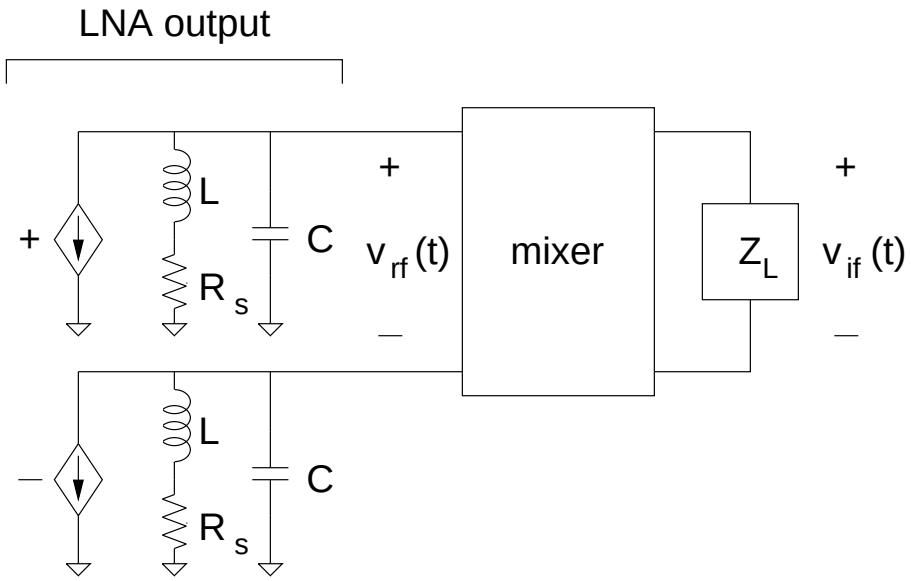
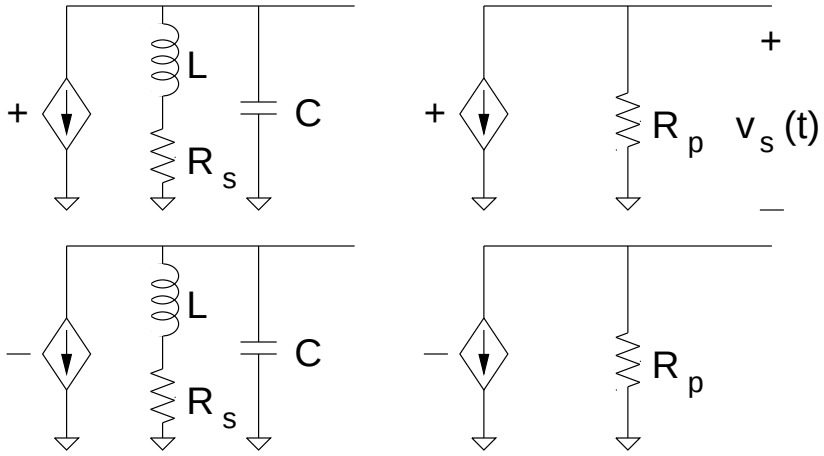


Figure 4.14: Mixer connections



(a) Figure 4.14's LNA circuit model

(b) LNA RF resistance

Figure 4.15: LNA transformation

Figure 4.14 models the LNA output with the mixer. The mixer and its IF termination have an influence on $v_{rf}(t)$, the voltage across the mixer's RF port, at frequencies where the RF source has a non-zero Thévenin impedance. Due to the mixer's time-varying nature, a circuit simulator needs to be used to design the interface between the LNA and mixer and to observe $v_{rf}(t)$. Once this design is completed, the reactive elements of the LNA's tuned output and any reactive loading presented by the mixer are of secondary interest. The LNA's output is therefore transformed at the RF to an equivalent parallel RLC circuit, and then the reactive elements are removed, as shown in Figure 4.15. This transformation allows us to find $v_s(t)$.

One must be aware that $v_s(t)$ can change when the mixer loads the LNA. If the mixer presents only a reactive impedance to the LNA, then $v_s(t)$ and $v_{rf}(t)$ will be identical. However, if there is a real component to the impedance, then the amplitude of $v_{rf}(t)$ is smaller than that of $v_s(t)$.

So, to summarize, the LNA's gain is computed using $v_s(t)$, and thus the mixer's voltage conversion gain will include the gain from $v_s(t)$ to $v_{rf}(t)$, $|A_{rf}|$ (which is less than or equal to one). The other part of the mixer's conversion gain, the conversion component, is analyzed in the remainder of this subsection. This component starts from $v_{rf}(t)$ and ends with $v_{if}(t)$.

4.3.1.2 LO Waveforms

Before proceeding further, we discuss the different ways the mixer will be driven during this conversion gain analysis, specifically the shapes of the LO and $\overline{\text{LO}}$ waveforms. $\overline{\text{LO}}$ is simply a time-shifted version of LO by $T_{LO}/2$, where T_{LO} is the LO period.

The first of four LO waveforms we consider is a square wave with 50% duty cycle and is mathematically expressed as

$$v_{lo}(t) = 2A_{LO}\Pi(2t/T_{LO}) * \sum_{n=-\infty}^{\infty} \delta(t - nT_{LO}) + B_{LO}, \tag{4.30}$$

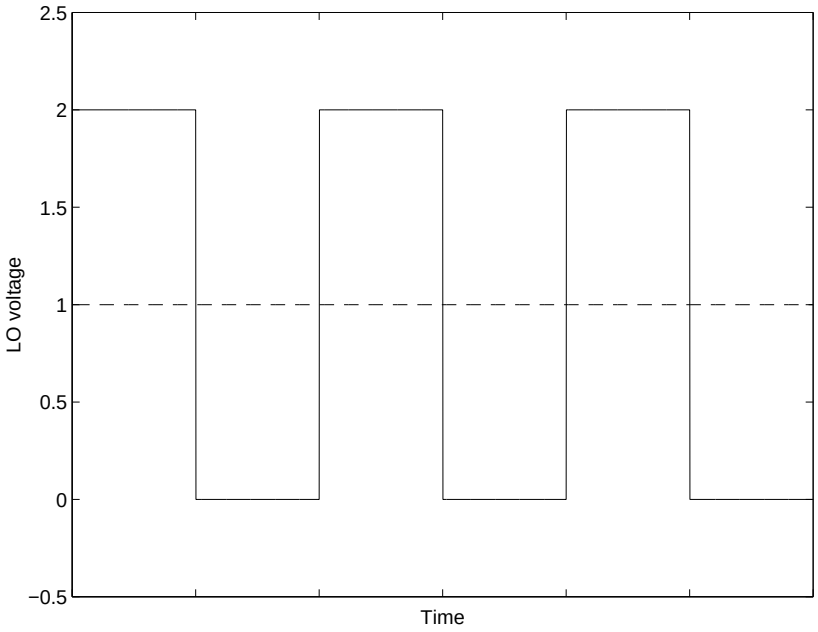
where A_{LO} is half the peak-to-peak voltage level, B_{LO} is the square wave's DC value (equal to the switch threshold voltage), $\Pi(t)$ is the rectangle function, and $\delta(t)$ is the delta function. This LO waveform will be the reference to which other types of LO waveforms will be compared, and is sketched in Figure 4.16(a).

As mentioned previously, a square wave drive is difficult to achieve, especially at high frequencies. A more practical and power efficient drive, presented in Figure 4.13, is achieved by resonating the gate capacitances of transistors M_1–M_4. Tuning the LO results in nearly sinusoidal waveforms which we approximate by

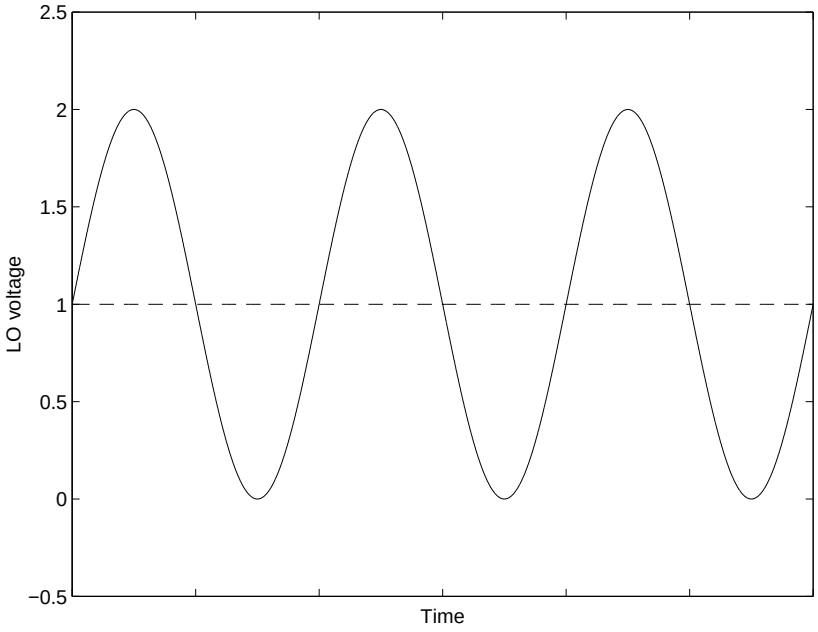
$$v_{lo}(t) = A_{LO}cos(2\pi f_{LO}t + \phi_{LO}) + B_{LO}, \tag{4.31}$$

where A_{LO} is the sinusoid's amplitude, B_{LO} is the DC level on the gates, f_{LO} is the LO frequency, and ϕ_{LO} is the LO phase. This type of waveform is the basis for the remaining three LO drives which are drawn in Figures 4.16(b), 4.17(a), and 4.17(b). In Figure 4.16(b), B_{LO} equals the switch threshold voltage, V_{th}; in 4.17(a), $B_{LO} < V_{th}$ illustrating break-before-make switching action; while in 4.17(b), $B_{LO} > V_{th}$ which is the opposite action, make-before-break.

Alternative drives can be similarly constructed and then used to explore amplitude mismatch between LO and $\overline{\text{LO}}$ and deviations from a perfect 180^o phase relationship between LO and $\overline{\text{LO}}$. However, these drives will require defining both LO and $\overline{\text{LO}}$, because $\overline{\text{LO}}$ is not simply the result of time-shifting LO by $T_{LO}/2$.

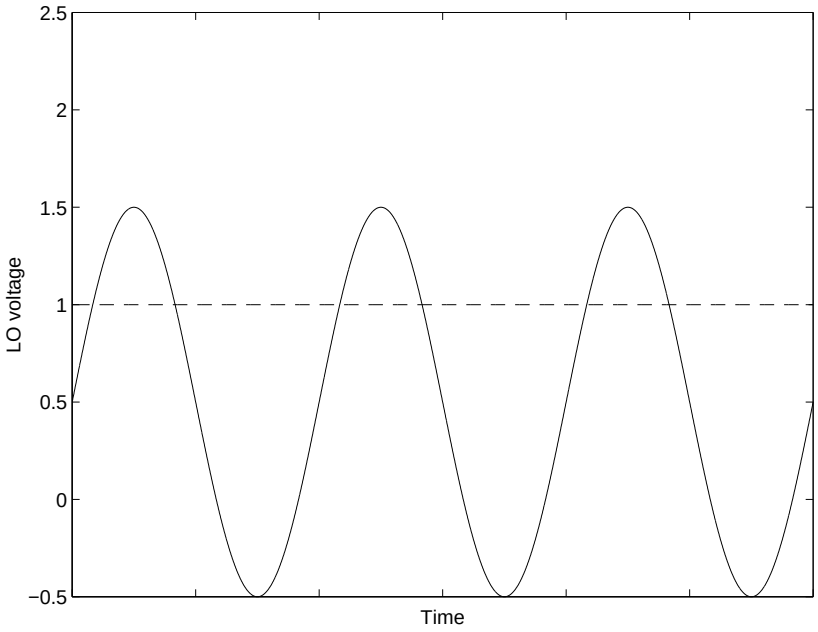


(a) Square drive ($A_{LO} = 1$, $B_{LO} = 1$, $V_{th} = 1$)

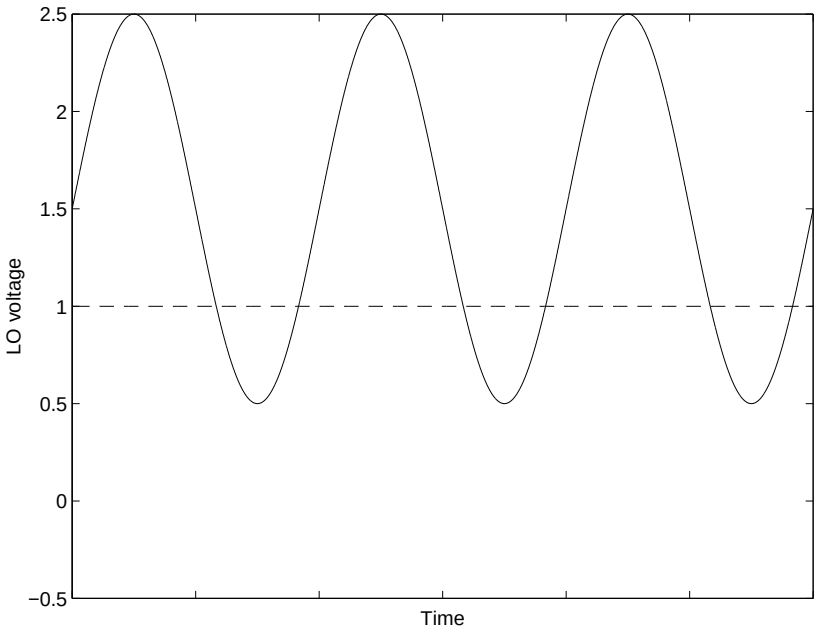


(b) Sinusoidal drive ($A_{LO} = 1$, $B_{LO} = 1$, $V_{th} = 1$)

Figure 4.16: Investigated LO waveforms



(a) Break-before-make drive ($A_{LO} = 1$, $B_{LO} = 0.5$, $V_{th} = 1$)



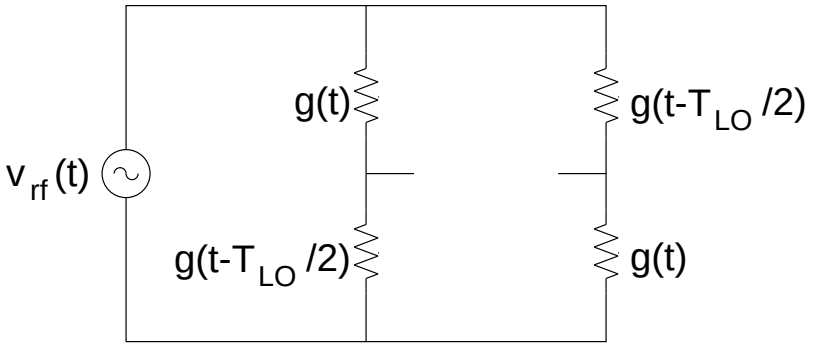
(b) Make-before-break drive ($A_{LO} = 1$, $B_{LO} = 1.5$, $V_{th} = 1$)

Figure 4.17: Investigated LO waveforms

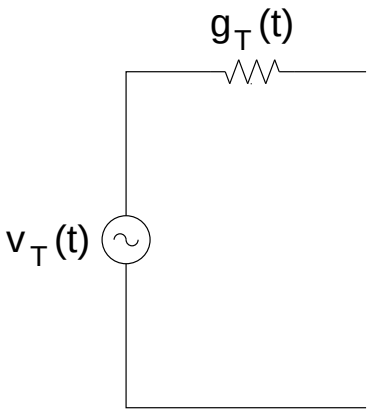
4.3.1.3 Mixer Model and its Thévenin Equivalent

Once the LO waveform has been defined, the mixer's switches can be viewed as time-varying conductances, as shown in Figure 4.18(a). The conductance can be roughly approximated as

$$g(t) \approx \begin{cases} k(v_{lo} - V_{th}) & v_{lo} > V_{th} \\ 0 & v_{lo} \leq V_{th}. \end{cases} \tag{4.32}$$



(a) Time-varying conductances



(b) Thévenin equivalent

Figure 4.18: Mixer model

It is possible to simplify the switch network with a Thévenin equivalent network. This representation is shown in Figure 4.18(b), where the open circuit voltage is given by

$$v_T(t) = \frac{g(t) - g(t - T_{LO}/2)}{g(t) + g(t - T_{LO}/2)} v_{rf}(t) = m(t) v_{rf}(t) \tag{4.33}$$

and the Thévenin impedance, written as a conductance, is

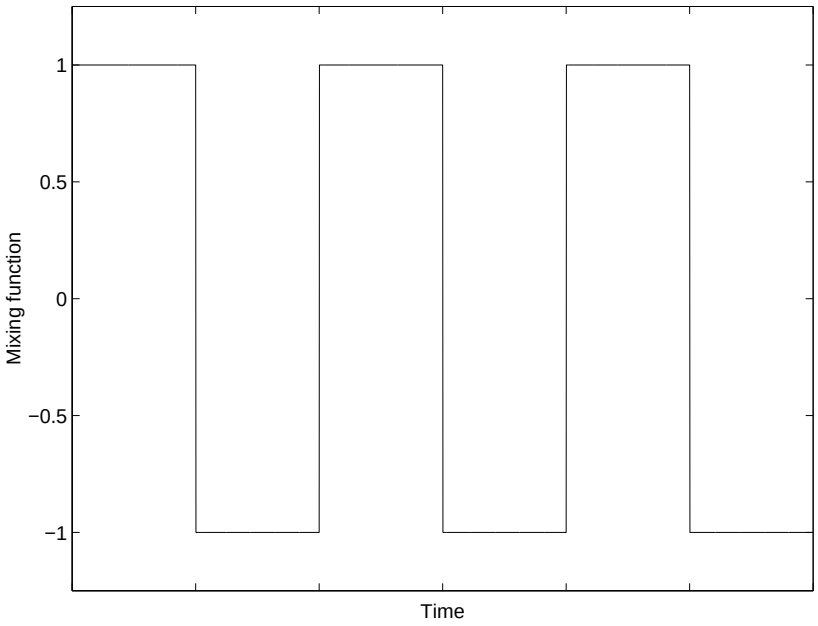
$$g_T(t) = \frac{g(t) + g(t - T_{LO}/2)}{2}. \tag{4.34}$$

The action of mixing, or frequency translation, is implicit in this transformation to a Thévenin equivalent. Frequency conversion is most easily seen in the reference case, when the LO drive is a square wave. The open circuit voltage is a square wave with zero DC value and unit amplitude multiplied by the RF voltage function, and so $v_T(t)$ is a mixed version of $v_{rf}(t)$. Figures 4.19–4.22 illustrate the mixing function

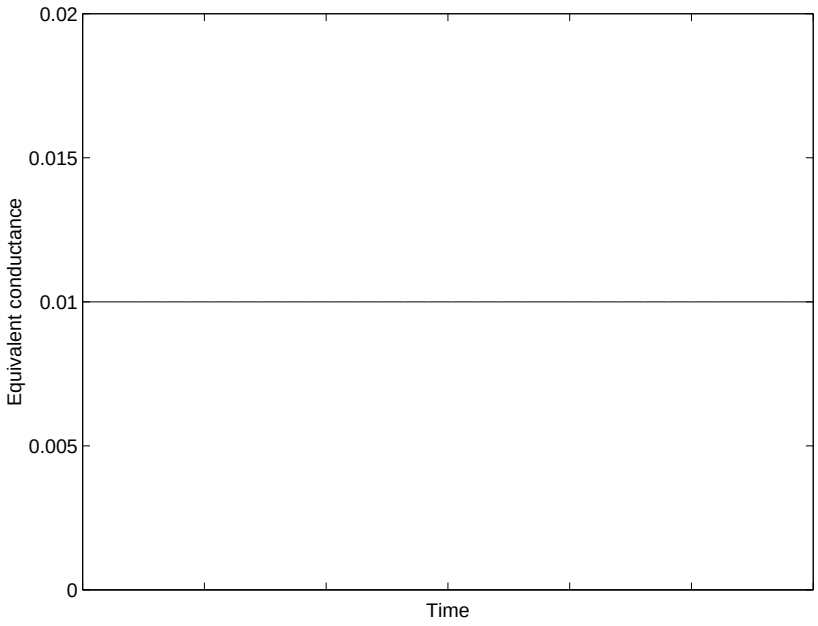
$$m(t) = \frac{g(t) - g(t - T_{LO}/2)}{g(t) + g(t - T_{LO}/2)} \tag{4.35}$$

and $g_T(t)$, for the four cases with $g_{max} = \frac{1}{50\Omega}$. g_{max} is the maximum conductance of each switch.

Both $m(t)$ and $g_T(t)$ exhibit important properties. The mixing function $m(t)$ has no DC component, is periodic with a period of T_{LO}, and has half wave symmetry, implying that it only has odd frequency content (nf_{LO}, where n is an odd integer). The conductance $g_T(t)$ has a DC component and is periodic with a period $T_{LO}/2$. We will see that this DC component together with a load capacitance yields the mixer bandwidth. Furthermore, $g_T(t)$'s particular periodicity foreshadows the importance of the frequency $2\omega_{LO}$ in the analysis of the next section.

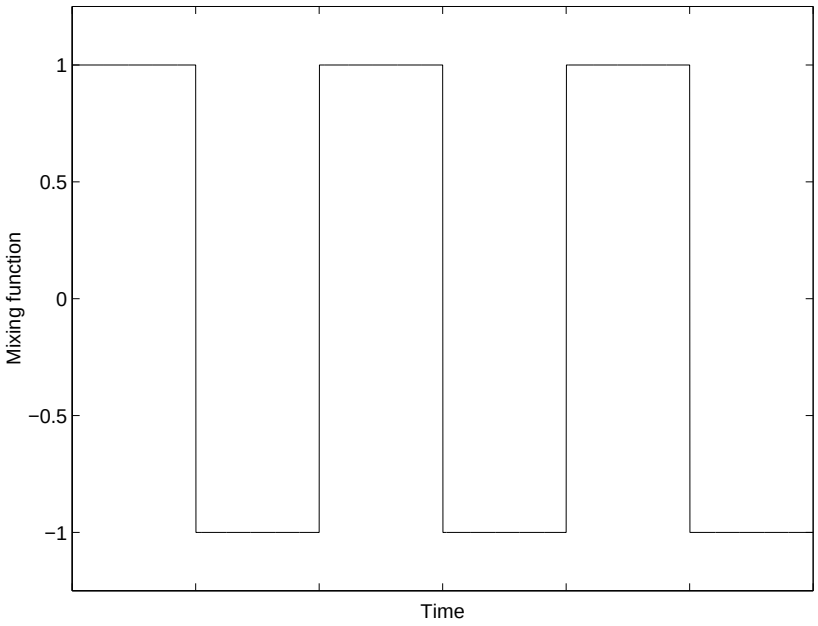


(a) Mixing function

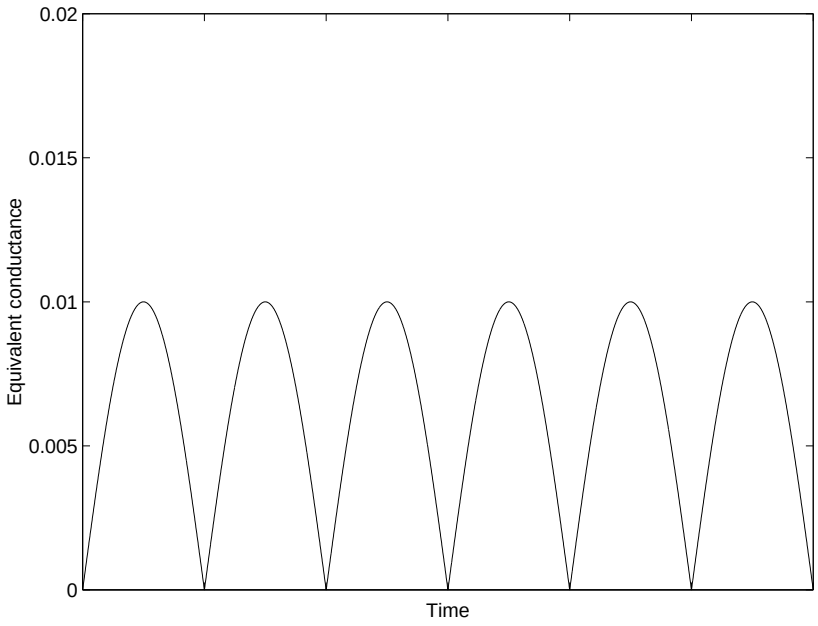


(b) Thévenin conductance for $g_{max} = \frac{1}{50\Omega}$

Figure 4.19: Square drive

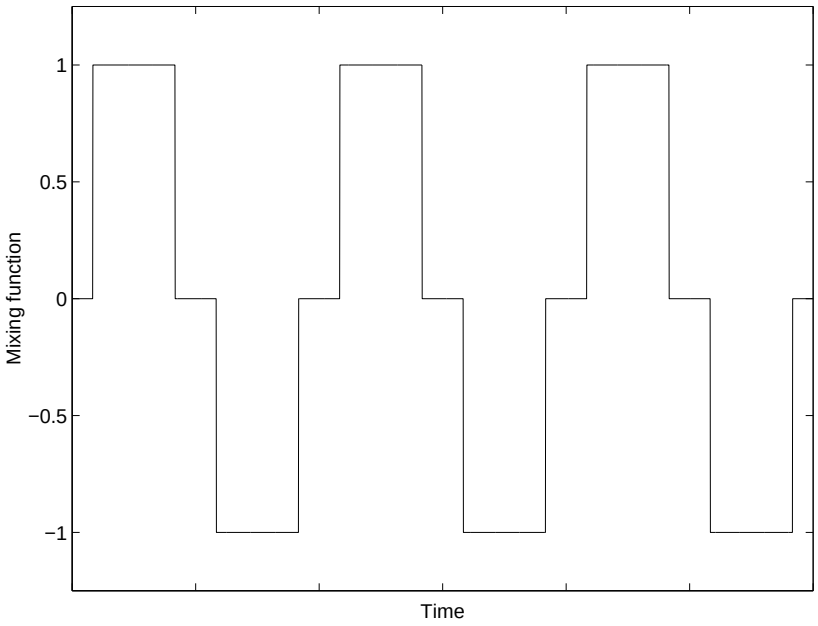


(a) Mixing function

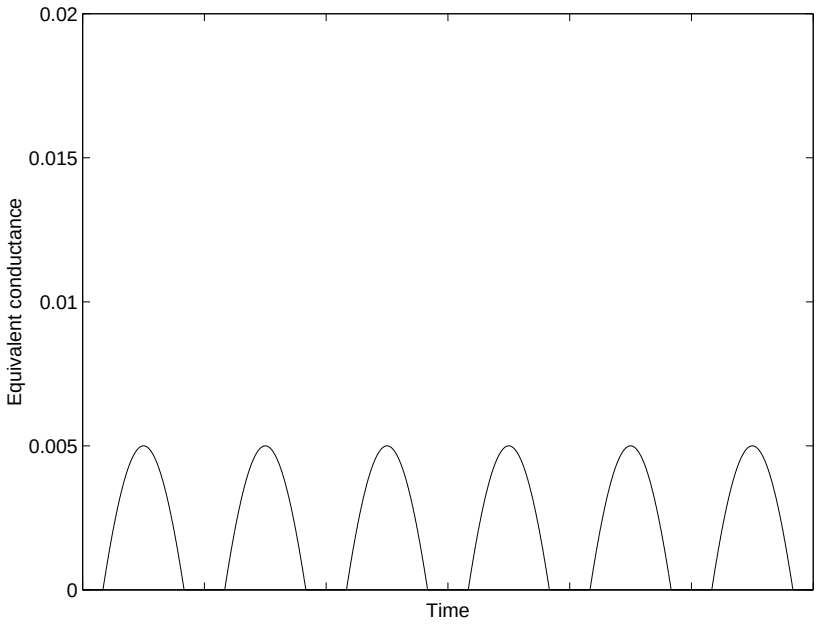


(b) Thévenin conductance for $g_{max} = \frac{1}{50\Omega}$

Figure 4.20: Sinusoidal drive

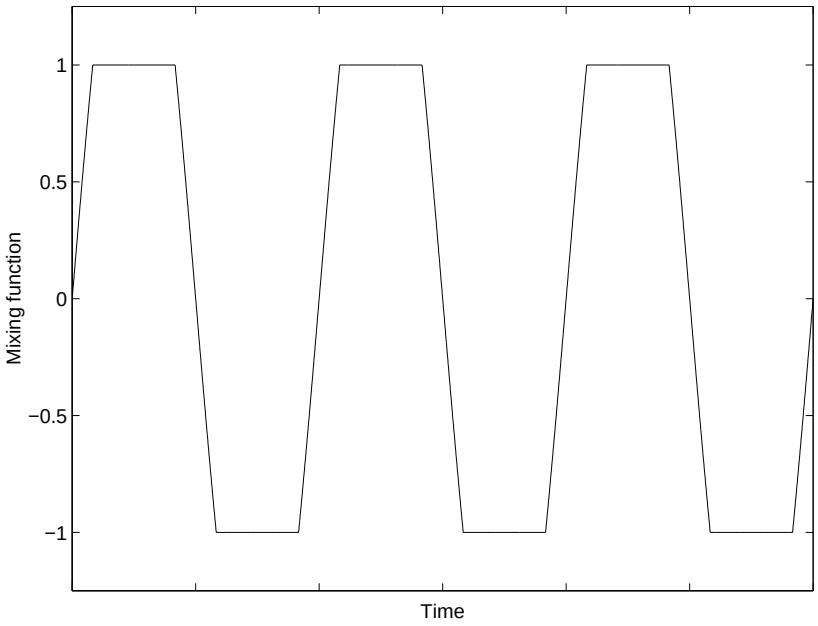


(a) Mixing function

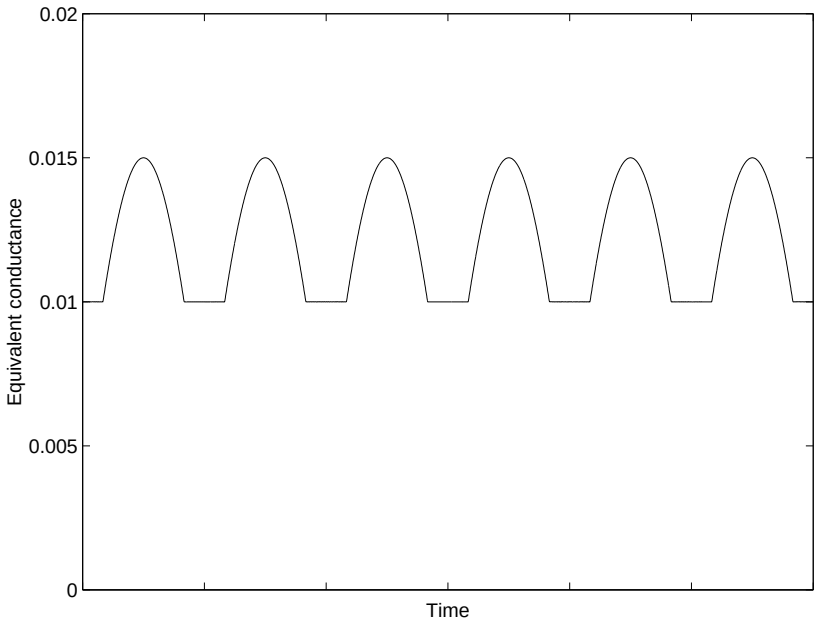


(b) Thévenin conductance for $g_{max} = \frac{1}{50\Omega}$

Figure 4.21: Break-before-make drive



(a) Mixing function



(b) Thévenin conductance for $g_{max} = \frac{1}{50\Omega}$

Figure 4.22: Make-before-break drive

4.3.1.4 Conversion Component

Once we have the Thévenin equivalent representation of the mixer, we can find $v_{if}(t)$ for arbitrary loads. The load that we are interested in is capacitive, as shown in Figure 4.23. Since the load is capacitive, we will compute the mixer's voltage conversion gain.

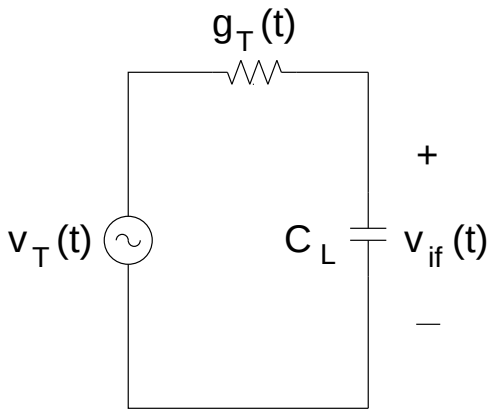


Figure 4.23: Mixer model with capacitive load

We first examine the extreme case where $C_L = 0$, using the Thévenin equivalent circuit developed in §4.3.1.3. If $C_L = 0$, then $v_{if}(t) = v_T(t) = m(t)v_{rf}(t)$[4]. Therefore, to find the voltage conversion gain from the RF port to the IF port, the Fourier transform of the mixing function must be evaluated at f_{LO}, which is shown in Table 4.5. For a square wave drive, the conversion gain is $2/\pi$, a classic result. Interestingly, a sinusoidal drive also has the same voltage conversion gain of $2/\pi$. Because both a square drive and a sinusoidal drive have the same voltage conversion gain, we consider $2/\pi$ (-3.92dB) as a baseline and use it throughout the remainder of this subsection on conversion gain. For the break-before-make and make-before-break drives, the voltage conversion gain is always less than or equal to $2/\pi$. Finally, it is interesting to note that $|M(f_{LO})|$ in the last two cases depends only on a single quantity that characterizes an LO waveform, r. In words, r is the

[4]From §4.3.1.1, $|A_{rf}| = 1$ when $C_L = 0$, since the mixer cannot load the LNA output in this case.

Table 4.5: $|M(f_{LO})|$

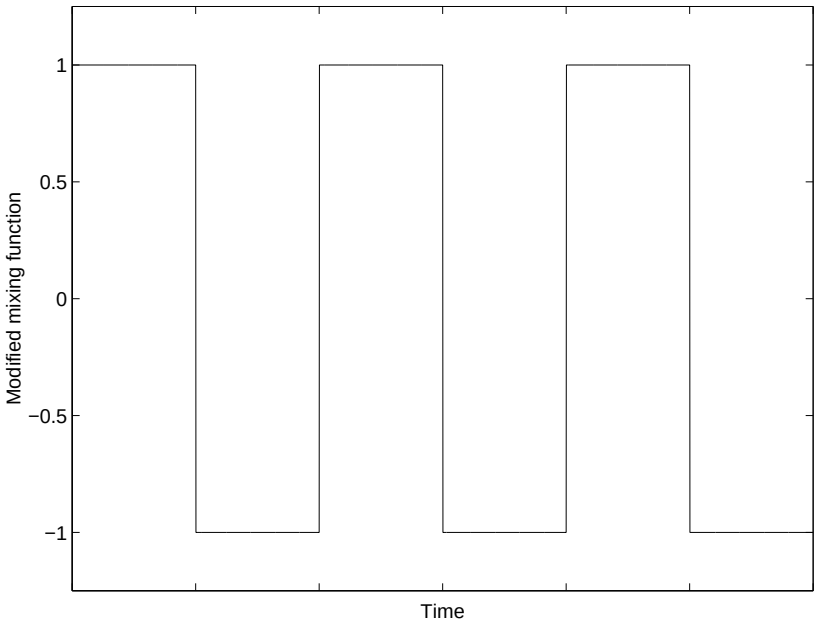
Square drive	$2/\pi$	
Sinusoidal drive	$2/\pi$	
Break-before-make drive *	$(2/\pi)\sqrt{1-r^2}$	$0 \leq r \leq 1$
Make-before-break drive *	$\frac{sin^{-1}(r)/r+\sqrt{1-r^2}}{\pi}$	$0 \leq r \leq 1$
	$1/(2r)$	$1 \leq r < \infty$

*$r = \frac{|V_{th}-B_{LO}|}{A_{LO}}$.

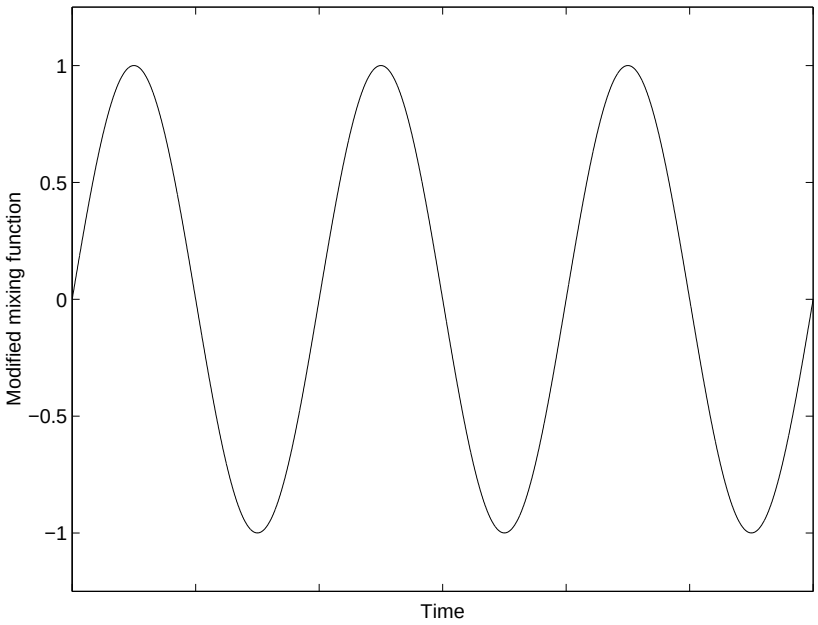
ratio of (the absolute value of) the difference between the switch threshold voltage and the DC level of the LO drive, to the amplitude of the LO drive. The absolute value of the difference is a measure of the drive's deviation from a sinusoidal drive, with a larger magnitude indicating a greater difference. A larger LO amplitude, on the other hand, causes the drive to appear more like a sinusoidal drive by making the difference less significant. So, r is an indication of a waveform's deviation from a sinusoidal drive ($r = 0$).

In general, C_L does not equal zero. But the conversion gain can still be solved through a more lengthy analysis. The superposition integral is used to find $v_{if}(t)$ as a function of $v_{rf}(t)$, after finding the network's impulse response. The detailed derivations are contained in Appendix A, while key results are presented here. The results indicate that under certain (but practically relevant) conditions, a very simple system can be used to analyze the voltage conversion gain. Furthermore, they also predict the somewhat surprising result that it is theoretically possible to achieve a voltage conversion gain of one.

The following discussion applies if $\frac{\overline{g_T}}{2\omega_{LO}C_L} \ll 1$, where $\overline{g_T}$ is the DC value of $g_T(t)$ and ω_{LO} is the angular LO frequency. Recall that $\frac{\overline{g_T}}{2\omega_{LO}C_L}$ is the ratio of the average mixer bandwidth to twice the LO frequency ($g_T(t)$ is periodic with a period $T_{LO}/2$). It is not surprising to find this ratio involved in an important condition, which is met by a bandlimiting mixer. Note that the average mixer bandwidth must be small compared to the LO frequency (not the IF) for this condition to be met,

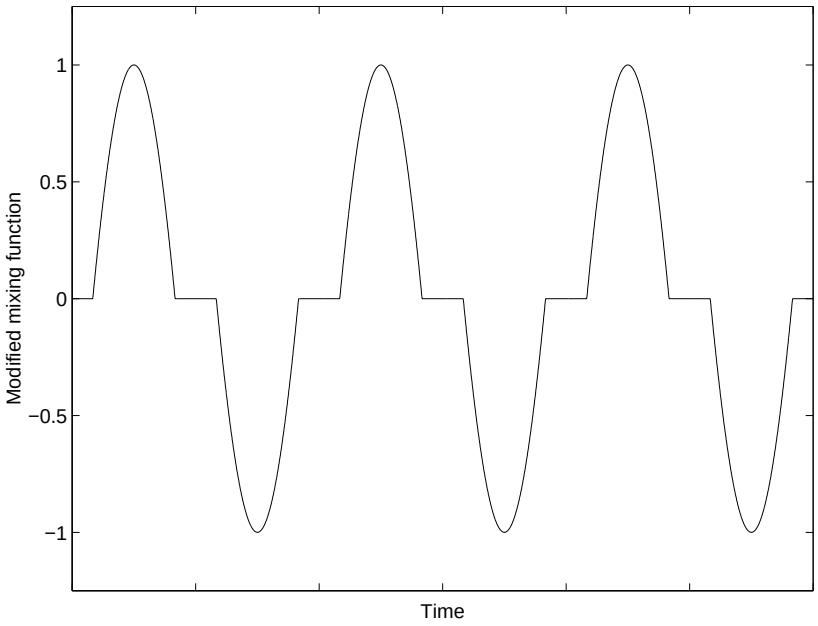


(a) Square wave drive

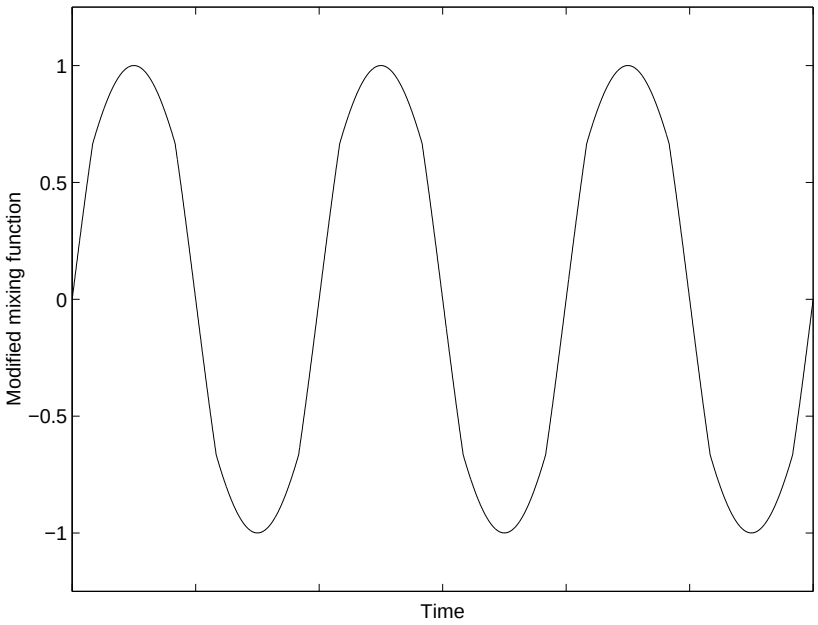


(b) Sinusoidal drive

Figure 4.24: Modified mixing functions



(a) Break-before-make



(b) Make-before-break

Figure 4.25: Modified mixing functions

although as we will see, the average mixer bandwidth as compared to the IF will affect conversion gain. For this case, the superposition integral reduces to

$$v_{if}(t) = \int_{-\infty}^{t} \frac{\overline{g_T}}{C_L} e^{-\frac{\overline{g_T}}{C_L}(t-\tau)} \left[\frac{g_T(\tau)}{\overline{g_T}} m(\tau) \right] v_{rf}(\tau) d\tau. \tag{4.36}$$

This equation looks involved, but it provides insight into the mixer's behavior. The RF voltage is evidently multiplied by a function that can be divided into a modified mixing function, which we will define as

$$m'(t) = \frac{g_T(t)}{g_{Tmax}} m(t), \tag{4.37}$$

and a gain term,

$$A = \frac{g_{Tmax}}{\overline{g_T}}. \tag{4.38}$$

g_{Tmax} is $g_T(t)$'s peak conductance, so normalization by g_{Tmax} forces the maximum magnitude of $m'(t)$ to one. Examples of the modified mixing function appear in Figures 4.24 and 4.25 for the four cases. The gain term, highlighted by g_{Tmax}, is the ratio of the peak conductance to the average conductance. With these definitions, (4.36) can be expressed as

$$v_{if}(t) = \int_{-\infty}^{t} \frac{\overline{g_T}}{C_L} e^{-\frac{\overline{g_T}}{C_L}(t-\tau)} A m'(\tau) v_{rf}(\tau) d\tau. \tag{4.39}$$

The remaining terms in (4.39) implement a very familiar component. A conventional single-pole low pass filter is shown in Figure 4.26. The superposition integral, which reduces to a convolution integral, for this filter is

$$v_{out}(t) = \int_{-\infty}^{t} \frac{g}{C} e^{-\frac{g}{C}(t-\tau)} v_{in}(\tau) d\tau. \tag{4.40}$$

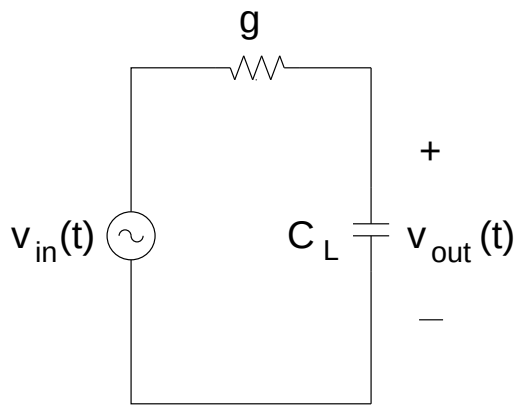


Figure 4.26: Single-pole low-pass filter

We see that (4.39) has the same form as (4.40), except g is replaced by the average conductance, $\overline{g_T}$. Now, (4.39) can be expressed as

$$v_{if}(t) = \int_{-\infty}^{t} h_{lpf}(t-\tau) A m'(\tau) v_{rf}(\tau) d\tau, \tag{4.41}$$

where h_{lpf} is the impulse response of a single-pole low-pass filter with a 3dB bandwidth of $\frac{\overline{g_T}}{C_L}$. In words, (4.41) indicates that the RF voltage is multiplied by the mixing function, then the product is scaled and filtered, as diagrammed in

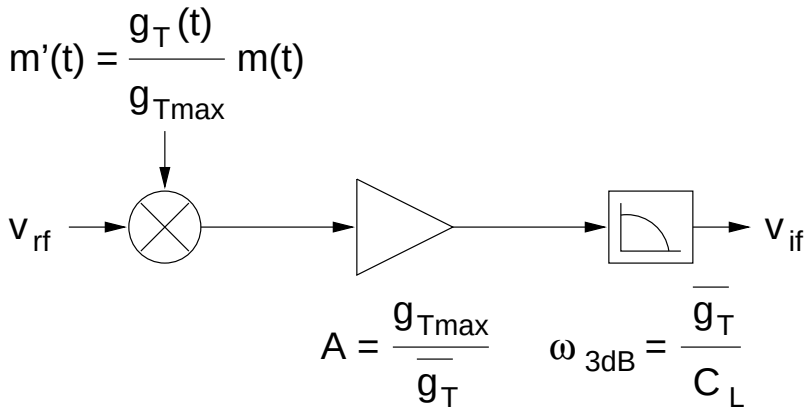


Figure 4.27: Equivalent block diagram for voltage conversion gain

Figure 4.27. Figure 4.27 helps to see that the voltage conversion gain[5] is just $A|M'(f_{LO})||H_{lpf}(f_{IF})|$.

For the remainder of this subsection, $A|M'(f_{LO})||H_{lpf}(f_{IF})|$, the conversion gain when $\frac{\overline{g_T}}{2\omega_{LO}C_L} \ll 1$, is evaluated for square, sinusoidal, and break-before make LO drives. It is very intriguing to discover that $AM'(f_{LO})$ for a sine wave with $B_{LO} = V_{th}$ gives rise to a *better* conversion gain, by a factor of $\pi^2/8$, than $AM'(f_{LO})$ for the reference square drive. For a sinusoidal drive, $|M'(f_{LO})| = 1/2$, whereas for a square drive, $|M'(f_{LO})| = 2/\pi$. But notice that the peak-to-average conductance is unity in the square wave case, while the gain, A, is $\pi/2$ in the sinusoidal case. When we take the product, $A|M'(f_{LO})|$, the total multiplication factor for a sinusoidal drive is $\pi/4$ (-2.1dB), which *exceeds* the $2/\pi$ (-3.9dB) value for a square drive.

Observing that the conversion gain of a sinusoidal drive is better than that for a square drive motivates looking at a break-before-make drive's conversion gain. It is possible, though slightly involved (see Appendix B), to express conversion gain as a function of r once again

$$\begin{aligned} A|M'(f_{LO})| =& \frac{cos^{-1}(r) - r\sqrt{1-r^2}}{2[\sqrt{1-r^2} - r\,cos^{-1}(r)]} \\ \approx& \left(1 - \frac{\pi}{4}\right) r + \frac{\pi}{4}, \end{aligned} \tag{4.42}$$

where $r = \frac{|V_{th} - B_{LO}|}{A_{LO}}$, as before. Figure 4.28 plots (4.42) as a function of r. The conversion gain actually improves as $r \to 1$, contrary to widely held beliefs. For reference, a sine wave with $B_{LO} = V_{th}$ corresponds to $r = 0$. The extreme of break-before-make action where each switch is on for just one instant of a LO cycle is $r = 1$. Although the conversion gain is higher for $r = 1$, linearity suffers, so this drive is not a practical one. Linearity degrades with increased break-before-make action because the steepness of the LO waveform at the instant of switching is reduced. We return to this topic and discuss it in further detail in §4.3.3.

[5] H_{lpf} is the single-pole low-pass filter's transfer function. Since a single-pole low-pass filter is a linear time-invariant (LTI) circuit, its transfer function can be found from the Fourier transforms of its circuit elements used in the voltage divider formula [24]. $|H_{lpf}(\omega)| = \frac{1}{\sqrt{1+(\frac{\omega}{\omega_{3dB}})^2}}$.

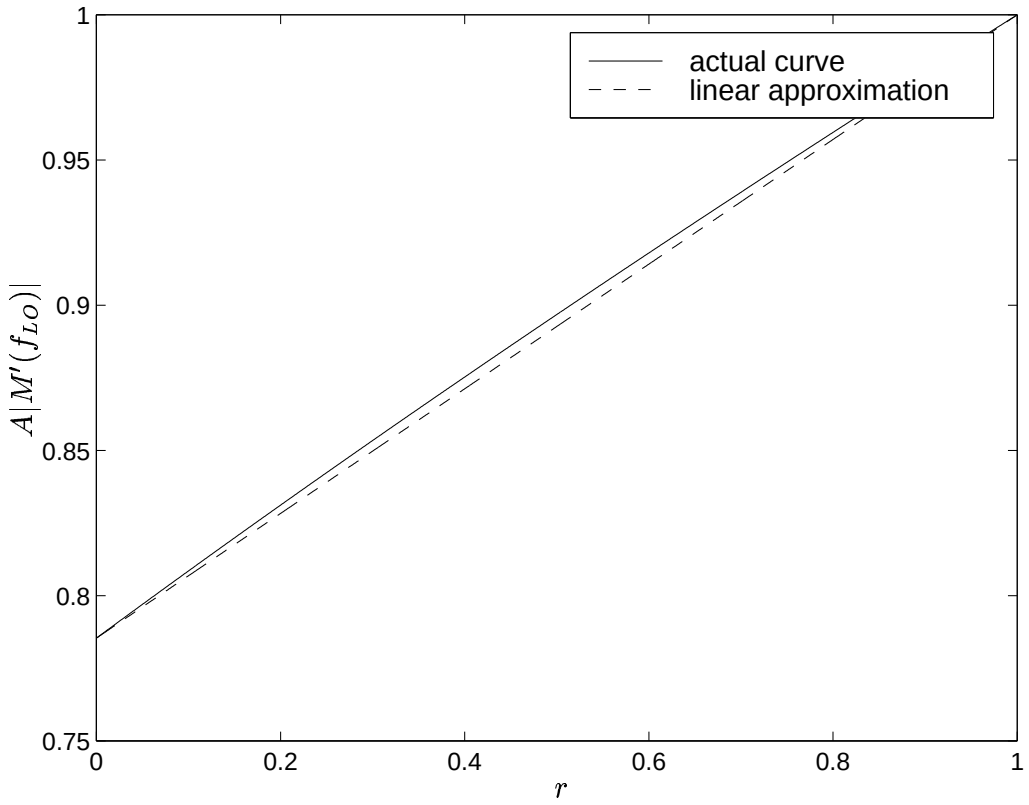


Figure 4.28: $A|M'(f_{LO})|$ vs. r for break-before-make drive

4.3.2 Noise Figure

In §4.1.3 we found the SSB noise figure for a single-response mixer to be

$$F_{SSB} = \frac{a\alpha}{4kT_oR_S}\frac{G_V^2 4kT_oR_S + \beta}{G_V^2 a\alpha} = 1 + \frac{\beta}{G_V^2 4kT_oR_S}. \tag{4.43}$$

§4.3.1 provides a method for finding G_V, but what is β for the voltage mixer?

To determine β, we start with the Thévenin equivalent model. Frequency conversion is implicit in the model, so we are only interested in spot noise at the IF. Instead of using a conductance, $g_T(t)$, we choose to deal with a resistance, $R_m(t)$, where

$$R_m(t) = \frac{1}{g_T(t)}. \tag{4.44}$$

This selection of a resistance instead of a conductance is due to the resistive source, and thus, with this choice, we anticipate a ratio of resistances in the expression for noise figure. The noise density due to the LNA output's narrowband source replaces $v_T(t)$ in the Thévenin equivalent model. Using notation established in §4.3.1.1, the noise density is that of a resistor of value $G_V^2 R_p$, assuming $\frac{\overline{g_T}}{C_L}$ exceeds the IF ($G_V = |A_{rf}||M(f_{LO})|$ or $G_V = |A_{rf}|A|M'(f_{LO})|$). This gives the picture in Figure 4.29 for determining the mixer's noise contribution. In Figure 4.29, to remove the cyclostationary nature of the mixer's noise, we use the average mixer resistance in its noise density[6] ($\overline{R_m} = \frac{1}{\overline{g_T}}$). Equation (4.43) indicates β is voltage noise in series with a source resistance $G_V^2 4kT_oR_S$. Examining Figure 4.29 and recognizing that

[6]Qualitative arguments indicate that this analysis yields a pessimistic noise figure. The switch noise of a particular switch appears at the output when its resistance is small, and therefore using an average resistance throughout, instead of a cyclostationary resistance, over estimates the mixer's noise contribution. However, the complexity of the analysis is significantly reduced, and the result is a first-order expression (i.e., within a few dB) for the mixer's noise figure.

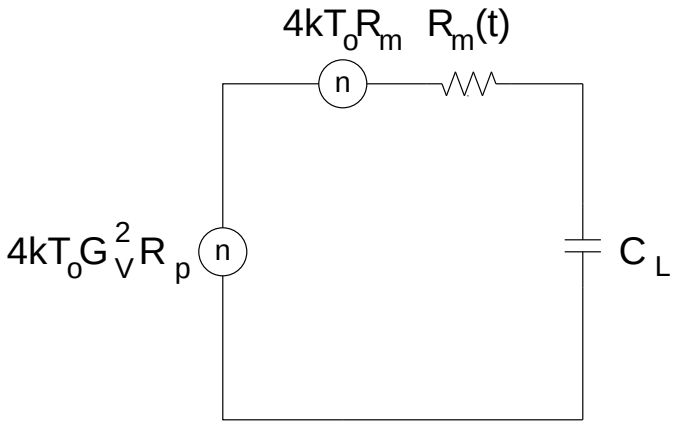


Figure 4.29: Mixer noise model

R_S is R_p, we find that β is $4kT_o\overline{R_m}$. The noise figure is easily calculated from this point to be

$$F_{SSB} = 1 + \frac{\overline{R_m}}{G_V^2 R_p}. \tag{4.45}$$

The foregoing is a simplified treatment of noise figure. However, the final expression, (4.45), does make intuitive sense. It indicates that the dominant source of noise is from the switch resistance and therefore reducing the mixer's average resistance improves its noise figure. For a more thorough treatment of the general subject of noise in mixers, the interested reader can refer to [25], where the authors present a complex theory for the analysis of noise performance in mixers. It should be emphasized that [25] outlines a general analysis applicable to any mixer architecture, whereas the analysis here is specific to the voltage mixer. A practical design formula, (4.45), results from limiting the noise figure analysis to this architecture. In contrast, to exploit the proposed general method requires implementation using a combination of available simulators and a separate two-dimensional fast Fourier transform (FFT) routine.

4.3.3 Linearity

We begin our treatment of the voltage mixer's linearity with a simple I-V relationship for a CMOS transistor operating in the linear region

$$I_d = k(V_{gs} - V_{th} - \frac{V_{ds}}{2})V_{ds}, \tag{4.46}$$

which can also be expressed as

$$I_d = k(V_{gs} - V_{th})V_{ds} - k\frac{V_{ds}^2}{2}. \tag{4.47}$$

This last expression provides important insight into the two major sources of distortion in the mixer: device nonlinearities and phase modulation (PM) of the switching instants. The first term has a factor $(V_{gs} - V_{th})$ that is a function of the device's source voltage, allowing for phase modulation of the switching instants, and the second term contains nonlinearities. The real device I-V relationship is much more complicated, and will contain other nonlinearities besides the second order nonlinearity of the simple I-V relationship (in particular, a cubic nonlinearity). Ideally, we desire an I-V characteristic of

$$I_d = k(V_g - V_{th})V_{ds} \tag{4.48}$$

to eliminate the distortion mechanisms. But since devices conforming to (4.48) are not available, we must find other ways to minimize the distortion mechanisms.

To improve the linearity of the transistors, it is most important to keep the current through the switches small to reduce nonlinear voltage drops across the devices [26]. One way to satisfy this criterion is to minimize the capacitive load, to present a high impedance to the output.

The second source of distortion arises from phase modulation of the mixing function by the RF voltage (sometimes referred to as cross-modulation), just as in diode ring mixers [27]. Borrowing from the research on diode rings, we may expect this type of distortion to diminish if larger LO drive levels are used to steepen the LO

waveform's slope as it passes through zero. A corollary is that square drives will lead to improved linearity over sinusoidal drives, if this phase modulation is the dominant source of nonlinearity. References [26] and [27] contain more detailed treatments of this type of distortion. In [27], the author accounts for cross-modulation by introducing PM terms proportional to the ratio between signal and LO amplitude into a square-wave switching function. Reference [26], extends the analysis in [27] to include switching functions with finite rise and fall times.

The remaining nonlinearities consist of parasitic junction capacitances, which are weak nonlinearities. Furthermore, at the RF port, the parasitic junction capacitances tend to be insignificant or small compared with the large, linear portion of the RF tank capacitance (e.g., from the on-chip spiral inductor's oxide capacitance). Therefore, we expect phase modulation to have the greatest impact on this mixer's linearity, implying that we should focus on the LO drive if linearity needs to be improved.

4.3.4 Switch Sizing

For M_1–M_4 in Figure 4.13, the only flexible parameter is their width, W; each device has length L_{min}, a transistor's minimum length in the given technology.

In order to resonate M_1–M_4's gate capacitance and provide a differential LO drive to the mixer, a narrowband amplifier with inductive loads is required for the mixer's LO-port interface. Figure 4.30 is a simple circuit diagram of a representative narrowband amplifier. Regardless of this amplifier's operation (e.g., class A, class AB, etc.), the power consumption is fixed once a specific amplifier current, I_A, is selected for each class of operation. However, the amplitude at each output node, A_{LO}, can vary depending on the load's resistive part at resonance. The ramification of having a fixed power consumption for different LO amplitudes becomes clear after looking at the mixer's noise figure. The analysis we perform is for a class A amplifier, but the insight it provides is independent of class. For other classes, the relationship between A_{LO} and I_A is the only difference.

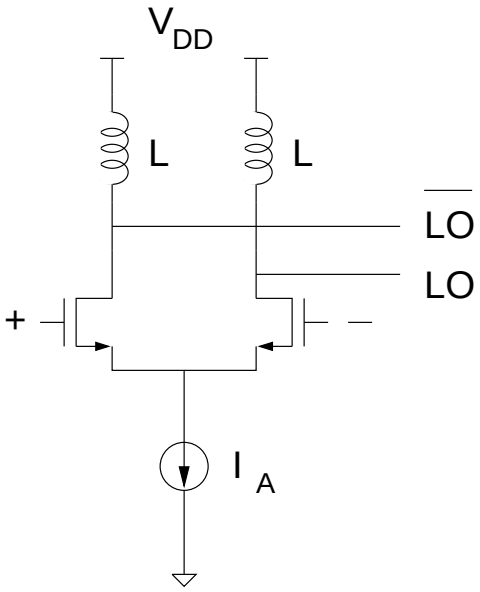


Figure 4.30: Narrowband amplifier circuit diagram

Using the quantitative expression for noise figure, (4.45), found in §4.3.3, we evaluate the integrated voltage mixer's F_{SSB}. (4.45) can be written as

$$F_{SSB} = 1 + \frac{2}{\overline{g} G_V^2 R_{LNA}}, \tag{4.49}$$

for square, sinusoidal, and break-before-make LO waveforms in Figures 4.16(a), 4.16(b), and 4.17(a), respectively, where $\overline{g}$ is the average conductance of a single device in the mixer and R_p of (4.45) is now designated with R_{LNA}. Furthermore, we plan to operate the integrated voltage mixer with a sinusoidal drive ($B_{LO} = V_{th}$), Figure 4.16(b). For this type of LO waveform,

$$\overline{g} = \frac{\mu_N C_{ox} W}{L_{min}} \frac{A_{LO}}{\pi}, \tag{4.50}$$

where μ_N is the low-field electron mobility and C_{ox} is the gate-oxide capacitance per unit area. Substituting (4.50) into (4.49) gives

$$F_{SSB} = 1 + \frac{2\pi L_{min}}{\mu_N C_{ox} W A_{LO} G_V^2 R_{LNA}}. \tag{4.51}$$

Equation (4.51) allows us to see the role of W and A_{LO}. F_{SSB} depends on the product WA_{LO}, which are mutually dependent variables. To gain more insight, we need to express F_{SSB} as a function of I_A.

To express F_{SSB} as a function of I_A, we return to the narrowband amplifier. For each inductive load,

$$Q = \frac{\omega_{LO}\mathcal{L}}{R}, \tag{4.52}$$

where $\mathcal{L}$ is the inductance and R is the inductor's parasitic series resistance. Performing a series to parallel conversion, the inductor's parallel resistance at the LO frequency is

$$R_A = (Q^2+1)R = \frac{Q^2+1}{Q}\omega_{LO}\mathcal{L}. \tag{4.53}$$

Thus,

$$A_{LO} = \eta\frac{I_A R_A}{2} = \eta\frac{I_A}{2}\frac{Q^2+1}{Q}\omega_{LO}\mathcal{L}, \tag{4.54}$$

where η is a constant factor between 0 and 1 describing the narrowband amplifier's steering efficiency. It accounts for parasitic losses in the LO driver itself that cause not all of the bias current to be available at the output.

Let us assume that the narrowband amplifier has no parasitic capacitance at either output node. Then, the capacitance on each output node is due to a pair of gates, either M_1 and M_4, or M_2 and M_3. When the mixer has a sinusoidal drive (Figure 4.16(b)), one switch in the pair is on and the other is off at any given time. Therefore, the capacitance due to each pair is

$$C_m = C_{ox}W(L_{min}+2L_{ov}) + C_{ox}W(2L_{ov}) = C_{ox}W(L_{min}+4L_{ov}), \tag{4.55}$$

where L_{ov} is the lateral-diffusion length. Capacitance C_m must resonate with $\mathcal{L}$ at the LO frequency, so

$$\omega_{LO} = \frac{1}{\sqrt{\mathcal{L}C_m}}. \tag{4.56}$$

Using (4.56) in (4.55) and solving for W gives

$$W = \frac{1}{\omega_{LO}^2 \mathcal{L} C_{ox}(L_{min} + 4L_{ov})}. \tag{4.57}$$

The WA_{LO} product is

$$WA_{LO} = \eta \frac{I_A}{2} \frac{Q^2+1}{Q} \frac{1}{\omega_{LO} C_{ox}(L_{min} + 4L_{ov})}, \tag{4.58}$$

in which the only variable is I_A. Inserting (4.58) into (4.51) yields

$$F_{SSB} = 1 + \frac{2\pi\omega_{LO} L_{min}(L_{min} + 4L_{ov})}{\mu_N G_V^2 R_{LNA} \eta} \frac{2}{I_A} \frac{Q}{Q^2+1}. \tag{4.59}$$

The interpretation of (4.59) is that the mixer's noise figure improves with increasing I_A. Once the narrowband amplifier's power is fixed, the mixer's noise figure is determined. But A_{LO} is still a free parameter. Thus, to maximize linearity, A_{LO} should also be maximized (assuming constant-Q inductors). The two design equations for the mixer are therefore

$$\mathcal{L} = \frac{1}{\eta} \frac{2A_{LO}}{I_A} \frac{Q}{Q^2+1} \frac{1}{\omega_{LO}} \tag{4.60}$$

from (4.54) and

$$W = \eta \frac{I_A}{2A_{LO}} \frac{Q^2+1}{Q} \frac{1}{\omega_{LO} C_{ox}(L_{min} + 4L_{ov})} \tag{4.61}$$

from (4.58).

In the preceding analysis, we assumed that the narrowband amplifier had no parasitic capacitance at either output node. Now, let us add a fixed capacitance,

C_f, to each output node to evaluate the effect of the narrowband amplifier's parasitic capacitance.

Equations (4.56)–(4.58) then change as follows:

$$\omega_{LO} = \frac{1}{\sqrt{L(C_m + C_f)}} \tag{4.62}$$

$$W = (\frac{1}{\omega_{LO}^2 L} - C_f)\frac{1}{C_{ox}(L_{min} + 4L_{ov})} \tag{4.63}$$

$$WA_{LO} = \eta\frac{I_A}{2}\frac{Q^2+1}{Q}\frac{1-\omega_{LO}^2 LC_f}{\omega_{LO}C_{ox}(L_{min}+4L_{ov})}. \tag{4.64}$$

If we designate the narrowband amplifier's self-resonant frequency as

$$\omega_f = \frac{1}{\sqrt{LC_f}}, \tag{4.65}$$

then we can write

$$WA_{LO} = \eta\frac{I_A}{2}\frac{Q^2+1}{Q}\frac{1-(\frac{\omega_{LO}}{\omega_f})^2}{\omega_{LO}C_{ox}(L_{min}+4L_{ov})}. \tag{4.66}$$

Now we have

$$F_{SSB} = 1 + \frac{2\pi\omega_{LO}L_{min}(L_{min}+4L_{ov})}{\mu_N G_V^2 R_{LNA}\eta\left[1-(\frac{\omega_{LO}}{\omega_f})^2\right]}\frac{2}{I_A}\frac{Q}{Q^2+1}. \tag{4.67}$$

Clearly, one design requirement is to ensure that the amplifier's self-resonant frequency exceeds the LO frequency. However, the mixer's noise figure is influenced by this self-resonant frequency, as well as the amplifier current. Once the narrowband-amplifier power is fixed, and hence I_A, ω_f remains a free parameter that can change F_{SSB}. Minimizing noise figure requires maximizing the narrowband amplifier's self-resonant frequency, leading us to select small inductors. But a small inductor yields a small LO amplitude resulting in minimum linearity.

If we try to increase the inductance, its self-resonant frequency decreases, but its parallel resistance, and thus A_{LO}, increases. These changes necessarily mean that

noise figure degrades somewhat, while linearity gets better. The real question is how much of an improvement occurs in linearity and what is the noise figure cost associated with it. Equation 4.67 quantitatively answers the latter portion of this question. Briefly examining it, we notice that the ratio of the LO frequency to the self-resonant frequency is responsible for the increase in noise figure. This ratio implies that if ω_f is much greater than ω_{LO}, then reducing it by a small amount has an insignificant effect on F_{SSB}. But, if ω_f is only slightly greater than ω_{LO}, then reducing it by a small amount has a much more pronounced effect on F_{SSB}.[7] This large cost may or may not be acceptable for the accompanying linearity boost.

Lastly, in both (4.59) and (4.67) we see that F_{SSB} improves as L_{min} decreases, indicating that noise figure improves as technology scales. However, as the term involving L_{min} approaches 1, diminishing returns sets in as L_{min} is reduced.

4.4 Summary

This chapter presented a new CMOS mixer architecture that uses four transistors to commutate signals in the voltage domain. The topology is simple and has no static power consumption. In a performance comparison to other architectures, the voltage mixer was shown to be the best, if low conversion gain is tolerable. Furthermore, scaling trends additionally favor this architecture.

To analyze the mixer, it was necessary to recognize its time-varying nature and use a linear time-varying mixer model. This model allowed closed-form expressions for the voltage conversion gain under different operating conditions. It was startling to discover through this analysis that the voltage conversion gain for a capacitively loaded voltage mixer can exceed $\frac{2}{\pi}$ and approach unity.

In addition to conversion gain, we have also presented a noise analysis of the mixer which indicates its ability for good noise performance. When we couple this information with the fact that the architecture can support large voltage swings on the mixer devices, the result is a topology that boasts of a wide dynamic range.

[7] The percentage of C_f due to the inductor's self-capacitance determines the degree by which a change in $\mathcal{L}$ affects ω_f.

Chapter 5

Experimental CMOS Voltage Mixers

THIS chapter describes two prototypes of the CMOS voltage mixer. The first prototype demonstrates this mixer architecture's performance, and its design is driven by the need to characterize the mixer as an individual block. The second prototype is a functional block within a GPS radio, and its design is coupled with that of the radio's signal path.

The first prototype mixer is packaged and placed on a test board, so consequently the mixer interfaces with structures on the test board, through bondwires and the package. Since the board and on-chip environments are very different, special care is taken with the circuits that connect off-chip. This standalone mixer is made in a 0.35μm, single-poly, double-metal, CMOS process that is mainly used for digital circuits.

The second, integrated prototype makes use of the performance information collected from testing the first prototype. The performance of the CMOS voltage mixer influences architectural choices for the radio; specifically, due to the negligible power consumption in this type of mixer, selecting architectures that use multiple mixers to achieve image rejection does not adversely affect the radio's power consumption. The GPS radio, which includes two voltage mixers in the signal path (see Figure 3.4), is fabricated in a 0.5μm, single-poly, triple-metal, CMOS process.

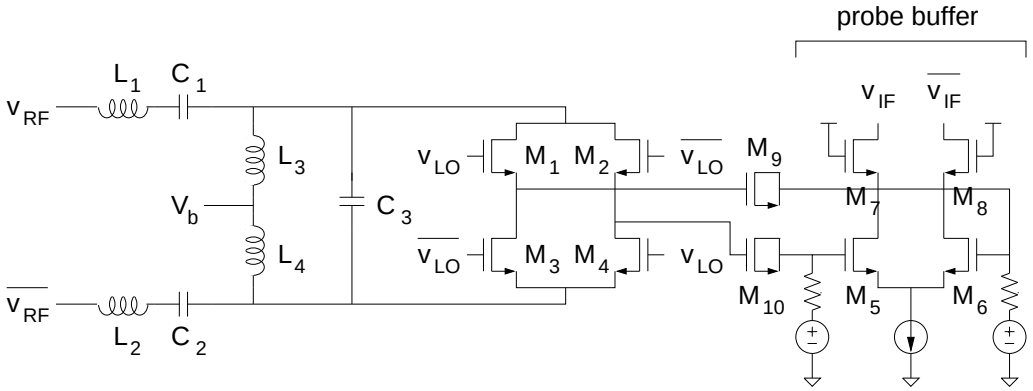


Figure 5.1: Standalone-mixer circuit diagram (shown with probe buffer)

5.1 A Standalone Mixer

Figure 5.1 shows the circuit schematic of the first mixer prototype. The mixer is formed by the four transistors, M_1–M_4, with reactive terminations on the mixer's RF port and IF port.

At the RF port, a passive filter network precedes the mixer and can be separated into two parts for convenient analysis: an L-match and an RF tank. The L-match is formed by inductors L_1 and L_2 with part of the tank capacitance, C_3, while the RF tank is formed by inductors L_3 and L_4 with the remainder of C_3. C_1 and C_2 are AC coupling capacitors implemented with parallel plate metal-to-metal capacitors, L_1–L_4 are implemented with bondwires, and C_3 is a metal-to-metal capacitor that incorporates lateral flux as well as vertical flux. The quad's DC bias is set through an off-chip voltage source and resistor that connect to the port labeled V_b. The filter network is analyzed in greater detail in §5.1.2.

At the IF port, the probe buffer provides a capacitive load for the mixer in the form of devices M_5 and M_6's gate capacitance. M_1–M_4's gate capacitance must also be accounted for, as well as parasitic capacitances, such as the junction capacitances at the quad's output. M_9 and M_{10} form AC coupling capacitors by operating in inversion with their respective source and drain terminals connected together. The probe buffer's cascoded output is left with open drains for testing.

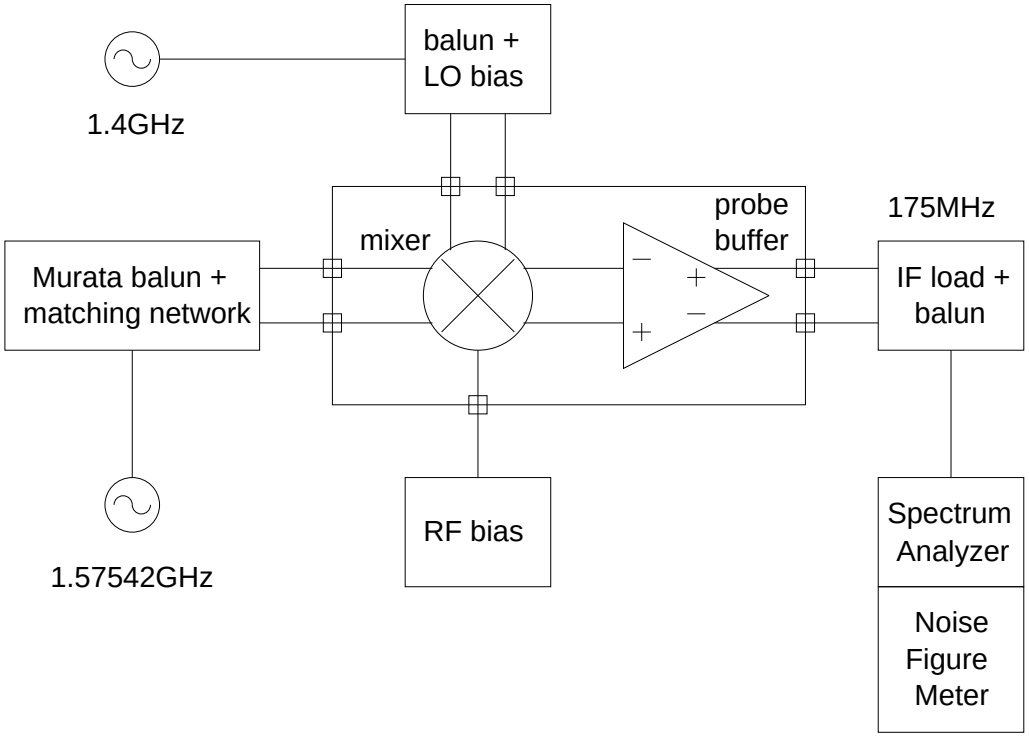


Figure 5.2: Standalone-mixer test setup

5.1.1 Testing Issues

Figure 5.2 shows the experimental setup used to test the standalone mixer. The two primary issues of interest are the conversion between single-ended and differential signals, and the 50Ω board environment. For the surface mount baluns used, the insertion loss is less than 0.8dB per balun. In a complete integrated radio, only one balun would be required to transform the single-ended signal from the antenna and RF filter into differential form. Some surface acoustic wave (SAW) filters naturally provide a differential output, in which case no balun is necessary.

The RF signal is converted from the single-ended output of a bench top signal generator to a differential drive by a Murata balun designed for the GPS carrier frequency. A matching network near the package is also necessary to enable the mixer's input impedance to be matched to 100Ω differential, since the balun provides

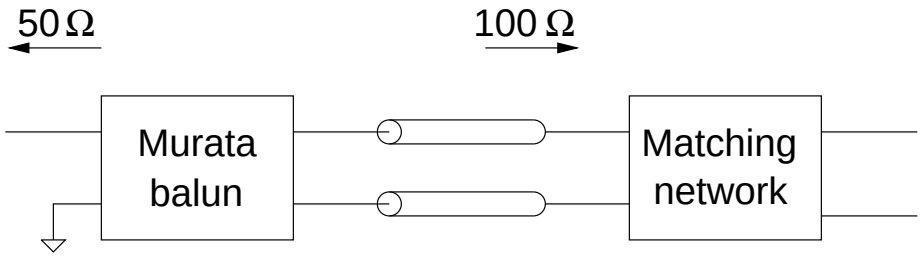


Figure 5.3: RF board structures

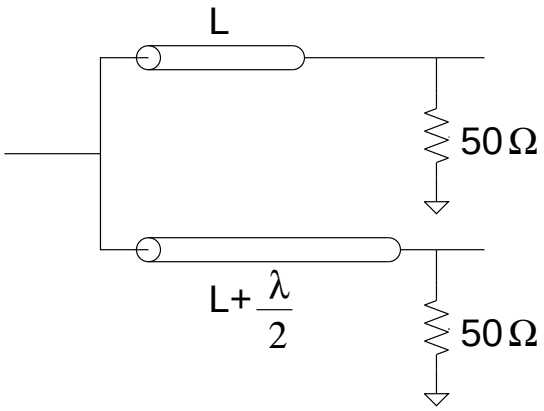


Figure 5.4: LO board structures

a 1:2 impedance ratio from the single-ended to differential side. Figure 5.3 illustrates this description.

The LO signal is also converted from the single-ended output of a bench top signal generator to a differential drive, but in this case by splitting the LO board trace into two that differ by half a wavelength from the fork to the package. Because the two traces produce a differential output at a frequency related to the length difference, a fixed LO at that frequency is used [28]. Each trace is terminated in 50Ω. Figure 5.4 pictures the LO's conversion from a single-ended to differential signal.

The IF output signal is sent off-chip in the form of a current, due to the open drain outputs of the probe buffer, and delivered into a 50Ω differential load. The load is created by an IF balun, also with a 1:2 impedance ratio from the single-ended to differential side, and two 50Ω resistors to ground on the differential side. The single-ended output of this balun can be connected to a spectrum analyzer or a

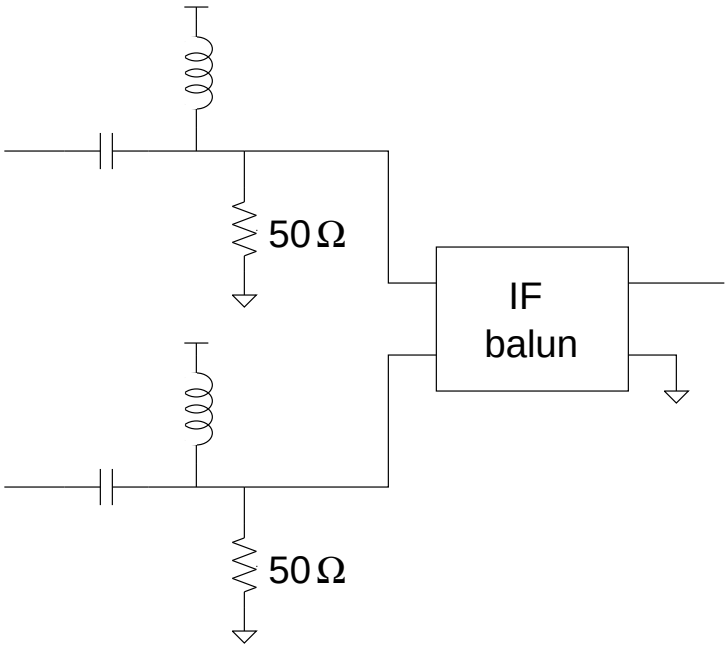


Figure 5.5: IF board structures

noise figure meter, both of which have 50Ω input connectors. Figure 5.5 shows the IF board structures.

The probe buffer is designed to attenuate so that its linearity does not interfere with the mixer linearity measurement. In a real chip, an IF amplifier (IFA) would replace the probe buffer, and since test equipment no longer needs to be driven, the IFA's design can proceed without having to drive a 50Ω differential load.

5.1.2 Filter Network

Because the mixer's load presents a high impedance, there is a negligible effect on the mixer's RF port voltage, $v_{rf}(t)$, during operation. Therefore, $v_{rf}(t)$ results from an LTI network driven by a voltage source, $v_s(t)$, which in the frequency domain is

$$V_{rf}(f) = H(f)V_s(f), \tag{5.1}$$

where $H(f)$ is the transfer function from the voltage source to the mixer's RF port.

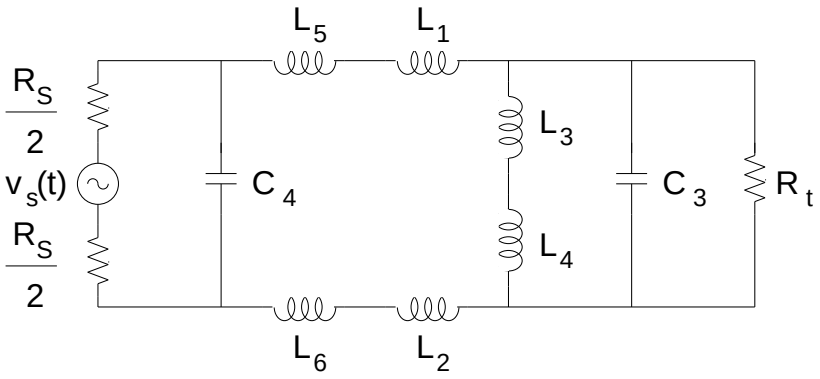
As discussed previously, the filter network consists of an RF tank and an L-match, which is redrawn in Figure 5.6(a) along with the external matching network, represented with inductors L_5 and L_6 and capacitor C_4. The differential value of the source resistance, R_S, equals 100Ω (Figure 5.6). By working from R_S towards the tank resistance, R_t, which models the tank parasitics, we will reduce the external matching network and L-match to find $H(f_{RF})$, through a transformation that is valid at the operating frequency.

A brief description of the transformation process follows. First, the RF tank is eliminated, as shown in Figure 5.6(b). The on-chip L-match, consisting of inductors L_1 and L_2 and capacitor C_5, boosts its input resistance to R_t, where the input resistance is an intermediate resistance, R_i. To fix the discrepancy between R_i and R_S requires an external matching network, consisting of inductors L_5 and L_6 and capacitor C_4, which lowers R_S to R_i. Together, the on-chip L-match and external matching network form a π-match between R_S and R_t. A matched condition exists at the end of our transformations, since the source drives the network at resonance.

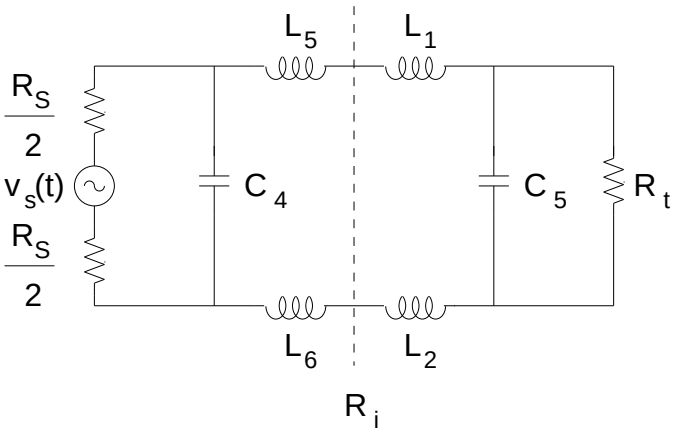
This condition is depicted in Figure 5.6(c) where the transformation is defined by the quality factor, Q, of the π-match, evaluated at the RF. From these simplifications, it is clear that

$$H(f_{RF}) = \frac{R_t}{(Q^2+1)R_s + R_t}\sqrt{Q^2+1}. \tag{5.2}$$

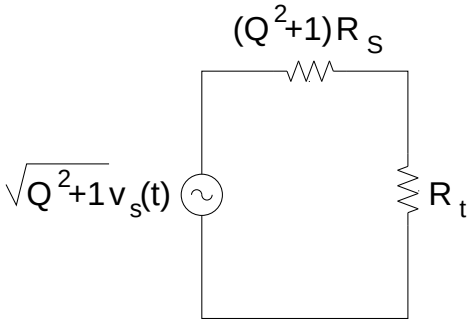
Note that if this network transforms the source resistance to match the tank resistance ($(Q^2+1)R_s = R_t$), then $H(f_{RF}) = \frac{\sqrt{Q^2+1}}{2}$.



(a) Filter network



(b) RF tank reduced



(c) π-match reduced

Figure 5.6: Passive-network reduction at resonance

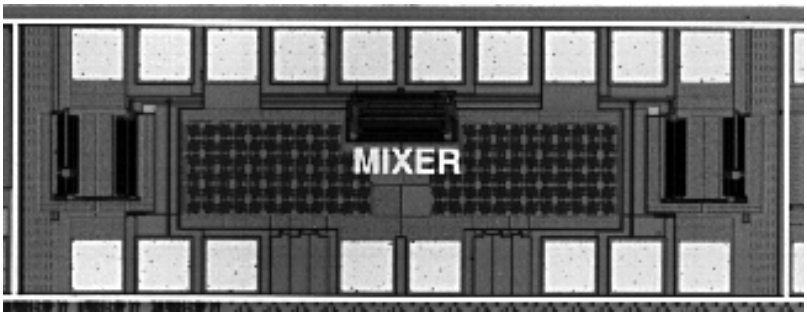


Figure 5.7: Standalone-mixer die photo

5.1.3 Experimental Results

In all of the discussion that follows, "dBm" is used in its original, rigorous sense: the signal power referenced to 1mW and expressed in dB. In cases where the impedance is not well known (and hence, the power difficult to quantify), voltage units are used explicitly to avoid confusion.

A die photograph of the standalone mixer is shown in Figure 5.7. It is made in a 0.35μm, single-poly, double-metal CMOS process. The aspect ratio of the silicon is somewhat unusual because this project was designed to fit in the scribe lane of a wafer that was primarily devoted to other dice. Accordingly, the dimensions are 350μm x 1mm[1].

Measurements are made with a differential LO amplitude of 300mV, biased about the switch threshold voltage. This LO amplitude is equivalent to -3.5dBm in a 100Ω impedance. Note, however, that the terminating impedance for the LO port need not be 100Ω if the LO were integrated with the mixer. Indeed, a higher impedance could be achieved with spiral inductor tuning of the LO port to further reduce LO power.

[1]Including the area of pads.

5.1.3.1 Voltage Conversion Gain

The voltage conversion gain, $|H(f_{RF})||M(f_{LO})|$, [2] is -3.6dB. The measurement is performed with the input port of the mixer impedance matched to 100Ω (differential). Note that, without the impedance transformation of the external matching plus L-match network, the expected voltage conversion gain should be close to -10dB. This value includes -6dB for the voltage attenuation from matching the input port and ≈ -4dB for the conversion component. We may infer that the Q of the external matching plus L-match network is approximately 1.8, resulting in a factor of 2.1 step up in voltage before the mixer. The RF tank thus presents an equivalent parallel resistance of about 440Ω at resonance, corresponding to a total network Q of about 11.

5.1.3.2 Noise Figure

Experimental Setup Figure 5.8 shows the noise-figure experimental setup using the HP8970B noise-figure meter.

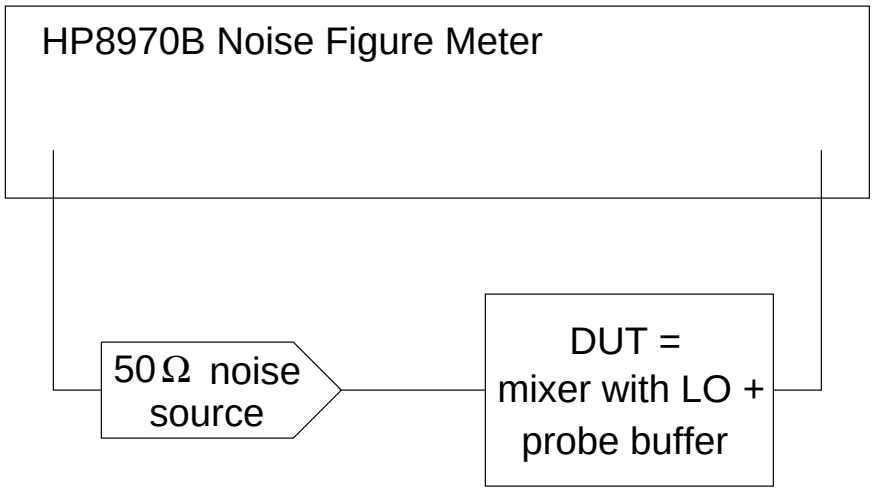


Figure 5.8: Noise-figure test setup

[2] $|M(f_{LO})|$ can be found from Table 4.5.

Data The mixer's SSB noise figure is estimated to be 10dB, based on noise figure measurements of the mixer/probe buffer combination. Appendix B contains a detailed description explaining how the mixer's noise figure is determined from the noise figure meter's readout. To put this mixer's noise figure in context, assuming an LNA with 17dB gain and 3.8dB noise figure, this mixer contributes 0.3dB to the noise figure of the LNA/mixer combination.

5.1.3.3 Linearity

Experimental Setup The linearity of the mixer was measured with a two-tone IP3 test with tones at 1.575GHz and 1.585GHz. Figure 5.9 shows the IP3 experimental setup. The tones are generated from two separate bench-top references, and combined through a resistive T-structure. Each reference generator has a 10dB attenuator placed on its output to reduce interference from the other generator (there is thus greater than 20dB of attenuation between generators).

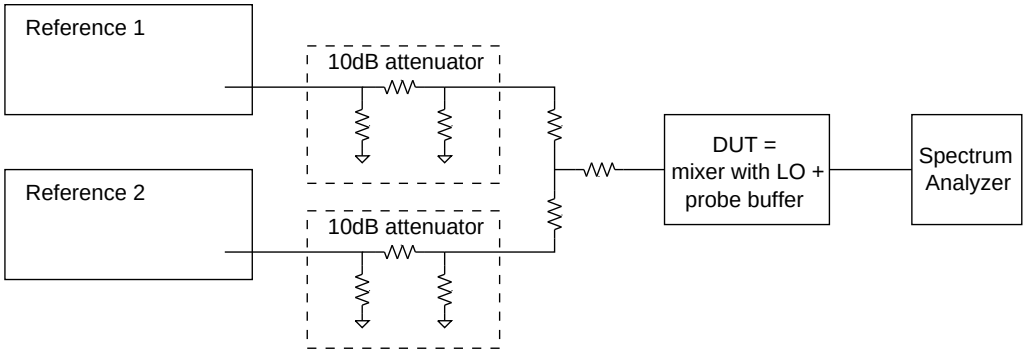


Figure 5.9: IP3 test setup

Data Figure 5.10 plots the collected data. The fundamental output amplitude is extrapolated along a line of unity slope, while the third-order intermodulation products are extrapolated along a line with a slope of three, using the products at higher source power as a reference. From the intersection of the lines, the input-referred IP3 is about 10dBm. The products at higher source power are used for extrapolation because they are well above the spectrum analyzer's noise floor, and therefore are more certain. The reason the products at low source powers are near the noise floor is a consequence of the attenuating probe buffer.

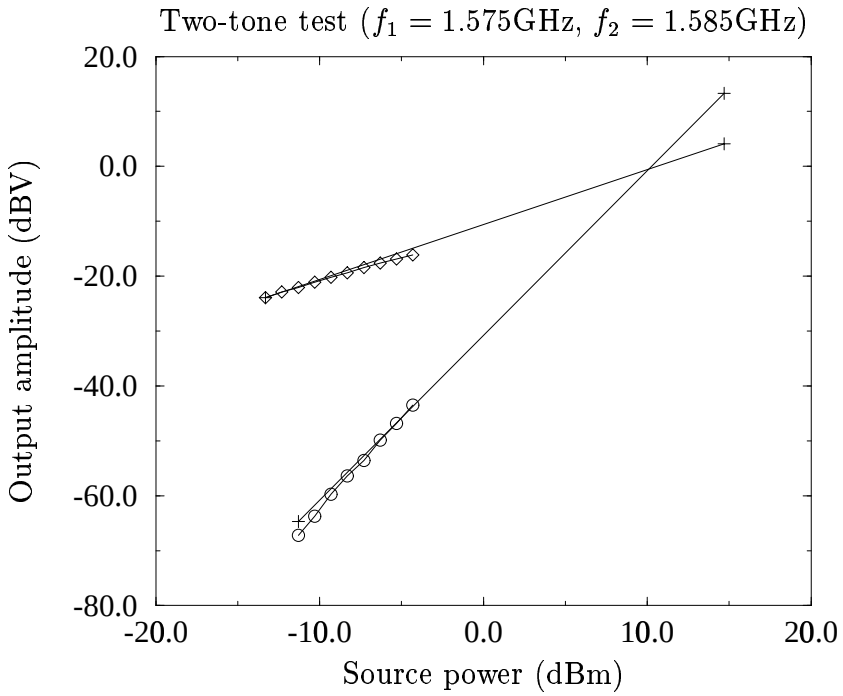


Figure 5.10: Mixer two-tone IP3 measurement

Table 5.1: Standalone-mixer performance summary

LO Frequency	1.4GHz
LO Amplitude *	300mV
Voltage Conversion Gain	-3.6dB
SSB Noise Figure	10dB
IIP3	10dBm
Input 1dB Compression	-5dBm
Supply Voltage	1.5V
Technology	0.35μm CMOS
Die Area †	0.35mm^2

*≈ -3.5dBm in 100Ω.

†The die area includes the area of pads.

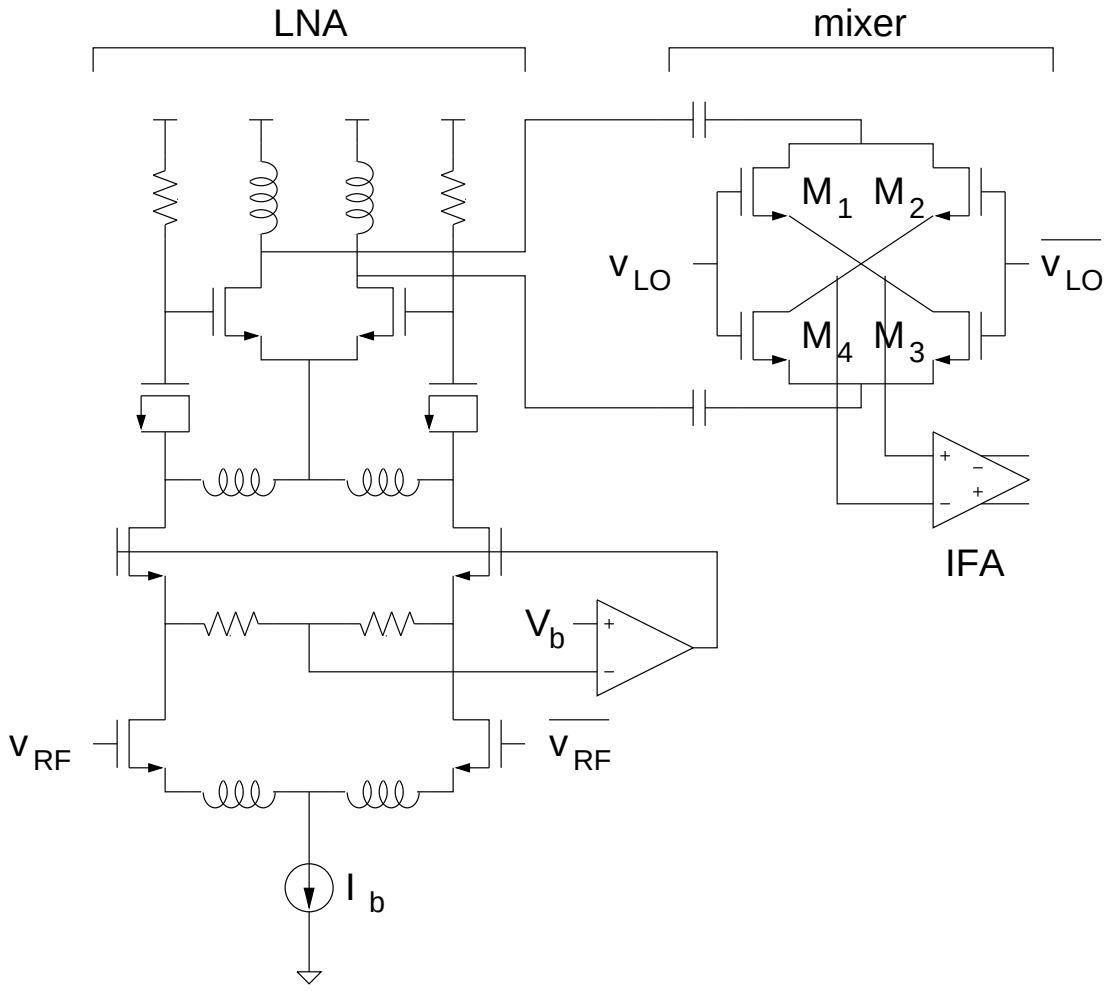


Figure 5.11: Integrated mixer circuit diagram

5.2 An Integrated Mixer

Figure 5.11 shows the circuit diagram of the second mixer prototype. The mixer is formed by the four transistors, M_1–M_4, and is placed between an LNA at the RF port and an IFA at the IF port. The IFA's input impedance is capacitive and the LNA's output impedance is that of a tuned circuit. Because we are in the on-chip environment, the LNA's output resistance at the RF can be large, and is $\approx 400\Omega$ differential in this case.

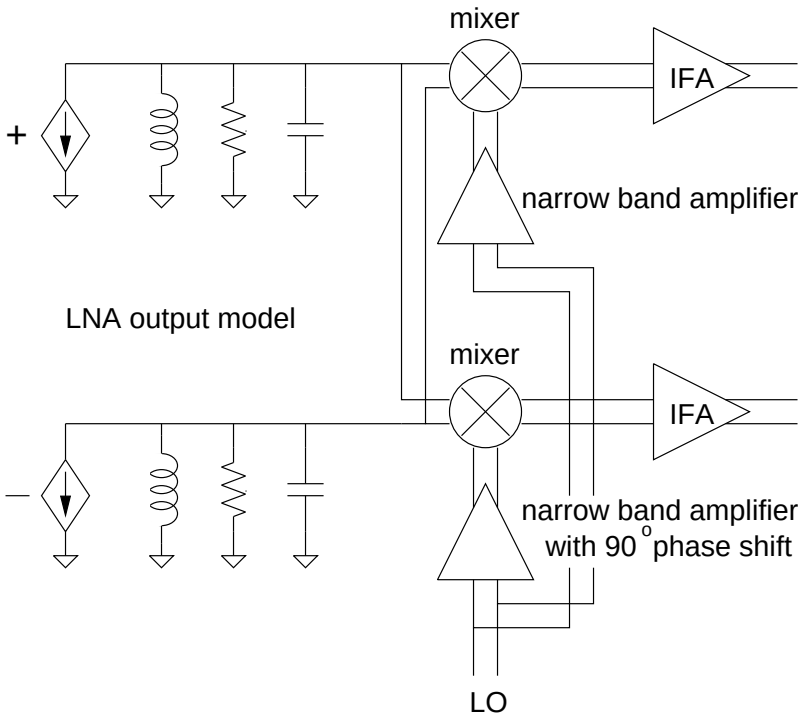


Figure 5.12: IP3 simulation setup

5.2.1 Simulated and Experimental Results

Two simulations are performed on the integrated voltage mixer in a SPICE based circuit simulator using BSIM3 transistor models. The first is an IP3 test, which provides the voltage conversion gain and IIP3. The second is a DC simulation of a single transistor from the mixer quad to find values for k and V_{th} in (4.32). Then, we use (4.49) to estimate noise figure.

Figure 5.12 shows the simulation setup in the circuit simulator. Instead of a single mixer driven from the LNA and loaded by an IFA, the simulation includes both in-phase and quadrature mixers. Using two mixers is due to the GPS radio architecture in Figure 3.4 where the in-phase and quadrature mixers share the same input port. Therefore, they interact with each other, and their influence on one another must be captured in the simulation. The IP3 simulation includes a circuit model of the LNA output, both in-phase and quadrature mixers, IFA loading, and two narrowband amplifiers, with quadrature outputs to one another, to drive the two mixers. Each

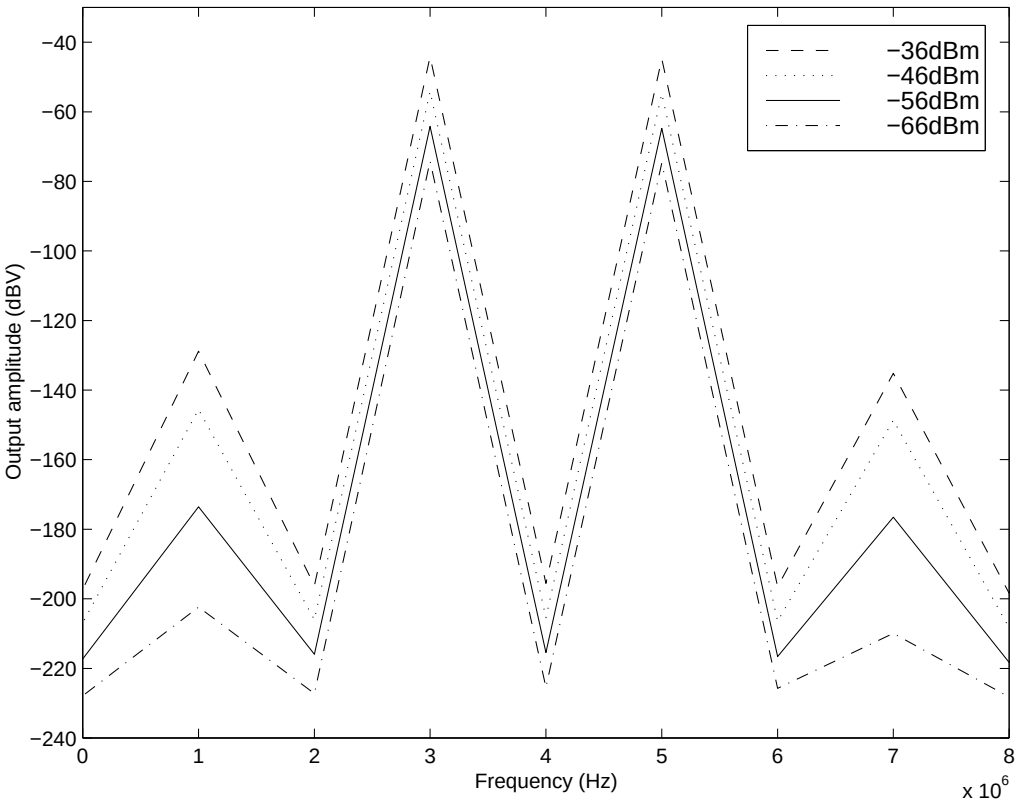


Figure 5.13: Output intermodulation products for different available source powers

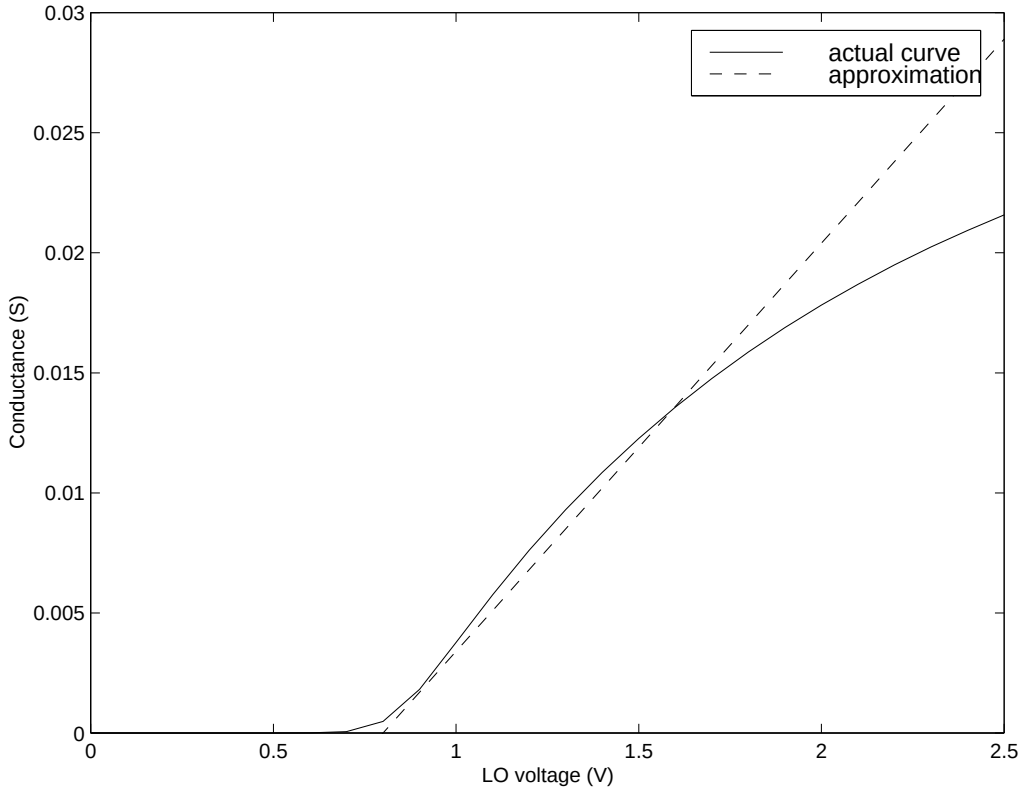


Figure 5.14: DC switch conductance

block is described at the transistor level in the simulator, but for the purpose of explaining the simulation, functional blocks are shown in Figure 5.12. Simulations at different power levels are performed and the fast Fourier transform (FFT) at the in-phase mixer output is executed to verify that the observed intermodulation products have a slope of three on a log-log plot. Figure 5.13 displays the FFT simulation results and Table 5.2 summarizes the simulated voltage conversion gain and IIP3.

The device conductance simulation is performed with $V_{ds} = 0.01$V and $V_{bs} = -1.7$V. Figure 5.14 displays the actual DC conductance vs. gate voltage and the approximation we use ($k = 0.017$ and $V_{th} = 0.8$). Now we have all that is required to estimate noise figure using (4.49). Our previous simulation provides us with G_V, in which $R_{LNA} = 400\Omega$ and $A_{LO} = 0.9$V. The value for V_{th} verifies that the mixer is

Table 5.2: Integrated-mixer performance summary

	Simulated	Measured
LO Frequency	1.573GHz	1.573GHz
LO Amplitude *	0.9V	0.9V
Voltage Conversion Gain	−7.3dB	−7.3dB
SSB Noise Figure	8.1dB	6dB
IIP3 †	−1.3dBm	4dBm
Supply Voltage	2.5V	2.5V
Technology		0.5μm CMOS
Die Area		0.0084mm^2

*Single-ended.
†Available source power.

DC biased at its threshold voltage, and the value for k is used with A_{LO} in (4.50) to evaluate $\overline{g}$. Table 5.2 shows the estimated noise figure.

Table 5.2 also contains the results of experimental measurements on the integrated mixer. The data in the table are inferred from signal-path measurements made at a test point between the filter and limiting amplifier [15]. Comparing the measured performance to the simulated performance, we find that our estimates for noise figure and linearity are conservative. This last result in not surprising, since the MOS models used in simulation can contain discontinuities for certain parameters at region of operation boundaries.

5.3 Summary

This chapter has shown the performance of two prototype CMOS voltage mixers. The standalone and integrated mixers confirm the expected performance presented in Chapter 4. Both designs validate the conversion gain analysis in Chapter 4, and the integrated design shows that the first-order quantitative noise figure analysis provides a good, but conservative, estimate.

We have seen that there is a dramatic difference between the design of a standalone versus an integrated mixer. In the former, the board, package, bond wires, and bonding pads all have an impact on performance, and if not accounted for, the effect can be severe. In contrast, an integrated design affords precise control over the mixer port terminations, in addition to greater flexibility with port impedances (e.g., the LNA's output resistance is no longer constrained by the 50Ω board environment). These benefits motivate integration of the mixer.

Chapter 6

Dividerless Frequency Synthesis

PART of offering a completely integrated radio solution involves providing a low-power, monolithic gigahertz local oscillator. A quartz crystal based oscillator cannot be used directly for the LO since the fundamental modes of inexpensive quartz crystals are limited to approximately 30MHz [29], and overtone orders of 50 are impractical. However, a crystal oscillator can be used as the reference in a static-modulus PLL frequency synthesizer. As is well known, the stability of the frequency-multiplied reference is retained by a wideband loop. This ability to synthesize a stable high frequency source is beneficial, but it usually comes at the expense of significant power consumption. This chapter addresses the power issue by introducing a new type of phase detector, called an aperture phase detector, capable of phase locking the synthesizer's frequency-multiplied output to its reference input, without the use of a divider. In addition to reducing power, removing the divider also reduces the PLL's on-chip interference that is caused by the divider's large high-speed digital transitions.

6.1 Loop Basics

Before proceeding to this new loop architecture, we must review the concept of phase lock. Figure 6.1 shows the block diagram of a general phase-locked loop, which consists of a phase detector and voltage-controlled oscillator. The input signal,

$$v_i(t) = A_i cos(\omega_i t + \phi_i) + B_i, \tag{6.1}$$

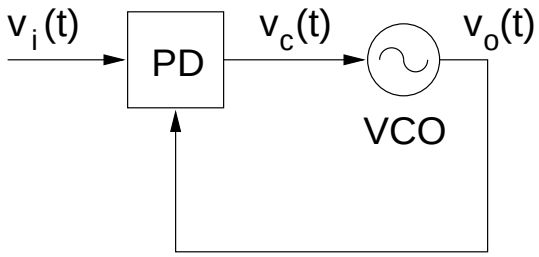


Figure 6.1: General PLL representation

is represented as a sinusoid with amplitude A_i, angular frequency ω_i, phase ϕ_i, and DC level B_i. The PLL forces the phase-locked signal, $v_o(t)$, to oscillate at ω_i with some fixed phase relationship between $v_i(t)$ and $v_o(t)$.

Phase lock is accomplished in the following manner. To make $v_o(t)$ oscillate at ω_i, the corresponding control voltage must be applied to the VCO. The phase detector produces a control voltage, $v_c(t)$, for the VCO that is a function of the phase difference between its two inputs. The control voltage servos the VCO's oscillation frequency to reduce this phase difference. Lastly, the phase detector maintains the right control voltage by adjusting to the proper fixed phase relationship between its inputs.

Once the loop has achieved lock, we desire a more comprehensive understanding of the phase relationship between input and output. Therefore, in lock, we model the phase transfer function with the linear time-invariant (LTI) system shown in Figure 6.2. In the phase domain, for small phase perturbations, the phase detector behaves as a subtractor, providing a current or voltage at its output proportional to the phase difference. The phase detector can also include a filter to smooth the current or voltage before being applied to the VCO. This filter is represented as an s-domain transfer function, $H(s)$, which can represent a filtering impedance (when transforming current to voltage) or a voltage filter (when the input is a voltage). The VCO is modeled as a pure integrator with its own proportionality constant, which is justified by the following. For a fixed voltage applied to the VCO, the output frequency is a constant, so phase accrues at a fixed rate. Thus, if the phase is plotted versus time, without introducing a modulo 2π, the result is a ramp. Because the step response of an integrator is a ramp, the VCO behaves as an integrator when

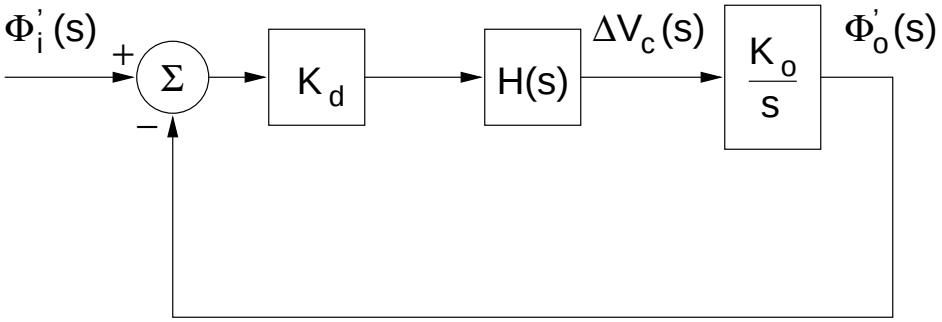


Figure 6.2: Basic-PLL LTI block diagram

phase is the output of interest.

The loop's closed-loop phase transfer function is readily found from Figure 6.2 using feedback theory [30] to be

$$\frac{\Phi_o'(s)}{\Phi_i'(s)} = \frac{K_d K_o H(s)}{s + K_d K_o H(s)}. \tag{6.2}$$

Equation (6.2) reveals that incremental input phase changes appear in the output phase with a gain of near unity for low frequencies. At higher frequencies, the PLL loses its effectiveness and the output phase is less sensitive to the input phase [31].

For feedback systems, both the closed-loop transfer function and the loop transmission are important. The loop transmission measures the frequency response (amplitude and phase) of signals that make one trip around the loop [32]. To find it, we break the loop after the subtractor and find the transfer function from the input, at the proportional gain block, to the output of the subtractor when the subtractor's positive input is zero:

$$\Gamma(s) = -\frac{K_d K_o H(s)}{s}. \tag{6.3}$$

The loop transmission is important because it indicates the PLL's loop bandwidth and governs the stability of the loop. The loop bandwidth is the frequency at which the loop transmission's magnitude is unity, and also corresponds to the frequency around which (6.2) departs from its low frequency (i.e., near unity) gain. As hinted, this frequency corresponds to a bandwidth, because to phase inputs the PLL acts

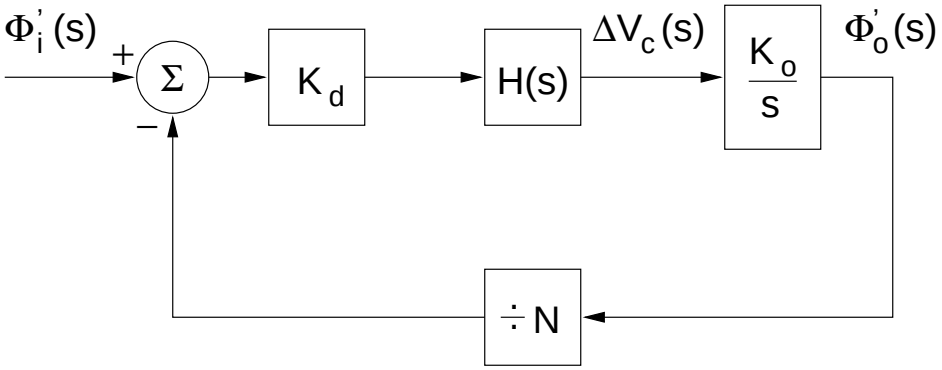


Figure 6.3: Integer-N-PLL LTI block diagram

as a filter whose 3dB bandwidth is governed by the loop transmission. The stability question uses both the loop transmission's magnitude and phase. The least rigorous criterion that must be met is to ensure in the design that there is some phase margin[1] between $\angle -\Gamma(j\omega)$ evaluated at the loop bandwidth and $-180°$.

If we modify Figure 6.2 by placing a divide-by-N block in the feedback path, as shown in Figure 6.3, the new closed-loop phase transfer function is

$$\frac{\Phi_o'(s)}{\Phi_i'(s)} = \frac{NK_dK_oH(s)}{sN + K_dK_oH(s)}. \tag{6.4}$$

Now, when the input phase appears in the output phase, it does so with a scale factor of N, within the loop bandwidth. A DC gain of N means that the output is N times the input frequency, since frequency is the derivative of phase. The fact that the phase-locked signal oscillates at a multiple of the input frequency motivates the implementation of Figure 6.3's divide-by-N block with a frequency divider (a frequency divider necessarily performs both frequency and phase division). But the divider is an explicit element that would have to be added to the loop. An alternate approach is to investigate if the phase division can be implicitly performed in the phase detector. As we see next, the result is that it can, thereby eliminating the need for an explicit divider to maintain phase lock.

[1]Typically, the loop bandwidth is designed to be less than one-tenth the loop's input frequency to ensure stability.

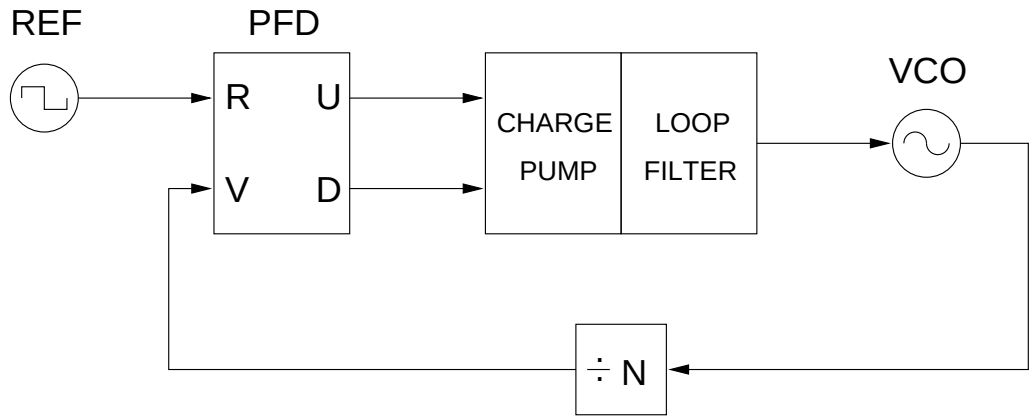
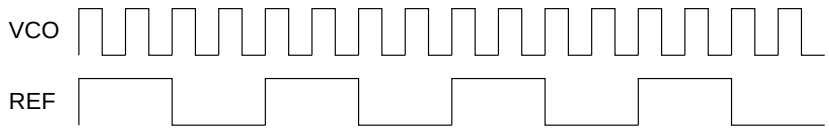


Figure 6.4: Integer-N synthesizer architecture

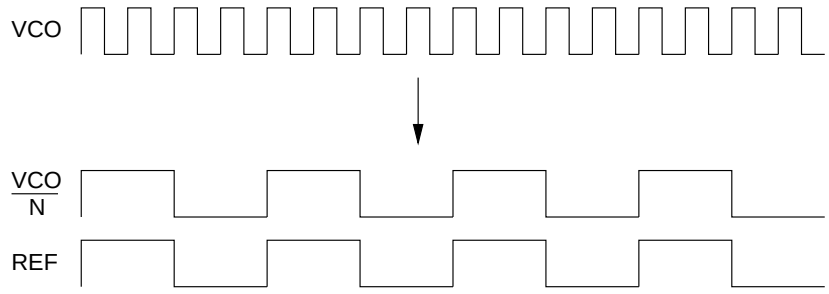
6.2 Architecture

A conventional and widely used implementation of the PLL frequency synthesizer modeled in Figure 6.3 is the integer-N synthesizer shown in Figure 6.4 [32]. As seen, it uses an explicit element, separate from the phase detector, to perform the feedback frequency division. One method to implement the divide-by-N block is with a single counter. However, power consumption and switching noise are two drawbacks associated with the divider. Power consumption is large, particularly at high frequencies, because of the well known CV^2f relationship [33]. For example, a recently published 1.6GHz integer-N synthesizer built in a $0.6\mu m$ CMOS technology reported a total power consumption of 90mW, of which 22.5mW were used by the divider [34]. A further disadvantage of the divider is the on-chip interference generated by its high-speed digital transitions. This noise is particularly worrisome if the synthesizer is to be integrated with the radio's sensitive LNA.

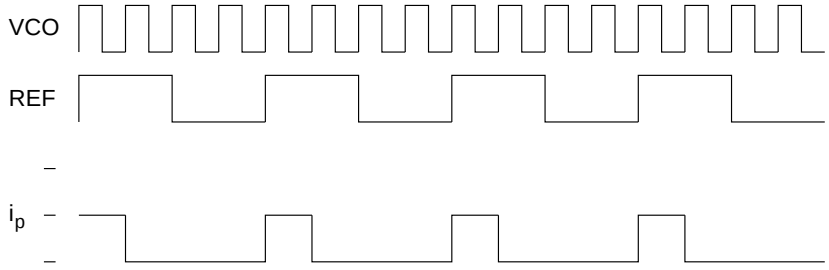
To reduce power consumption and high-frequency interference, a windowing technique that eliminates the divide-by-N block for phase comparisons is investigated. To appreciate how windowing may be of benefit, it is worthwhile to revisit the phenomenon of locking in a conventional PLL. To retain phase lock, it is necessary to align every Nth rising or falling edge of the VCO with a corresponding reference edge. Phase lock is demonstrated in Figure 6.5(a) for $N = 4$, where every 4th rising



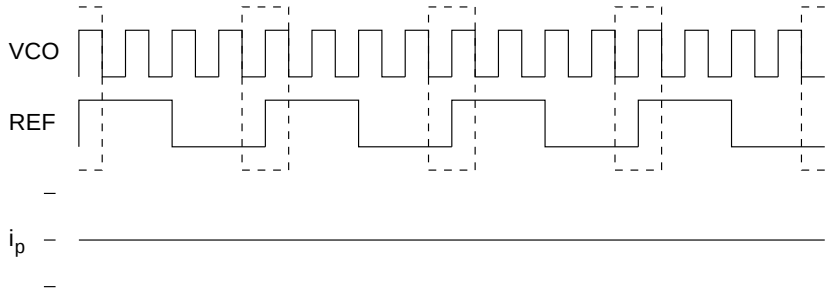
(a) Phase-locked signals



(b) Phase lock with a divider and PFD



(c) PFD alone; negative charge pump current commands the VCO to decrease its frequency, breaking phase lock



(d) Phase lock with an APD

Figure 6.5: Phase-lock techniques

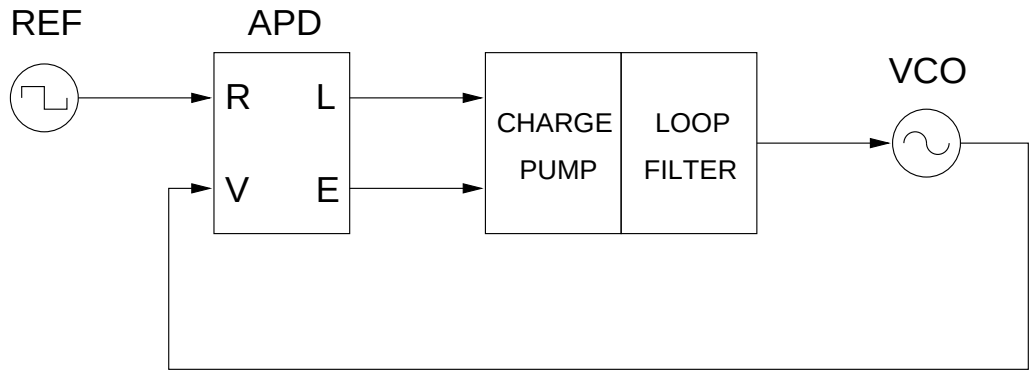


Figure 6.6: APD synthesizer architecture

VCO edge lines up with a rising reference edge. A divider with a phase/frequency detector (PFD) accomplishes edge alignment by first dividing down the VCO by the right multiple so that edge alignment is unambiguous, as pictured in Figure 6.5(b). Because the PFD compares phase over the entire reference cycle, a PFD cannot phase lock two inputs at different frequencies. In fact, it is precisely this property that makes the PFD popular.

Now consider using a PFD without a divider. Clearly, there would be an edge ambiguity problem, rendering the PFD quite ineffective as seen in Figure 6.5(c). The reason is that the PFD responds to every edge of the VCO, evidenced by the charge pump (CP) current's net negative value. This current erroneously commands the VCO to decrease its frequency. However, by restricting the time interval during which phase is examined, one may eliminate the edge ambiguity, and hence the frequency divider. The dashed boxes in Figure 6.5(d) define the window during which phase may be compared, even if the two inputs are of different frequency. The window can be controlled by the reference timebase since it periodically opens at that rate. Furthermore, the window need only be wide enough so that a VCO edge falls within it (i.e., the window's duration equals the instantaneous VCO period), however a wider window is easier to implement. No dividers are thus necessary to maintain phase lock, and this phase detector, called an aperture phase detector, can operate with two inputs that are at different frequencies, as shown in Figure 6.6.

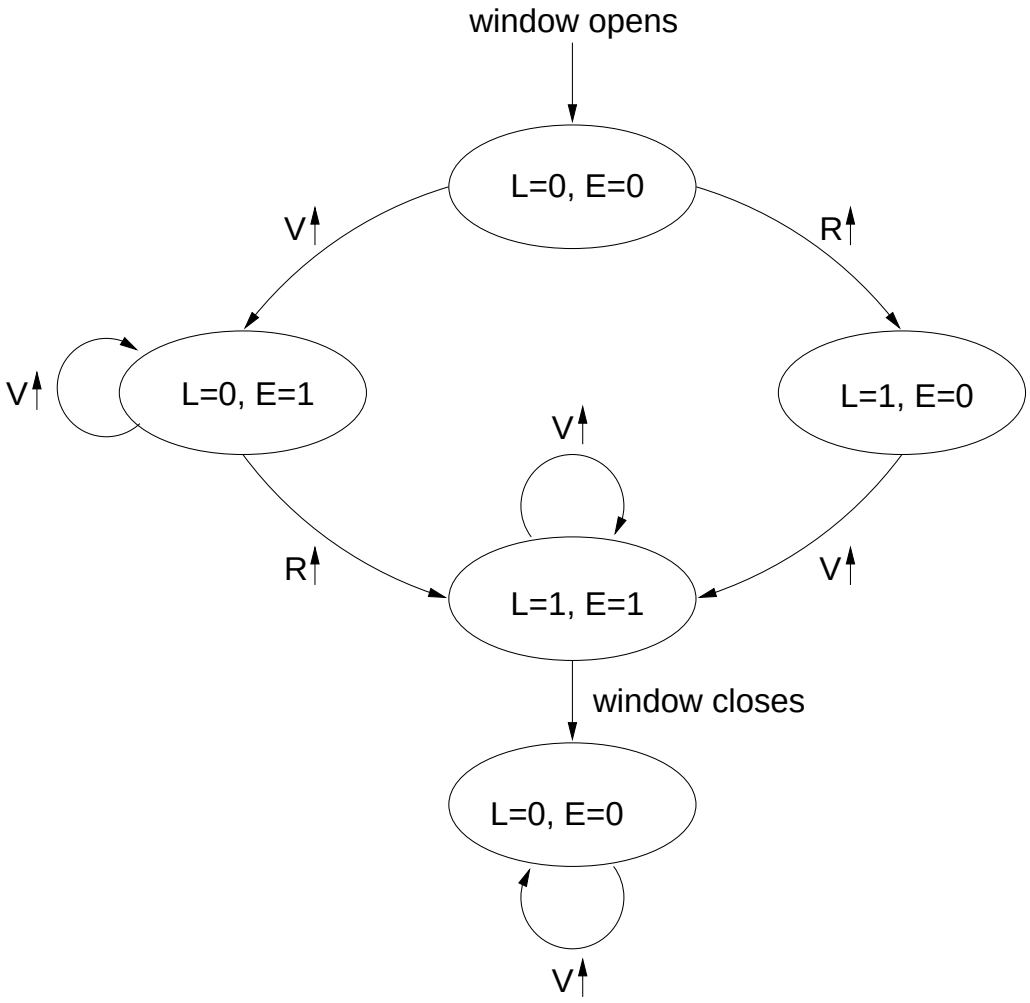


Figure 6.7: Ideal APD state diagram

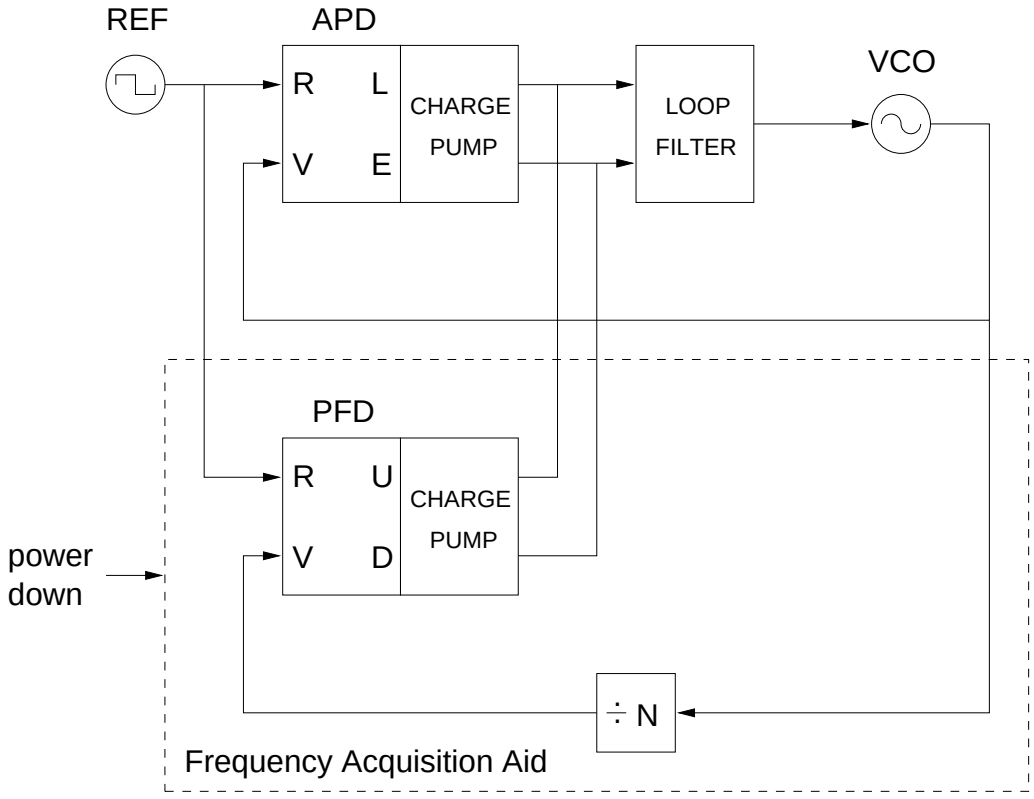


Figure 6.8: APD synthesizer architecture with integer-N FAA

A more substantive description of the APD's operation is provided in Figure 6.7, which illustrates the state diagram for an idealized APD. When the window opens, the phase detector becomes active. The R-input rising edge sets the L (denoting "late") terminal true, and the V-input rising edge sets the E (denoting "early") terminal true. Subsequent edges of the V-input are ignored until the next window opens. The time difference between the rising edges of the L and E signals is proportional to the phase error between the reference and VCO phases. If L is set first, the VCO phase is late, and conversely, if E is set first, the VCO phase is early. When the window closes, the L and E terminals are reset (to false).

A consequence of restricting phase comparisons to a window is that some type of frequency acquisition aid (FAA) is required to bring an APD based loop initially into lock. Figure 6.5(d) not only shows phase lock, but also that using windowing eliminates the phase detector's ability to perform frequency detection. This issue is discussed in further detail in §6.5. For this work, an external acquisition aid was

used for experimental purposes. Figure 6.8 shows an integrated implementation of the acquisition aid that uses the traditional divider with PFD to lock the loop, and then powers down the FAA, transferring control to the low-power APD. An APD can be used once in lock because the reference is derived from a stable crystal oscillator.

6.3 Loop Theory

Having provided an overview of the APD's operation, we now develop a linearized APD PLL model relating input and output phases. This model is important for quantitative loop design and ensures that the synthesized output has the desired stability and noise performance.

From the description of the late and early APD signals given in §6.2, the average charge pump current over one reference cycle is given by

$$i_d = I_p(t_v - t_r)\frac{\omega_r}{2\pi}, \tag{6.5}$$

where I_p is the magnitude of the charge pump current, t_v is the time of the first VCO rising edge in the window, t_r is the time of the reference rising edge in the window, and ω_r is the angular reference frequency. The current, i_d, can be expressed as a function of the reference and VCO phases by relating these phases to t_r and t_v, assuming small phase errors. Expressions relating edge time to signal phase are

$$t_r = -\frac{\theta_r}{\omega_r}, \tag{6.6}$$

where θ_r is the reference phase, and

$$t_v = -\frac{\theta_v}{\omega_v}, \tag{6.7}$$

where θ_v is the VCO phase and ω_v is the angular VCO frequency. Substituting (6.6) and (6.7) into (6.5) shows that

$$i_d = \frac{I_p}{2\pi}\left(\theta_r - \theta_v\frac{\omega_r}{\omega_v}\right). \tag{6.8}$$

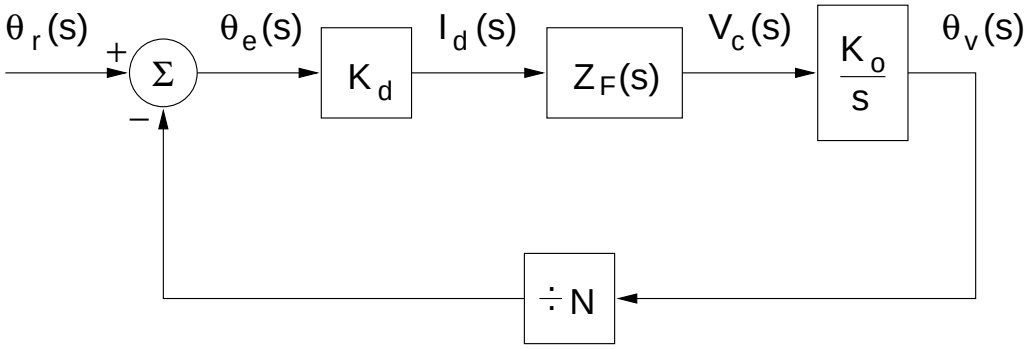


Figure 6.9: APD-PLL LTI block diagram

When the acquisition aid is powered down, after locking the loop to the desired integer ratio N, $\omega_v = N\omega_r$ for the APD, giving

$$i_d = \frac{I_p}{2\pi}\left(\theta_r - \frac{\theta_v}{N}\right) = \frac{I_p}{2\pi}\theta_e = K_d\theta_e, \tag{6.9}$$

where K_d is the phase detector gain constant. Note that even though there is no explicit divider in the loop, the VCO phase is divided by N in (6.9), just as in a conventional loop.

This model from (6.9) can be used in place of the APD in Figure 6.6 and the other blocks in the same figure can be replaced by their corresponding LTI models, yielding the overall system model shown in Figure 6.9. Figure 6.9 is a representation of the APD PLL in lock, from which the phase transfer function is readily found to be

$$H(s) = \frac{\theta_v}{\theta_r} = \frac{NK_dK_oZ_F(s)}{Ns + K_dK_oZ_F(s)}, \tag{6.10}$$

where K_o is the VCO gain constant, and $Z_F(s)$ is the loop filter's impedance, expressed in the s-domain.

6.4 APD+CP Characteristic (i_d vs. θ_e)

6.4.1 Ideal Edge-Detecting APD

The derivation in §6.3 treats the APD for small phase errors. For completeness, it is instructive to examine the response of the APD to arbitrary phase errors. Now, the delay between the time the window opens and the time at which the reference edge occurs becomes important. This delay is designated by d, which is a positive quantity whose least restrictive range is limited to $[0, T_r)$, where T_r is the reference period. However, the loop can lock if and only if d is in the interval $[0, T_v]$, where T_v is the VCO period. Otherwise, the first VCO edge within the window will always precede the reference edge.

From Figure 6.10 it is apparent that the characteristic will be periodic in VCO phase, because when the VCO waveform has moved one VCO period to the right, the situation is identical to the start. As the VCO waveform moves to the right, the time difference, $t_v - t_r$, varies proportionally with phase error, θ_e. Therefore, to find the APD+CP's characteristic, θ_e and i_d need to be calculated at only two points and the remainder of the characteristic is generated by connecting these endpoints. First, θ_e and i_d are calculated for $t_v - t_r = -d$:

$$\theta_e = -\omega_r d, \tag{6.11}$$

$$i_d = -\frac{I_p}{2\pi}\omega_r d. \tag{6.12}$$

Next, θ_e and i_d are calculated at the other extreme where $t_v - t_r = T_v - d$:

$$\theta_e = \omega_r(T_v - d), \tag{6.13}$$

$$i_d = \frac{I_p}{2\pi}\omega_r(T_v - d). \tag{6.14}$$

From this information, the portion of the APD+CP characteristic shown in Figure 6.11 can be constructed.

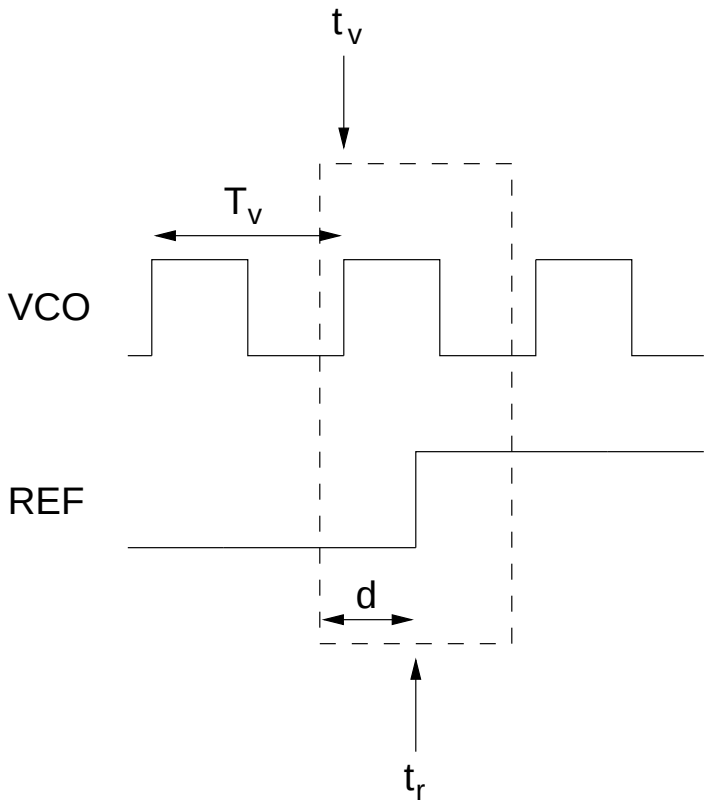


Figure 6.10: VCO and reference edge positions in window

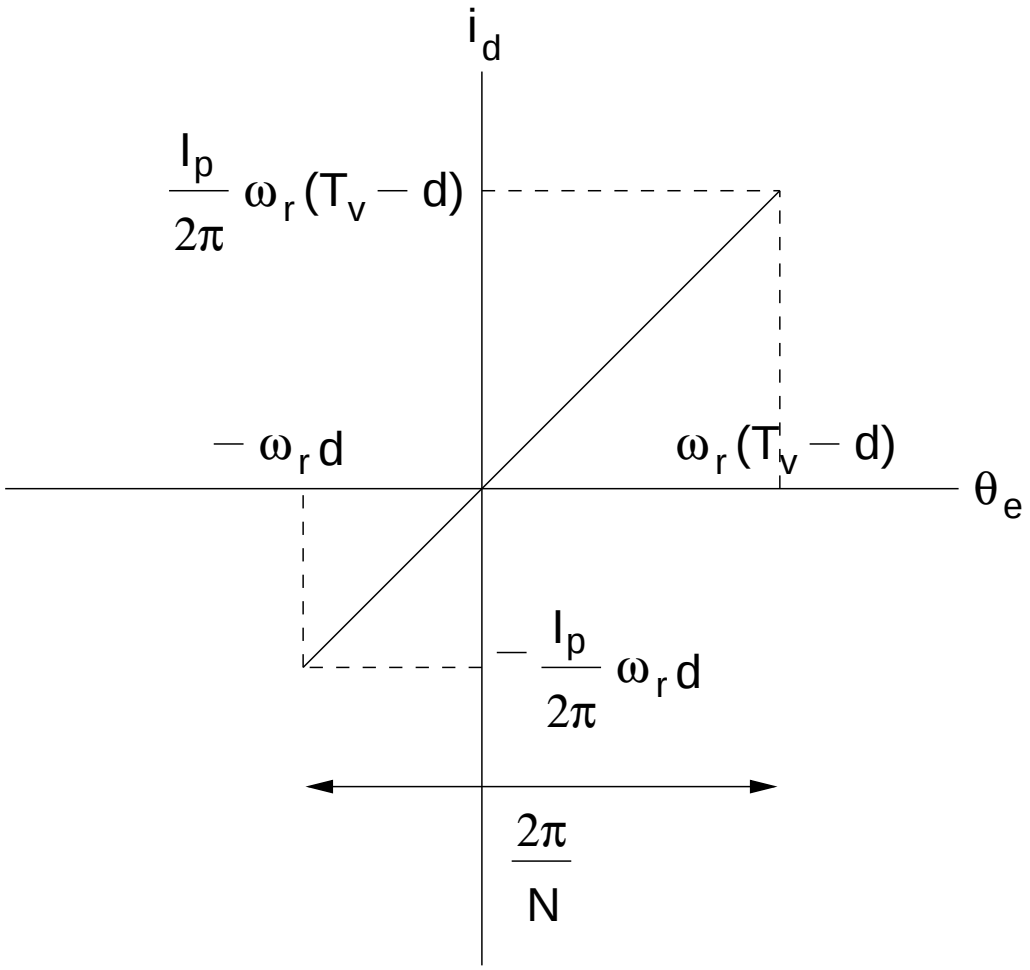


Figure 6.11: Ideal edge-detecting characteristic over $\frac{2\pi}{N}$ interval

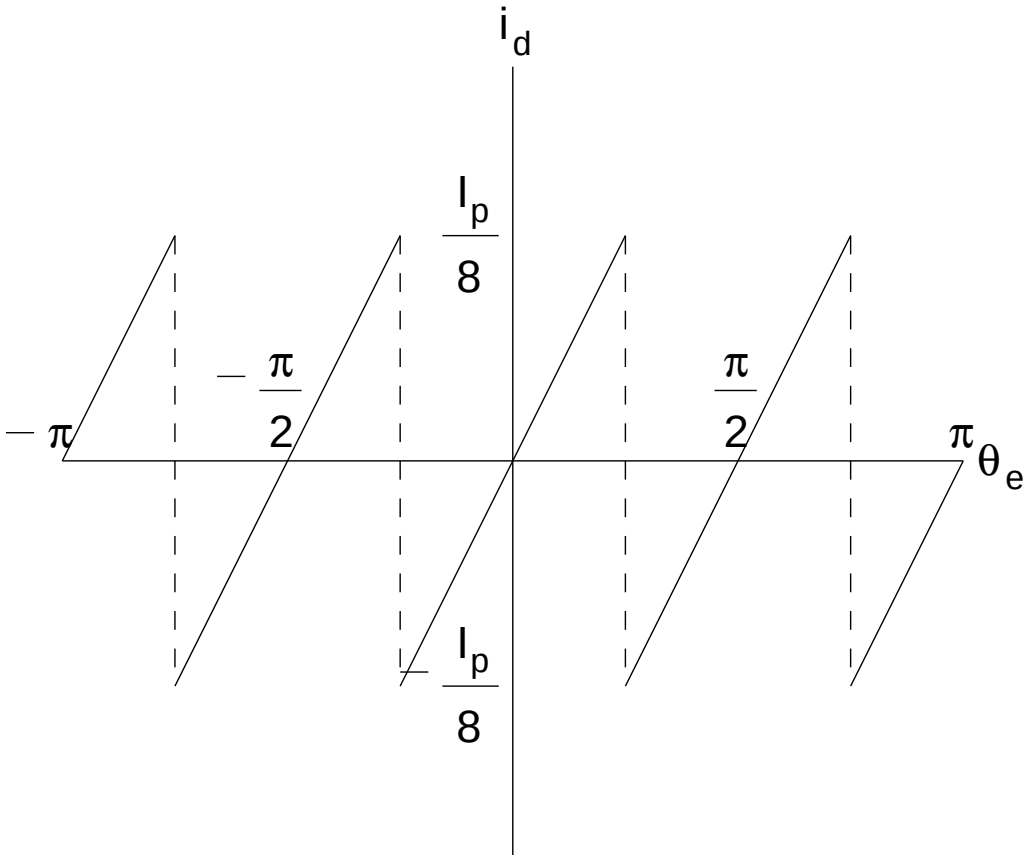


Figure 6.12: APD+CP characteristic for $d = \frac{T_v}{2}$ and $N = 4$

The influence of two parameters, d, the delay between the time the window opens and when the reference edge occurs, and N, the ratio between the VCO and reference frequencies, warrants special attention. Decreasing d shifts the characteristic diagonally up (along the line of the characteristic), and increasing d shifts the characteristic diagonally down. It is desirable to have d equal to half the synthesized frequency's period. By designing for this condition, the APD+CP characteristic will be centered about $i_d = 0$, to provide a symmetrical correction range. The parameter N affects the phase error's periodicity, with larger values increasing the periodicity. Figure 6.12 shows the complete APD+CP characteristic (a 2π variation in θ_e) for the specific case where $d = \frac{T_v}{2}$ and $N = 4$.

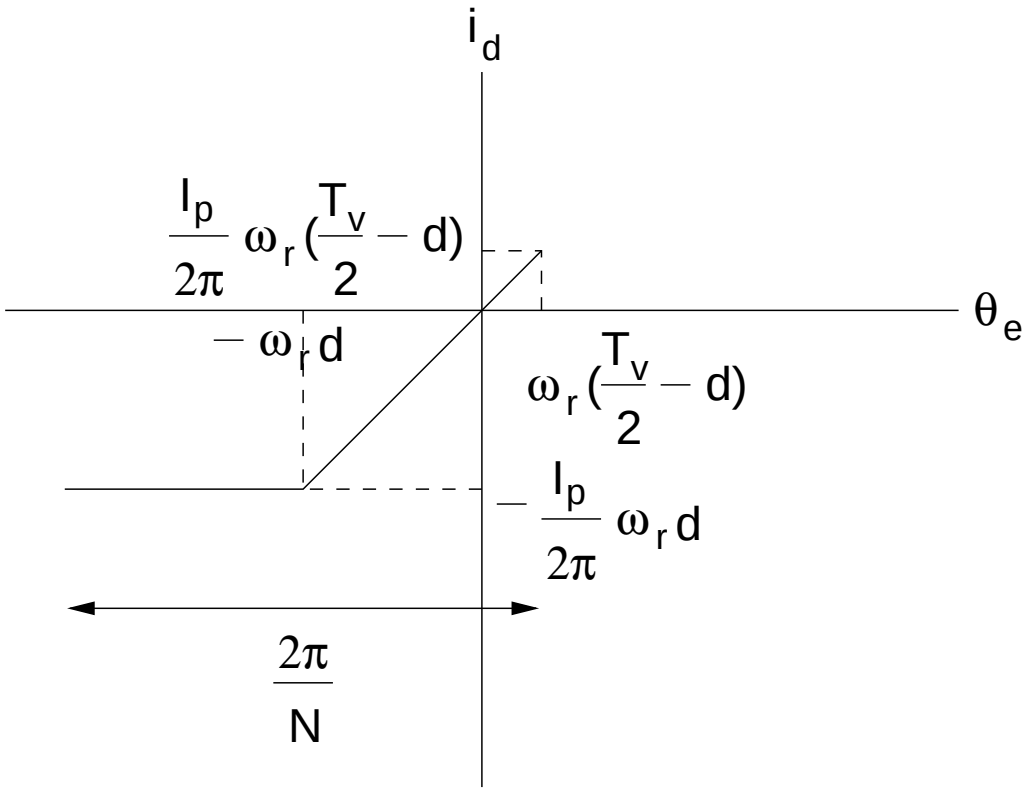


Figure 6.13: Ideal level-detecting characteristic over $\frac{2\pi}{N}$ interval

6.4.2 Ideal Level-Detecting APD

Implementation of the APD with edge detectors is preferred, but the APD can also be built around level detectors. Naturally, changing the detector to sense levels instead of edges has ramifications on the APD's behavior. This subsection addresses the differences.

The first difference is the range of values of d over which the loop can maintain lock. Because an ideal level detector triggers on sensing a high or low signal, the VCO's level detector must trigger by $\frac{T_v}{2}$ after the window opens. Thus, the loop can lock if and only if d is in the interval $[0, \frac{T_v}{2}]$.

The APD+CP characteristic is still periodic in VCO phase, but, as we will see, it is flat for half the VCO phase. For the present analysis, the definitions of t_v and t_r need to be modified slightly. t_v corresponds to the time of the first high VCO level in

the window and t_r corresponds to the first high reference level in the window. Thus, t_r still equals d. As the VCO waveform moves to the right, the time difference $t_v - t_r$ remains constant with phase error for $\frac{\pi}{N}$ radians (while the high VCO level crosses the left window boundary in Figure 6.10) and varies proportionally with phase error for another $\frac{\pi}{N}$ radians (while the low VCO level crosses the left window boundary in Figure 6.10). Again, to find the APD+CP's characteristic, θ_e and i_d need to be calculated at only two points and the remainder of the characteristic is generated by connecting these points to the right, and extending the first point horizontally to the left. First, θ_e and i_d are calculated when a VCO rising edge is positioned on the window's left boundary:

$$\theta_e = -\omega_r d, \tag{6.15}$$

$$i_d = -\frac{I_p}{2\pi}\omega_r d. \tag{6.16}$$

Next, θ_e and i_d are calculated when a VCO falling edge is positioned on the window's left boundary:

$$\theta_e = \omega_r(\frac{T_v}{2} - d), \tag{6.17}$$

$$i_d = \frac{I_p}{2\pi}\omega_r(\frac{T_v}{2} - d). \tag{6.18}$$

Figure 6.13 shows the portion of the APD+CP characteristic generated from this information. We can see that for symmetric maximum up and down correction currents, it is desirable to have $d = \frac{T_v}{4}$.

6.5 Subharmonic Lock Modes

During each window, which opens periodically at the reference rate, the APD makes a single phase comparison. It is this property that allows an APD to phase lock the VCO's output to an integer multiple of the reference input. But the ability to compare the phase of two signals with different frequencies introduces more modes than just the desired integer lock modes. Additional subharmonic modes occur if the net current delivered over multiple cycles of the reference is zero, allowing the loop to stay locked at an undesired frequency [31].

If we designate by M the number of reference cycles over which the net charge delivered to the loop filter is zero, then an expression relating the reference frequency to the VCO frequency in lock is $M\omega_v = N\omega_r$. Figure 6.14 displays the points on the APD characteristic between which the loop ping-pongs for the specific case where $M = 2$ and $N = 7$. Because $M = 2$, the charge pump alternates between pumping up on one cycle and pumping down on the next cycle, balancing the charge to the loop filter over two cycles.

These subharmonic lock modes are problematic because they are spaced, in frequency, closer than the neighboring integer modes. However, the APD favors integer over subharmonic modes for two reasons. First, the loop's bandwidth itself imposes a limit on M. If the number of reference cycles over which the charge pump current averages to zero grows too large, the loop acts on partial information, because the loop responds to signals averaged over a loop period (defined here as the reciprocal of the closed-loop bandwidth). Another reason the APD favors integer over subharmonic modes is that the subharmonic modes have a lower detector gain, K_d, because the VCO edge arrives at a different time in each of the M cycles. If the APD characteristic is nonlinear, then the overall detector gain is the average of the M individual linearized detector gains.

Using a frequency acquisition aid to ensure frequency lock eliminates the concern of locking in a subharmonic mode. Once lock has been achieved, and control transferred to the APD, the APD is capable of maintaining lock at the desired frequency.

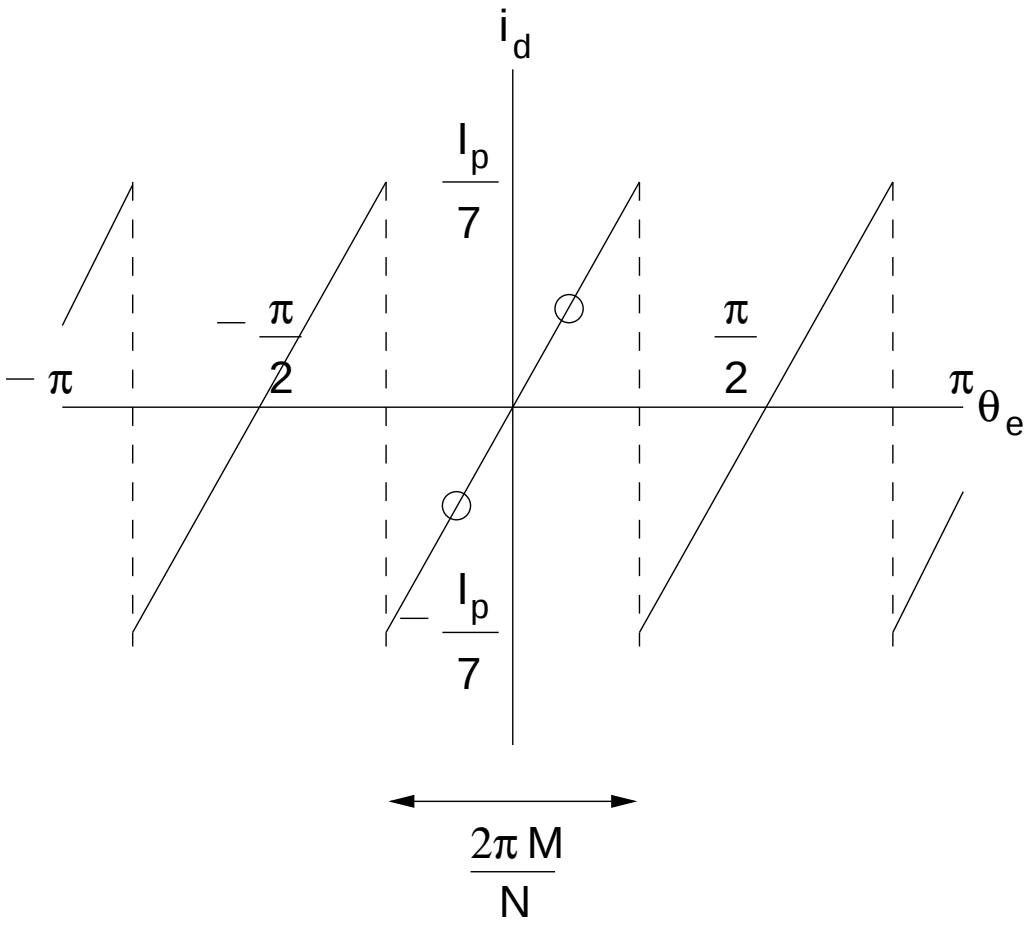


Figure 6.14: $M = 2$, $N = 7$ subharmonic lock mode

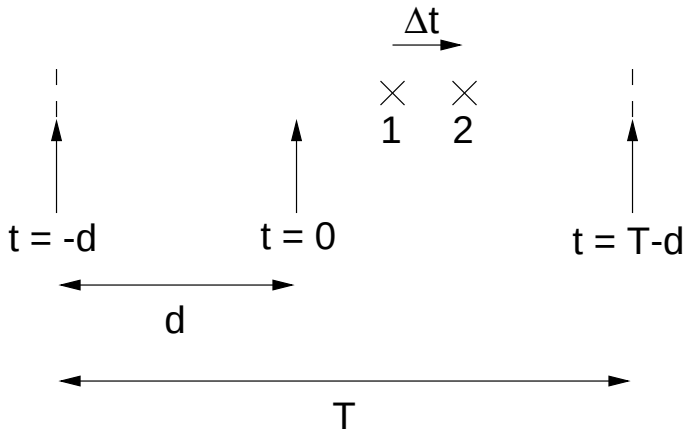


Figure 6.15: "Effective" APD window

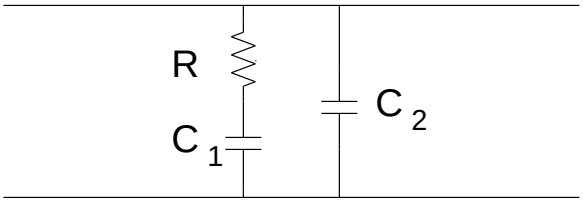


Figure 6.16: Common loop filter

6.6 Out-of-Lock Behavior

Analyzing the behavior of the APD when it is not in lock reveals a unique frequency-detection response. The procedure to determine the out-of-lock behavior is the following. The loop is broken at the VCO input and a fixed control voltage is applied to the VCO. The average charge pump current over one reference cycle changes each cycle when the VCO frequency is not harmonically related to the reference frequency. By determining the average charge pump current over many cycles of the reference (on the order of a loop period), we get an indication of the direction and rate of change to VCO frequency that the loop is trying to make.[2] Using this information in conjunction with the mathematical expression for a VCO yields the APD PLL's frequency-detection response.

[2]Having a fixed VCO frequency during this interval is valid, because macroscopic loop behavior occurs on a time scale commensurate with the loop period.

To develop this open-loop analysis, it is instructive to begin with Figure 6.15, which is a picture of the "effective" APD window. The window's effective time duration is one VCO period, since a VCO edge must arrive within one period of the window's opening. The time $t = 0$ is selected to coincide with the reference rising edge, which always occurs a fixed time d after the window opens. If we designate the positions of the VCO rising edges with Xs, then after a loop period the Xs appear uniformly distributed throughout the window. To see this edge distribution, Figure 6.15 shows the position of the first VCO rising edge observed in this analysis (the X with a 1 under it). The second edge occurs at a different location (assuming the VCO frequency is not an integer multiple of the reference frequency), a distance Δt from the first. Subsequent edges propagate with a separation of Δt, rotating through the window, to the point where it appears the edges are evenly distributed in the window. Therefore, the average time difference between a VCO edge and a reference edge is

$$\overline{t_v - t_r} = \frac{T_v}{2} - d. \tag{6.19}$$

From (6.5), the average charge pump current is

$$\overline{i_d} = \frac{I_p \omega_r}{2\pi} \left(\frac{T_v}{2} - d \right). \tag{6.20}$$

Now, assuming a typical 2nd-order loop filter as shown in Figure 6.16, the average control line voltage is

$$\overline{v_c} = \frac{1}{C_1 + C_2} \int \overline{i_d} dt + const = \frac{I_p \omega_r}{2\pi (C_1 + C_2)} \left(\frac{T_v}{2} - d \right) \int dt + const. \tag{6.21}$$

Returning to the VCO, its instantaneous output frequency is

$$f_v(t) = \frac{\Omega_0}{2\pi} + \frac{K_o}{2\pi} \Delta v_c(t), \tag{6.22}$$

which represents a linearized VCO characteristic. Differentiating (6.22) to get an expression that describes the rate of change of the VCO's output frequency gives

$$\frac{df_v}{dt} = \frac{K_o}{2\pi}\frac{d\Delta v_c}{dt}. \tag{6.23}$$

If we close the loop from the open-loop condition, the VCO's control voltage begins to move, affecting the VCO's output frequency. Since the dynamics of this process are governed by the loop filter, (6.21) can be differentiated and that result used for $\frac{d\Delta v_c}{dt}$ [35]. Replacing $\frac{d\Delta v_c}{dt}$ with $\frac{dv_c}{dt}$ in (6.23) yields the following differential equation describing the APD PLL's frequency response:

$$\frac{df_v}{dt} = \frac{K_o I_p \omega_r}{(2\pi)^2(C_1 + C_2)}\left(\frac{1}{2f_v} - d\right). \tag{6.24}$$

We now consider analytical and graphical solutions to (6.24).

6.6.1 Analytic Solution

Separating variables and integrating (6.24) results in a transcendental equation:

$$\frac{-f_v}{d} - \frac{1}{2d^2}\ln|1 - 2df_v| = \frac{K_o I_p \omega_r}{(2\pi)^2(C_1 + C_2)}t + const, \tag{6.25}$$

Notice that the argument of the natural log term becomes zero when $f_v = \frac{1}{2d}$. It turns out that $f_v = \frac{1}{2d}$ is the steady-state solution of (6.24), and it is therefore clear that the natural log term should dominate towards the end of the frequency-sweep process.

Thus, to proceed further with the analytical solution, the linear term in (6.25) is dropped, allowing f_v to be solved in terms of t

$$f_v(t) = \frac{1}{2d} + (f_v(0) - \frac{1}{2d})e^{-\frac{K_o I_p \omega_r d^2}{2\pi^2(C_1+C_2)}t}, \tag{6.26}$$

where $f_v(0)$ is the initial instantaneous VCO frequency. The time constant of the exponential in (6.26),

$$\tau_a = \frac{2\pi^2(C_1 + C_2)}{K_o I_p \omega_r d^2}, \tag{6.27}$$

allows us to estimate roughly the amount of time the loop takes to reach $f_v = \frac{1}{2d}$, i.e. $\approx$ 4–5 time constants, during its sweep.

In §6.6.2, we will graphically compare the exact solution to this approximate solution. We shall see that the approximate solution is a good estimate of the real solution, and is thus a useful engineering approximation.

6.6.2 Graphical Solution

The solution to (6.24) with specific parameter values (see Table 6.1) and initial condition is presented in Figure 6.17, along with (6.26). Both curves settle to the same oscillation frequency of $\frac{1}{2d}$. From (6.20), it is clear that $\overline{i_d} = 0$ if $T_v = 2d$. Intuitively, if there is no average current delivered to the loop filter, the control voltage does not change and the VCO's output frequency remains constant. This frequency, $f_v = \frac{1}{2d}$, is thus the steady-state solution of (6.24). For $f_v > \frac{1}{2d}$, the VCO's frequency decreases towards $f_v = \frac{1}{2d}$; for $f_v < \frac{1}{2d}$, the VCO's frequency increases towards $f_v = \frac{1}{2d}$.

It is important to realize that the steady-state frequency to which the APD PLL sweeps, $\frac{1}{2d}$, may not be the same as the desired lock frequency, $N\omega_r$. It is difficult to design the APD to have $\frac{1}{2d} = N\omega_r$ in general. Also, as frequency sweeps, the loop passes through other potential lock modes (both integer and subharmonic), and there is always some chance that the loop may enter one of these modes. This uncertainty reinforces the need for an FAA to ensure that the loop reaches the desired lock frequency.

Table 6.1: Parameter values and initial condition for Figure 6.17

f_r	100MHz
d	300ps
I_p	100μA
K_o	$2\pi(1.2 * 10^9)$rad/s/V
$C_1 + C_2$	20pF

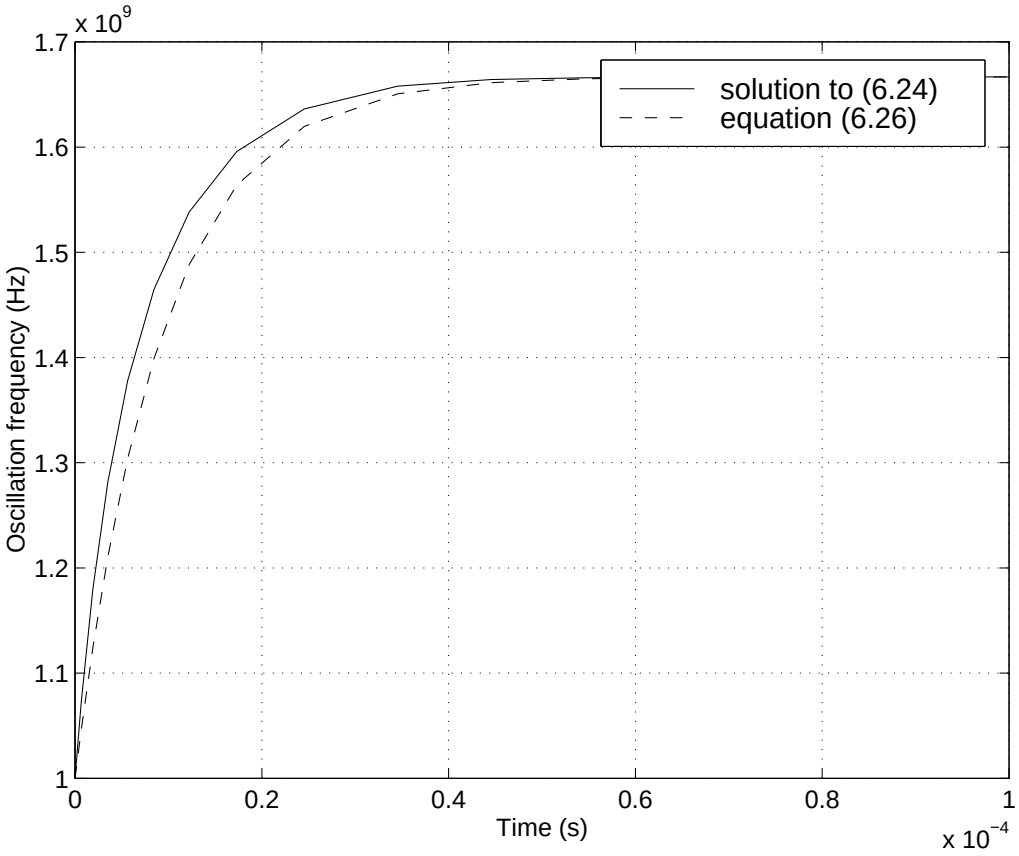


Figure 6.17: APD-PLL frequency sweep

6.7 Summary

In this chapter, we have presented a new technique for frequency synthesis that uses an aperture phase detector. It is interesting to discover that even though a divider is not present in the feedback path, the LTI system in lock is identical to that of a loop with a divider. This architecture trades off improved power consumption and reduced interference for added complexity in frequency acquisition. The requirement of a frequency acquisition aid follows from the analysis which reveals the loop's ability to lock in subharmonic modes.

A final point is that, if an FAA is present, it can be made programmable to allow synthesis of multiple output frequencies, since the APD can maintain phase lock to any integer multiple of the reference frequency. Again, the FAA only needs to be active when switching between integer lock modes.

Chapter 7

An Experimental APD PLL

THE new loop architecture described in Chapter 6 was fabricated in a 0.5μm single-poly, triple-metal, CMOS process that has high-density capacitors and salicided-poly resistors. The APD PLL is the frequency synthesizer for the GPS radio shown in Figure 7.1, and as such, its frequency plan was tailored to the radio's low-IF architecture. Eliminating the need for a divider in the PLL allows the synthesis of a 1.573GHz output using only 36mW of power in this technology.

Whereas Chapter 6 focused on theory, this chapter focuses on implementation, on relevant simulations for loop characterization, and application of the existing body of knowledge for loop design. Lastly, we conclude with a presentation of measurement results.

7.1 Design

7.1.1 Aperture Phase Detector

Figure 7.2 shows a circuit implementation of an APD. The reference clock is shaped by the structure preceding the delay to have fast falling edges, since these are the edges that enable the precharged gates. When the reference input is low, M_1 is off and M_2 is on, causing the output to be high. After the reference rises, M_2 shuts off before M_1 turns on due to the two inverter delays. Therefore, M_1 does not fight

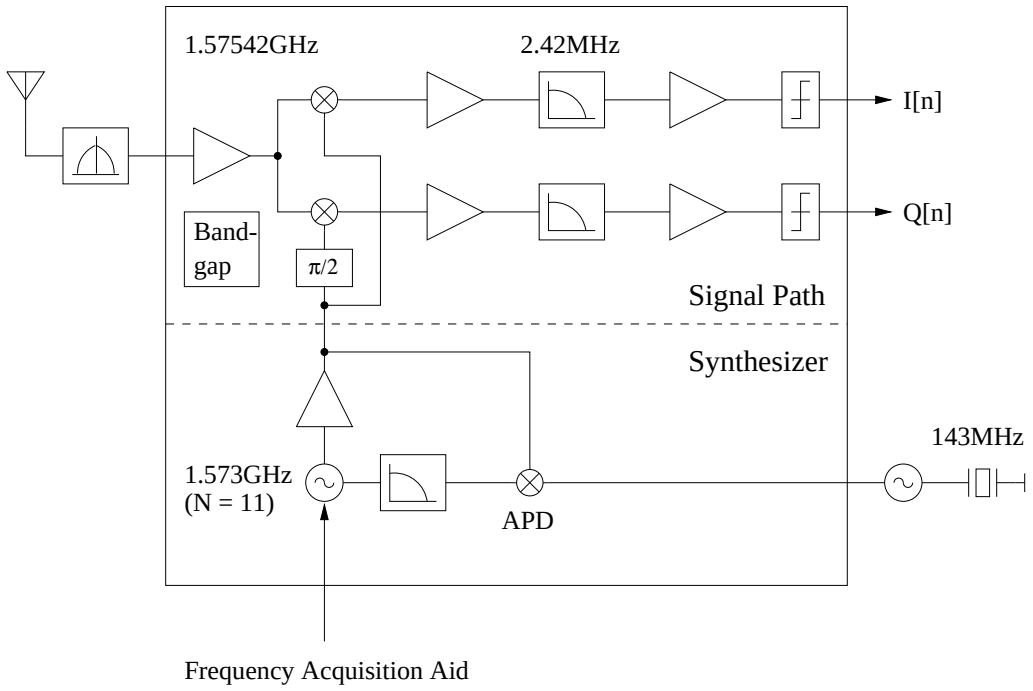


Figure 7.1: GPS radio architecture

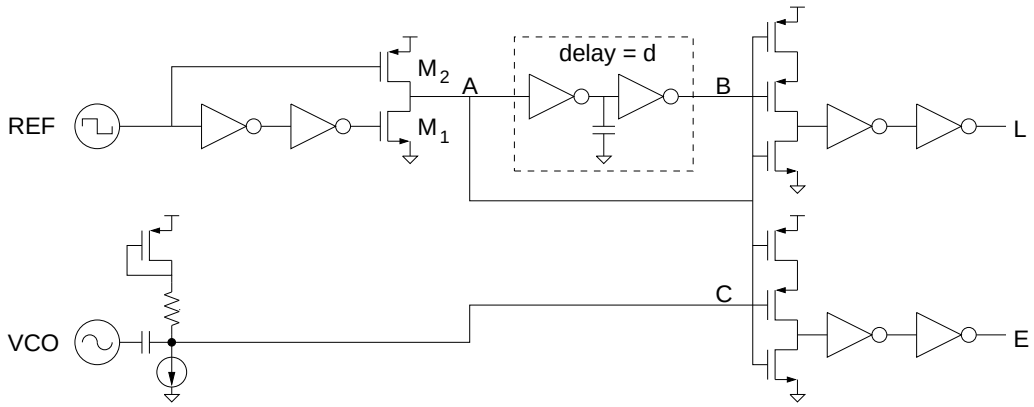


Figure 7.2: APD circuit diagram

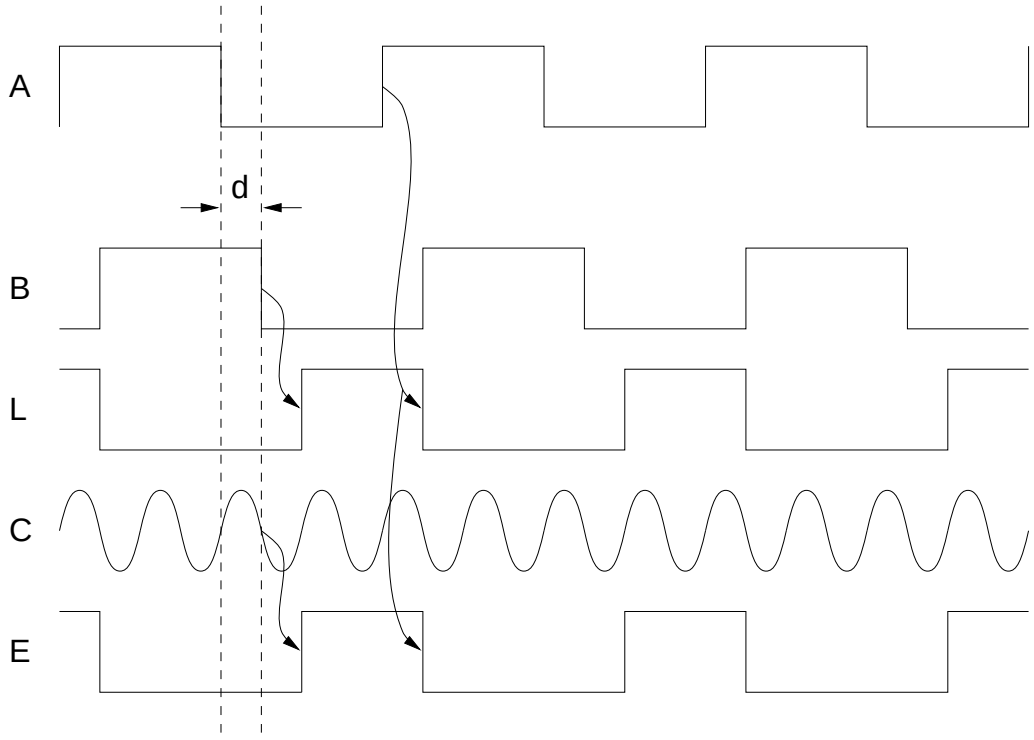


Figure 7.3: APD timing diagram

M_2 to pull the output low, and creates a fast falling edge which opens the window. The delay, d, between the opening of the window and the location of the reference edge is determined by the subcircuit shown within the dashed box in Figure 7.2. The APD uses two precharged gates to evaluate the reference and VCO phases. An advantage of such gates is that they respond only once when activated, which is ideal for examining the VCO phase. In this case, the gates are precharged low, and rise on detection of low levels (a better implementation would use a precharged flipflop [36] to enable edge detection). Also, the precharge action does not affect the loop to first order, because the state L=1, E=1 has the same action as the state L=0, E=0. That is, for both these cases, the charge pump does not deliver any current to the loop filter.

To further understand this circuit, Figure 7.3 displays the APD's operational waveforms at the nodes labeled A, B, C, L, and E in Figure 7.2. The waveform at node A is the windowing signal that indicates when the precharge gates are active. In this case, the gates precharge when A is high and are active when A is low. The waveform labeled B is just a delayed version of A, and feeds into Figure 7.2's top precharge gate. Because the gates respond to low levels, the falling B edge causes a rising L edge a brief time later. The action between the C and E signals is similar to the effect B has on L. The C signal is the VCO output, which is a full-swing sinusoid. The first low period of C after the bottom precharge gate is activated causes E to rise. Both L and E are reset at the same time when A falls, initiating the precharge mode. Notice that in Figure 7.3, we see a phase lock condition for a VCO signal at four times the reference frequency.

It is important to identify the differences between this circuit implementation and the ideal edge-detecting APD circuit, discussed in §6.4.1. One difference is that the circuit implementation responds to falling edges instead of rising edges. Changing the type of edge to which the phase detector responds does not alter it from being ideal. But if we look at the precise operation of the APD in Figure 7.2, we see that the precharged gates act as level detectors instead of as edge detectors, which is less desirable.

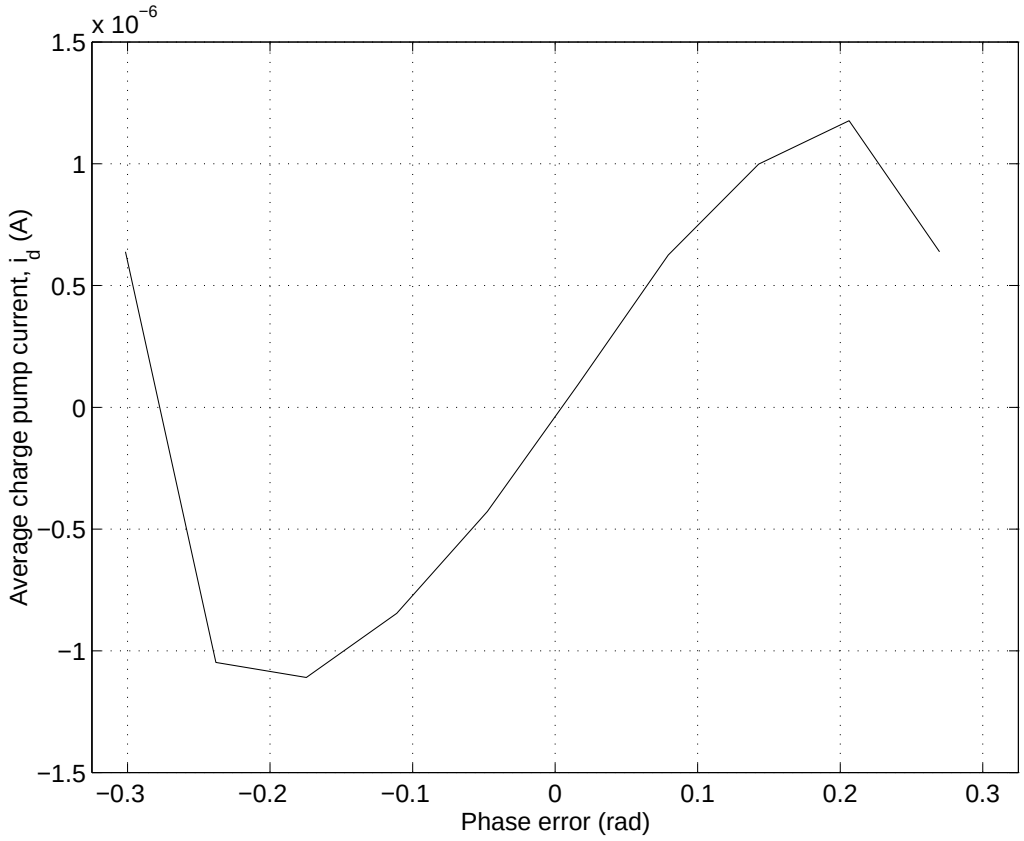


Figure 7.4: Simulated APD+CP characteristic for circuit of Figure 7.2

A simulation of the APD+CP characteristic over a $\frac{2\pi}{N}$ interval is shown in Figure 7.4, where $N = 11$. The phase error, θ_e, is plotted against the average charge-pump current over one cycle, i_d, and includes the nonidealities of the charge pump as well. The flat section near –0.2 radians is where the E signal driving the charge pump is compressing due to the level-detection nature of the precharged gates. Another imperfection in this circuit's APD characteristic is the section with finite negative slope, instead of a discontinuity. From the characteristic, the phase detector's gain constant, K_d, in a state of zero static phase error ($\theta_e = 0$) is evaluated to be 7.4μA/rad.

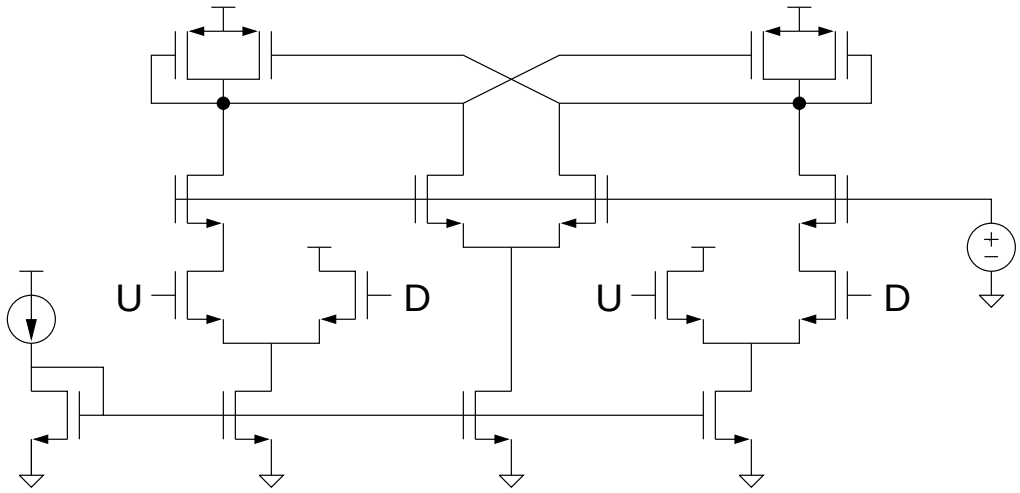


Figure 7.5: Charge-pump circuit diagram

7.1.2 Charge Pump

Figure 7.5 is a schematic of the loop's differential charge pump. The current-source loads employ a static pull down that determines the common-mode output level (see Figure 7.6). The pump circuit proper is shown in Figure 7.7.

An important design requirement for locking is to ensure that the pump current can overcome any static offset currents.[1] The source of the unwanted currents is mismatch in the bias circuitry. A mathematical expression of this requirement is

$$min(max(i_d), max(-i_d)) \geq I_s, \tag{7.1}$$

where I_s is the static offset-current magnitude. The expression on the left points out that since I_s is random, the worst case scenario needs to be accommodated. In this worst case, the charge pump uses its smaller extreme (see Figure 7.4) to balance I_s. Appendix C quantifies I_s, which determines the minimum pump current.

We now focus on the differential-pair structure on the bottom of Figure 7.7. The pump circuit pulls current from the loop filter if and only if D is high and U is low.

[1] It is possible that this criterion is automatically met when the PLL is designed to meet specified levels for its reference spurs.

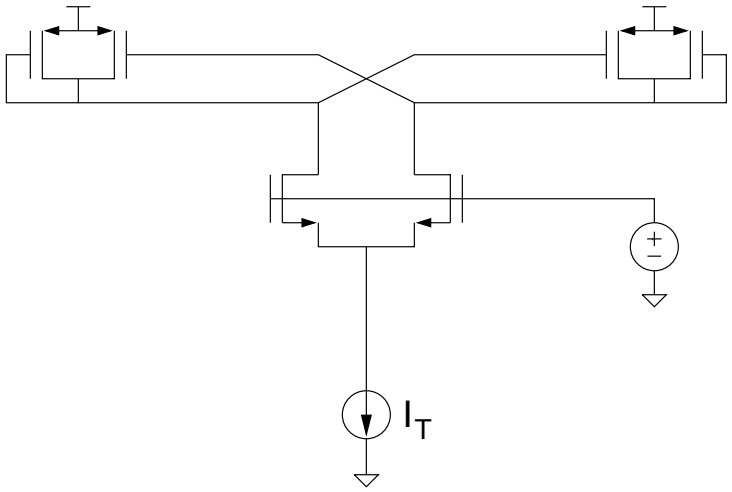


Figure 7.6: Charge-pump bias circuitry

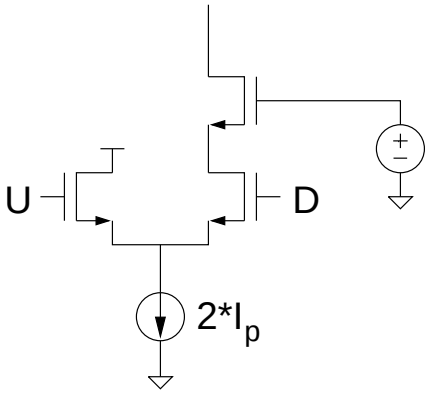


Figure 7.7: Pump circuit

Anytime U is high, current is drawn instead from the supply. The net result is that the loop-filter voltage is decreased in proportion to the time that D is high while U is low. The complementary pump circuit, with U and D reversed, is seen in Figure 7.5.

A cascode device included in the pump current's path to the loop filter provides two benefits. First, it eliminates feedthrough of the U and D transitions to the output. On a low-to-high transition, the cascode device is off, so an increase in the cascode's source voltage has no effect. Also, the high-to-low transitions of U and D occur simultaneously, so the balanced nature of the charge pump causes the differential output to remain unchanged. Second, the cascode device reduces loop-filter leakage by increasing output resistance. It allows lower leakage without having to increase the lengths of the differential-pair devices, which would increase the capacitance of the U and D inputs.

A worthwhile pump-circuit simulation, which indicates how close to ideal it behaves, is to keep U fixed and sweep the D input-signal low-to-high transition time, t_D. This simulation yields the total charge delivered to the loop filter versus the time difference between the D and U transitions, $t_D - t_U$. Ideally, this plot would be a straight line with a slope equal to the charge pump current. We performed this simulation sweeping t_D from -500ps to 50ps relative to t_U. The D transition extends to 50ps past the U transition because non-negligible charge can affect the loop filter until this point.

Figure 7.8 displays the results of this simulation, accounting for the operation of both pump circuits. Note that the slope is not constant, and increases by nearly a factor of two for small separations between U and D. The increase in slope is due to an overpull effect present in the pump circuit of Figure 7.7. When D rises, there is significant charge pulled to bring the differential pair's source voltage to its steady-state value, in addition to the charge ordinarily pulled from the loop filter. For larger separations between U and D, the overpull leads to a fixed offset in the amount of charged delivered, but does not affect the slope. The increase in slope near $t_D - t_U = 0$ translates into a higher K_d, which can be twice what is expected from equation (6.8).

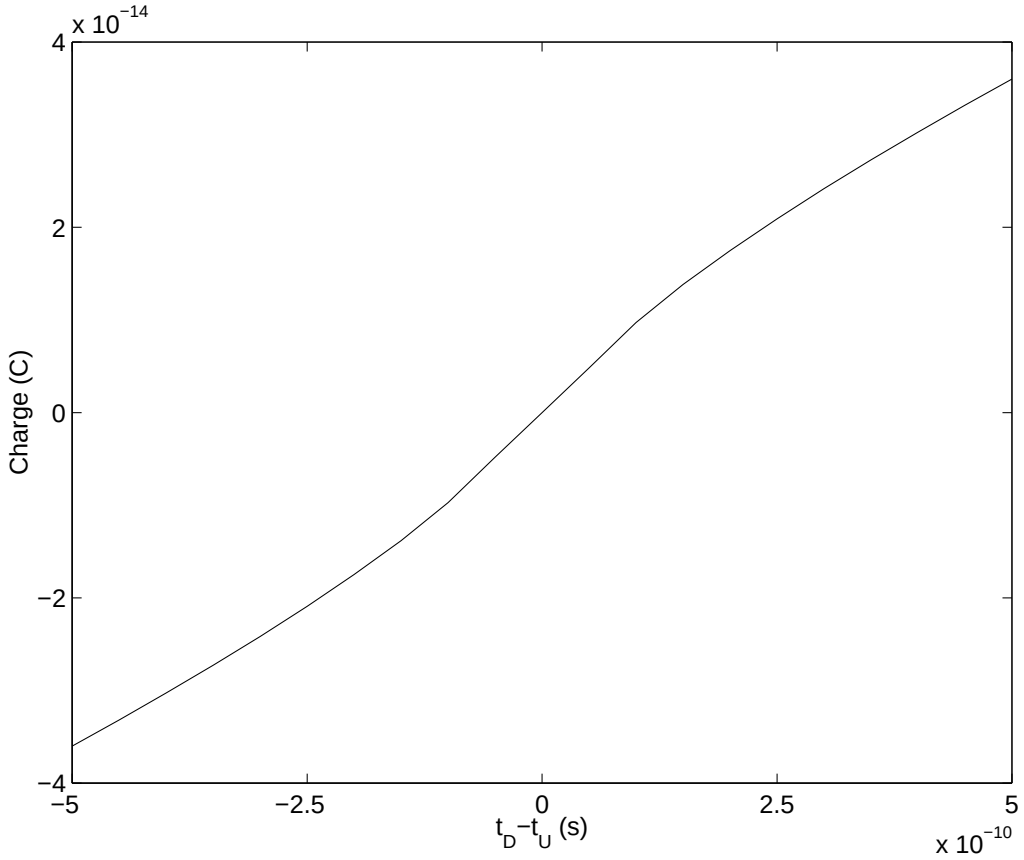


Figure 7.8: Charge delivered vs. $t_D - t_U$

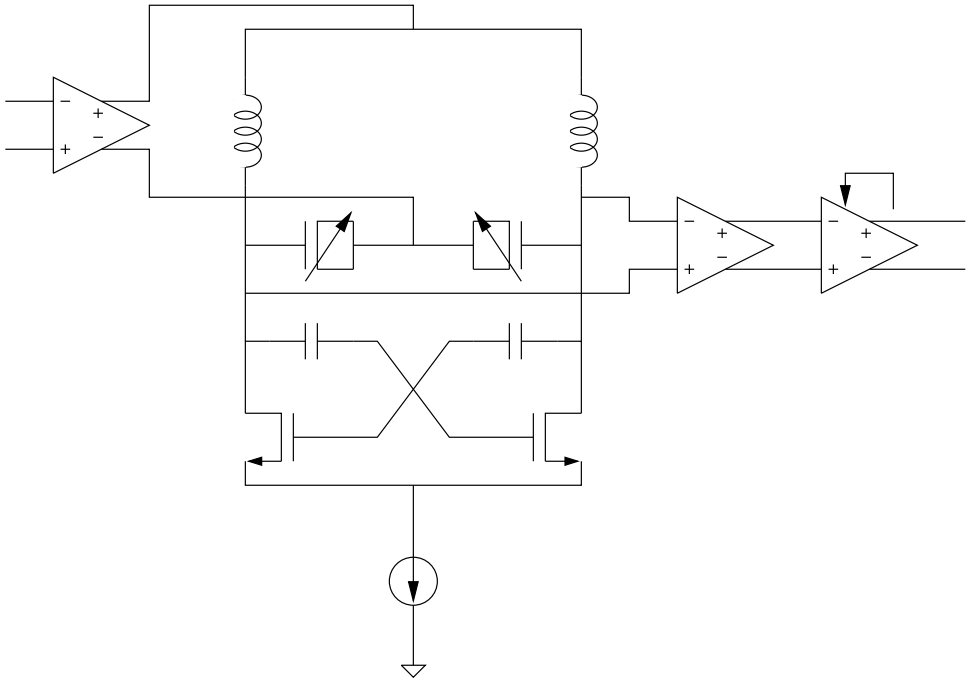


Figure 7.9: VCO circuit diagram

7.1.3 Voltage-Controlled Oscillator

A fully integrated VCO was designed for the PLL. It can be divided into four parts as shown in Figure 7.9: a preamplifier, the oscillator core, a postbuffer, and another postbuffer with automatic gain control (AGC).

The VCO architecture is built around a cross-coupled NMOS pair that has inductive loads. The inductor design is a key part of the VCO design, and techniques to improve on-chip spiral inductors are used. An example of one technique is the placement of a patterned ground shield below the spiral inductor [37]. The inductors have an inductance of 6.9nH with $\omega L/R$ equal to 8.5 at 1.573GHz. The parasitic capacitance between an inductor and its patterned ground shield is absorbed as part of the LC tank capacitance [38]. Simulations indicate that the effective parallel resistance of the tank is 570Ω, yielding an effective Q of 8 at 1.573GHz.

The core's bias current is provided through the preamplifier, which contains a feedback loop that servos a current source connected to the core. By comparing

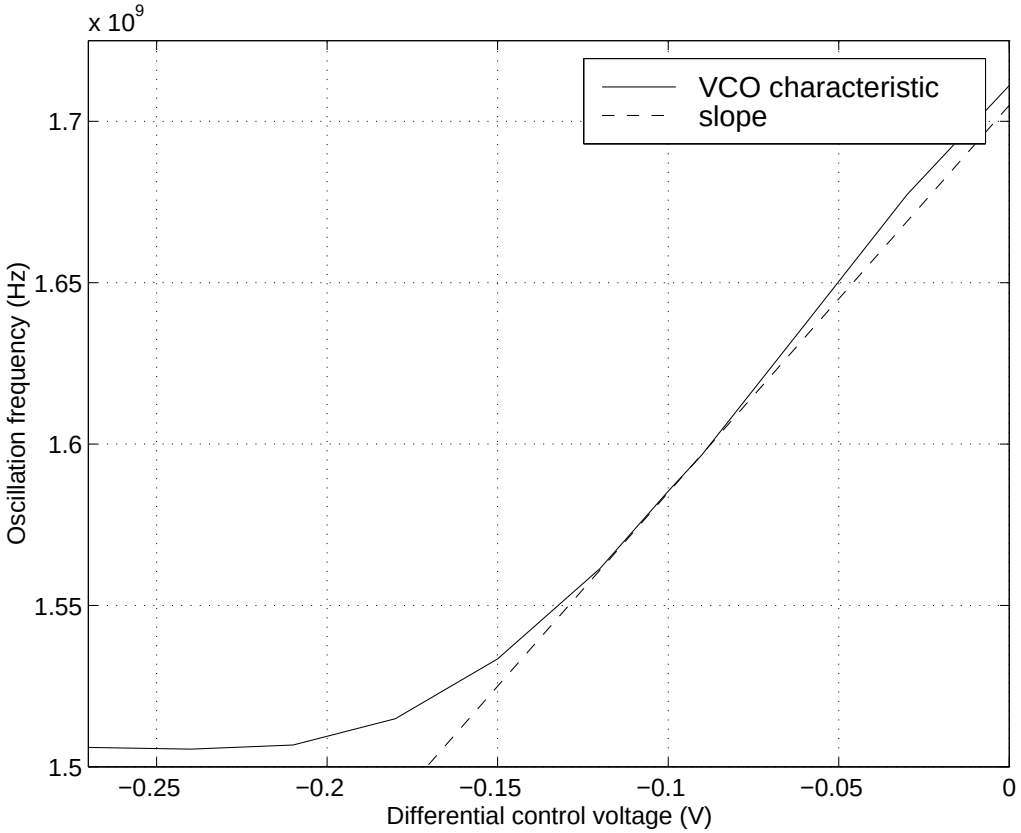


Figure 7.10: Simulated VCO characteristic for circuit of Figure 7.9

the common-mode voltage at the preamplifier's output to a reference, the correct current is delivered to the oscillator. This bias scheme also allows the differential output of the preamplifier to be used as the VCO's voltage control mechanism.

The core's tuning is achieved with the use of n-type accumulation-mode capacitors to keep Q high [39]. Models for n-type accumulation-mode capacitors were not available, so PMOS transistor models were used in the VCO and APD-PLL simulations. Figure 7.10 shows the simulated VCO characteristic from which the VCO's gain constant, K_o, at an oscillation frequency of 1.57542GHz is calculated to be $2\pi(1.2 * 10^9)$ rad/s/V.

7.1.4 Phase-Locked Loop

Section 6.3 developed a general model for a locked APD PLL, expressing the closed-loop phase transfer function in terms of the loop filter's s-domain impedance and an idealized VCO. We now provide specific expressions for the loop as actually implemented.

The loop filter used is the conventional network shown in Figure 6.16, whose s-domain impedance is

$$Z_F(s) = \frac{1 + sRC_1}{(C_1 + C_2)s\left(1 + sR\frac{C_1C_2}{C_1+C_2}\right)}. \tag{7.2}$$

For the VCO, a single-pole amplifier interfaces with the VCO's varactor, thus the ideal VCO transfer function, $\frac{K_o}{s}$, must be modified to

$$\frac{K_o}{s\left(1 + \frac{s}{2\pi f_{3dB}}\right)}, \tag{7.3}$$

where f_{3dB} is the 3dB bandwidth of the VCO's preamplifier. Using (7.2) and (7.3) in (6.10) enables us to write the complete phase transfer function of the implemented APD PLL as

$$H(s) = \frac{NK_dK_o(1 + sRC_1)}{N(C_1 + C_2)s^2\left(1 + \frac{s}{2\pi f_{3dB}}\right)\left(1 + sR\frac{C_1C_2}{C_1+C_2}\right) + K_dK_o(1 + sRC_1)} \tag{7.4}$$

In §7.2, we will compare measured data to (7.4).

Table 7.1 shows the designed values of the seven loop parameters. The loop's reference frequency is 143MHz, from which a 1.573GHz signal is synthesized, so that N is 11. The phase detector's gain constant, K_d, is taken from Figure 7.4 for zero static phase error ($\theta_e = 0$). The VCO's gain constant, K_o, is taken from Figure 7.10, where the slope is found for an oscillation frequency of 1.573GHz. And finally, the loop filter parameters, R, C_1, and C_2, are set to their designed values and f_{3dB} is calculated from the technology data.

Table 7.1: Designed loop-parameter values

N *	11
K_d	7.4μA/rad
K_o	$2\pi(1.2 * 10^9)$rad/s/V
R	9kΩ
C_1	15.4pF
C_2	1.2pF
f_{3dB}	15MHz

* $f_v = 1.573$GHz, $f_r = 143$MHz.

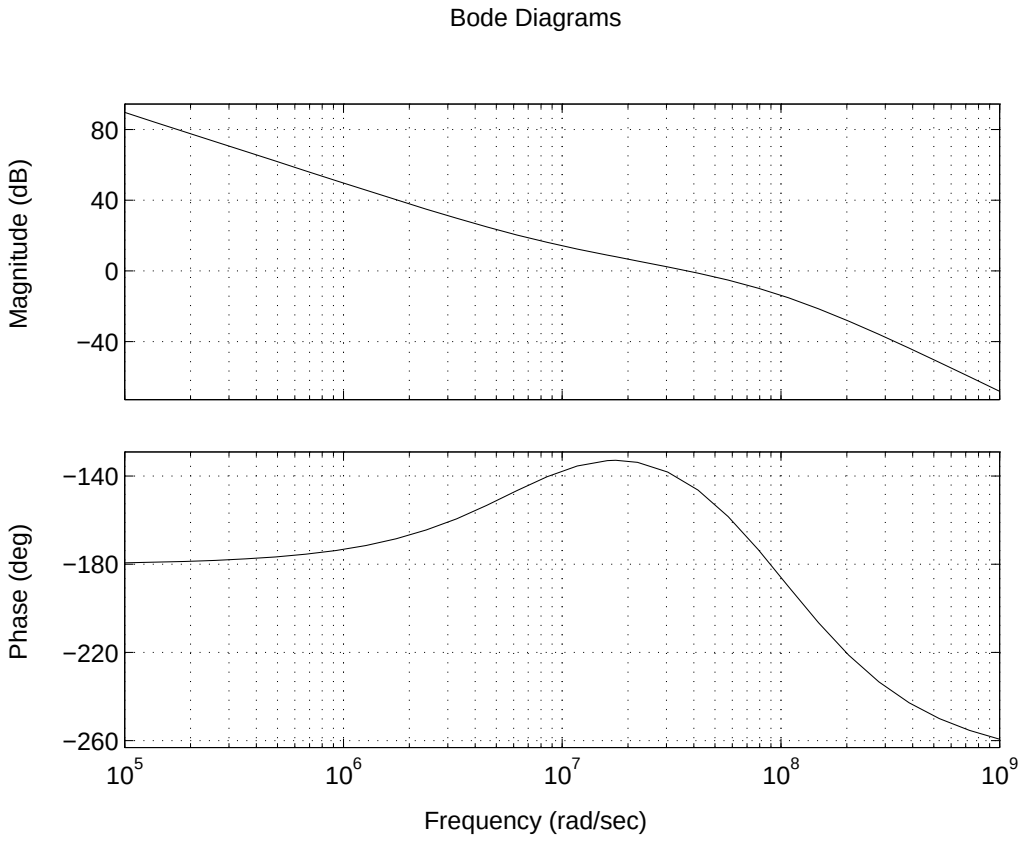


Figure 7.11: Loop-transmission magnitude and phase vs. frequency

Using the design values in Table 7.1, Bode plots of the loop transmission, $|\Gamma(j\omega)|$ and $\angle\Gamma(j\omega)$, are generated and displayed in Figure 7.11. They indicate a design loop bandwidth of 5.9MHz (the frequency at which the loop-transmission magnitude crosses 0dB), and a design phase margin of 37° (the loop-transmission phase at the loop bandwidth).

7.2 Experimental Results

A test chip containing a copy of the GPS-radio APD PLL, shown in Figure 7.12, is used for evaluation. On the same die, there is also a copy of the PLL's internal VCO, shown in Figure 7.13. Three separate tests are performed: one to characterize the VCO, one to verify the APD PLL's derived closed-loop transfer function, (7.4), and one to observe the synthesized LO spectrum for the GPS radio. In the last test, the synthesized LO is also checked with a microwave frequency counter to verify its long-term stability.

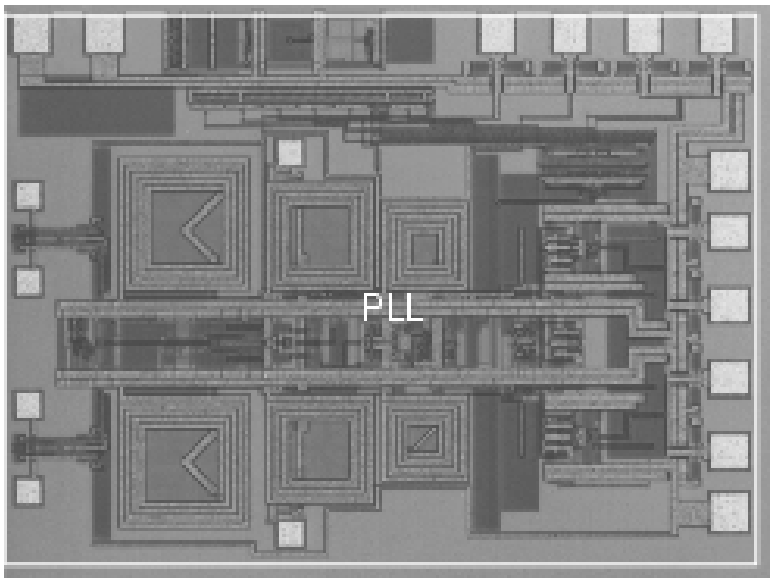


Figure 7.12: GPS-radio test-chip die photo showing APD PLL

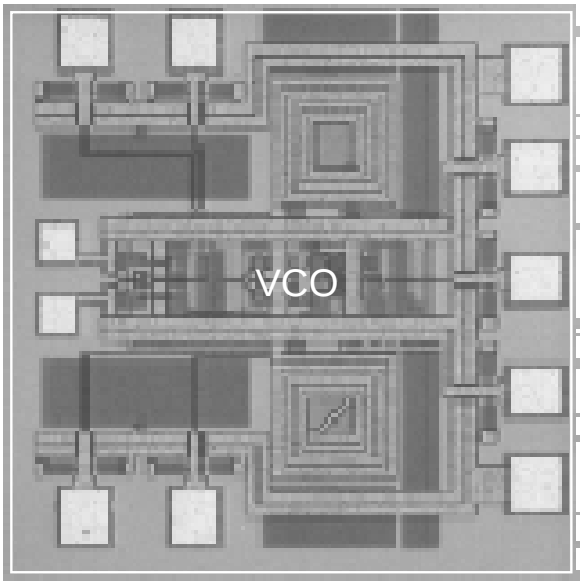


Figure 7.13: GPS-radio test-chip die photo showing VCO

7.2.1 VCO Characteristic

Figure 7.14 diagrams the experimental setup for the first test. Oscillation frequency is measured with the HP8563E spectrum analyzer. The recorded data is displayed in Figure 7.15. The difference between the x-axis in Figure 7.10, the simulated VCO characteristic, and Figure 7.15, the measured VCO characteristic, is due to a static offset added to the preamplifier output for the purpose of making a 1.573GHz oscillation frequency correspond approximately to a 0V differential control voltage. Because the preamplifier has a gain of -5, the control voltage must be multiplied by this factor to find the voltage across the accumulation-mode capacitors.

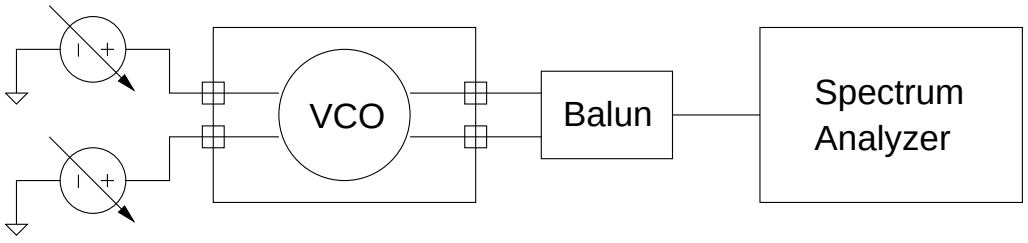


Figure 7.14: VCO test setup

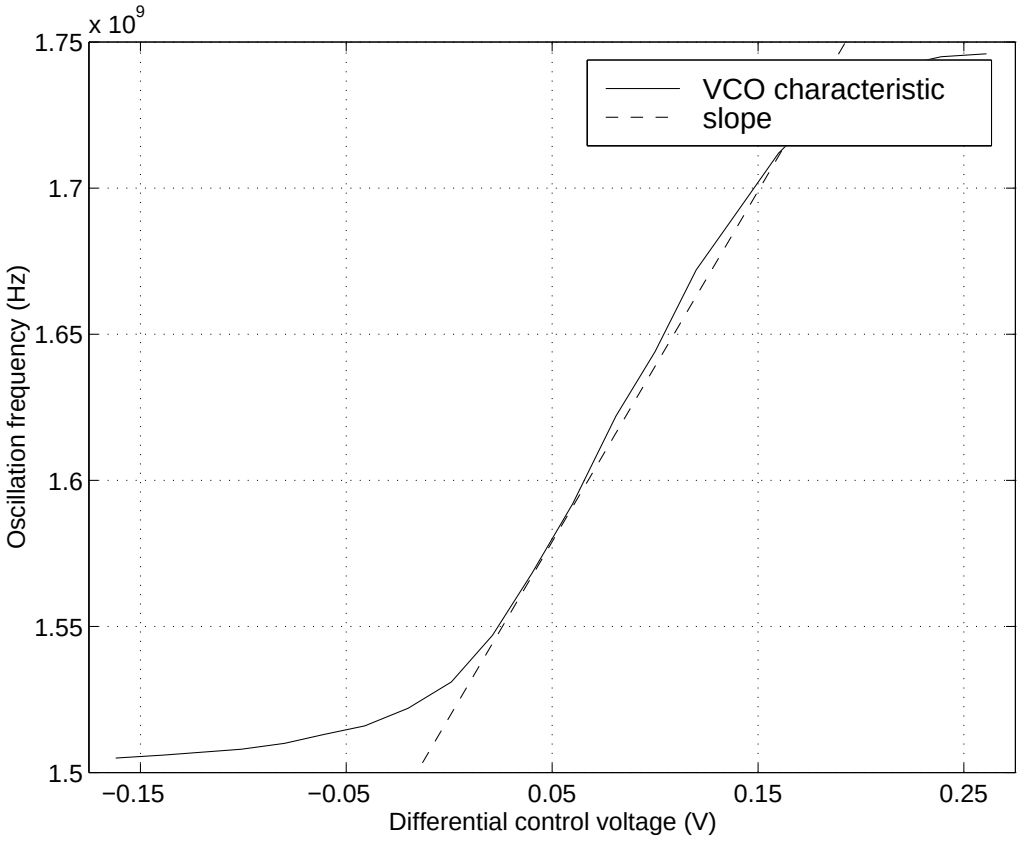


Figure 7.15: Measured VCO characteristic for circuit of Figure 7.9

The straight line in Figure 7.15 has a slope of 1.2GHz/V resulting in

$$K_o = 2\pi(1.2 * 10^9)\frac{rad}{sV}. \tag{7.5}$$

The minimum and maximum oscillation frequency, f_{min} and f_{max}, are 1.505GHz and 1.745GHz, respectively. The VCO tuning range is therefore

$$2\frac{f_{max} - f_{min}}{f_{max} + f_{min}}100\% = 14.8\%. \tag{7.6}$$

In the process of characterizing the signal path (see Figure 7.1), the radio's performance in the presence of a strong interferer, or blocker, is measured (see Figure 7.16). For the offsets shown, the limiting mechanism is the phase noise of

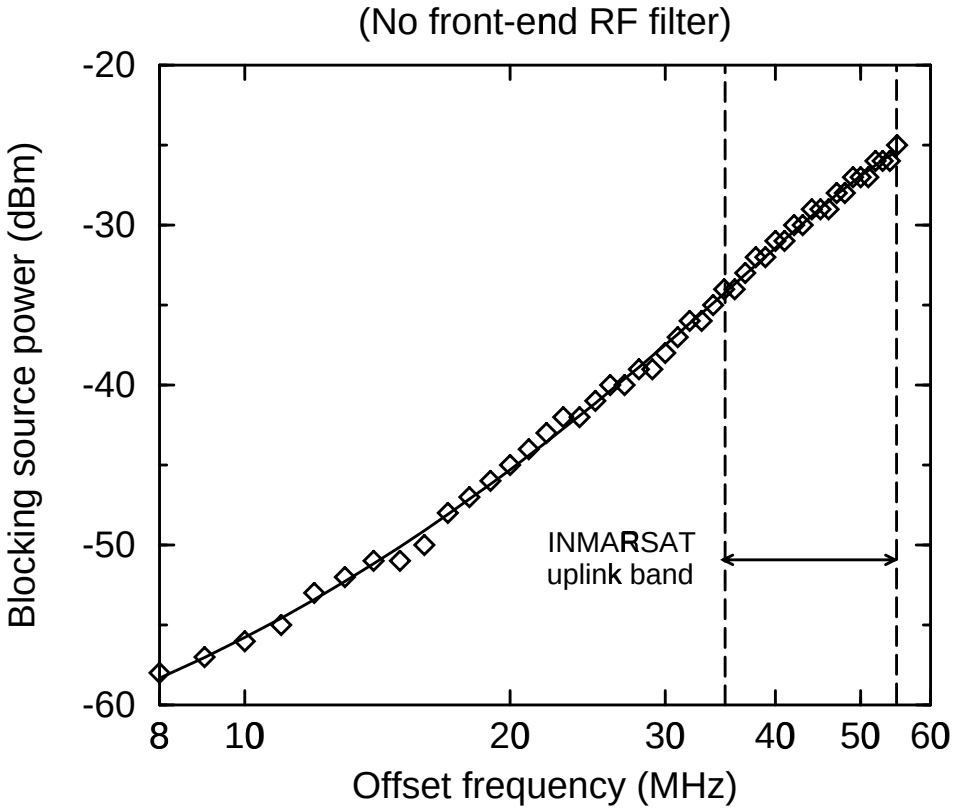


Figure 7.16: Radio 1dB blocking performance

the PLL's VCO, as these offsets are outside the loop bandwidth. At 35MHz, the blocking source power that results in a 1dB increase in the noise floor is -35dBm. Using -174dBm/Hz as the input thermal-noise level and a radio noise figure of 2.9dB results in a maximum phase noise of -140dBc/Hz at 35MHz offset. This performance is not the best that can be achieved, and can be attributed to noise from the preamplifier in the oscillator core.

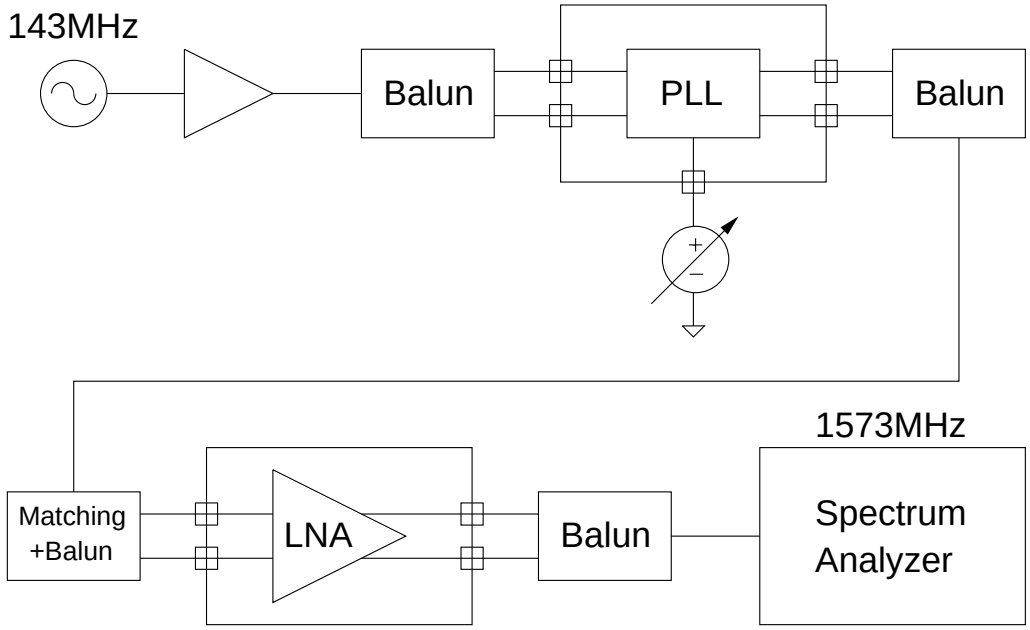


Figure 7.17: PLL test setup #1

7.2.2 Closed-Loop Phase Transfer Function

Figure 7.17 shows the experimental setup for the second test. Phase noise is measured for offsets from 1kHz to 10MHz with the HP8563E spectrum analyzer, which has special phase-noise measurement software. Ten MHz is used as the upper limit since the loop is designed to have a bandwidth less than 10MHz. Beyond the loop bandwidth, the PLL phase noise is determined by the VCO phase noise, making measurement of the PLL transfer function difficult. One of the largest factors affecting measurement accuracy is the instrument noise floor. To minimize this error source, measurements of the floor with a clean source are performed first. These results are later used to calibrate the data. The reference's phase noise and the APD PLL's phase noise are also both measured. After some data processing, the closed-loop phase transfer function, $H(s)$, is determined.

Figure 7.18 shows the measured $H(f)$ as well as the theoretical predictions of (7.4), for the case where the reference frequency is 143MHz and the VCO frequency is 1.573GHz ($N = 11$). The seven loop parameters are set as follows in (7.4): N

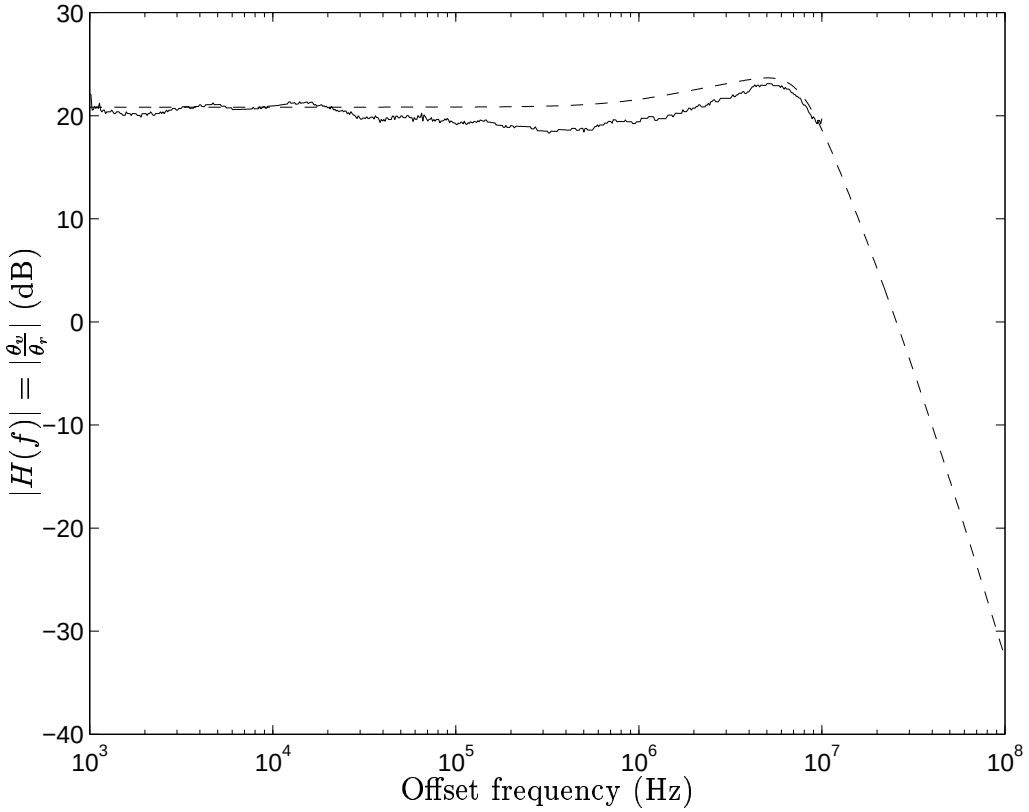


Figure 7.18: Measured and predicted $|H(f)|$

is known; K_o is taken from measured VCO data; R, C_1, and C_2 are taken to be their designed loop-filter values; f_{3dB} is calculated from the technology data; and K_d is fit. The fit value of K_d, 6.6μA/rad, is a little less than the simulated value noted in Table 7.1, 7.4μA/rad. Since $K_d = \frac{I_p}{2\pi}$ for an ideal APD, one might argue that the discrepancy in K_d is due to an actual pump current that is lower than the value used in simulations. But when I_p is measured, it is found to be correct. Still, the discrepancy in K_d is readily explained. The simulation establishes an upper bound on K_d, because it is measured in a state of zero static phase error. But the simulated APD+CP characteristic of Figure 7.4 illustrates that the detector gain decreases the farther one departs from zero radians. The charge pump is known to have some offset, and there is thus some static phase error in lock, resulting in a slightly lower K_d.

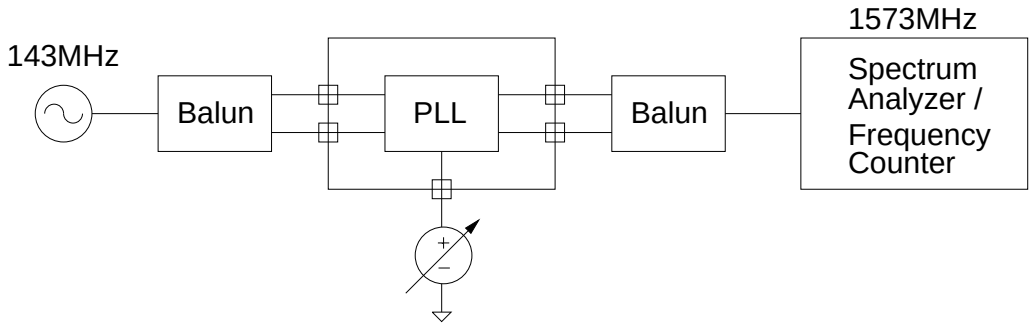


Figure 7.19: PLL test setup #2

7.2.3 Local Oscillator Spectrum

Figure 7.19 shows the experimental setup for the third test. The LO spectrum is measured with the HP8563E spectrum analyzer and the frequency is checked with an HP5350B microwave frequency counter. Figure 7.20 displays the synthesized output spectrum, in which the PLL's ability to track the low close-in phase noise of the reference can be seen. The visible skirts are due to the VCO's phase noise outside the PLL's 6MHz bandwidth. Spurious tones at -47dBc are primarily due to control-line ripple resulting from charge pump leakage. For GPS applications the measured spurious level is acceptable because of the absence of blockers at the corresponding offset frequencies. In more demanding applications, one may reduce control-line ripple through improved charge-pump design and the use of analog phase interpolation (API) [32].

Table 7.2 provides a summary of the APD PLL performance. The APD PLL has a wide bandwidth of 6MHz and the APD circuit consumes only one-quarter of the total 36mW synthesizer power. With the elimination of the divider, the main power consumer in the synthesizer is now the VCO.

In order to interpret this loop power consumption properly, one must recognize that power is saved by removing a nearly fixed consumer from the loop. In this 0.5μm technology, a frequency divider operating with an input frequency of roughly 1.6GHz at 2.5V, and implementing divides greater than 10, consumes approximately 15–20mW. Thus, if the loop described here included a divider, its total power consumption would be about 51–56mW. The savings ratio in this case can be misleading,

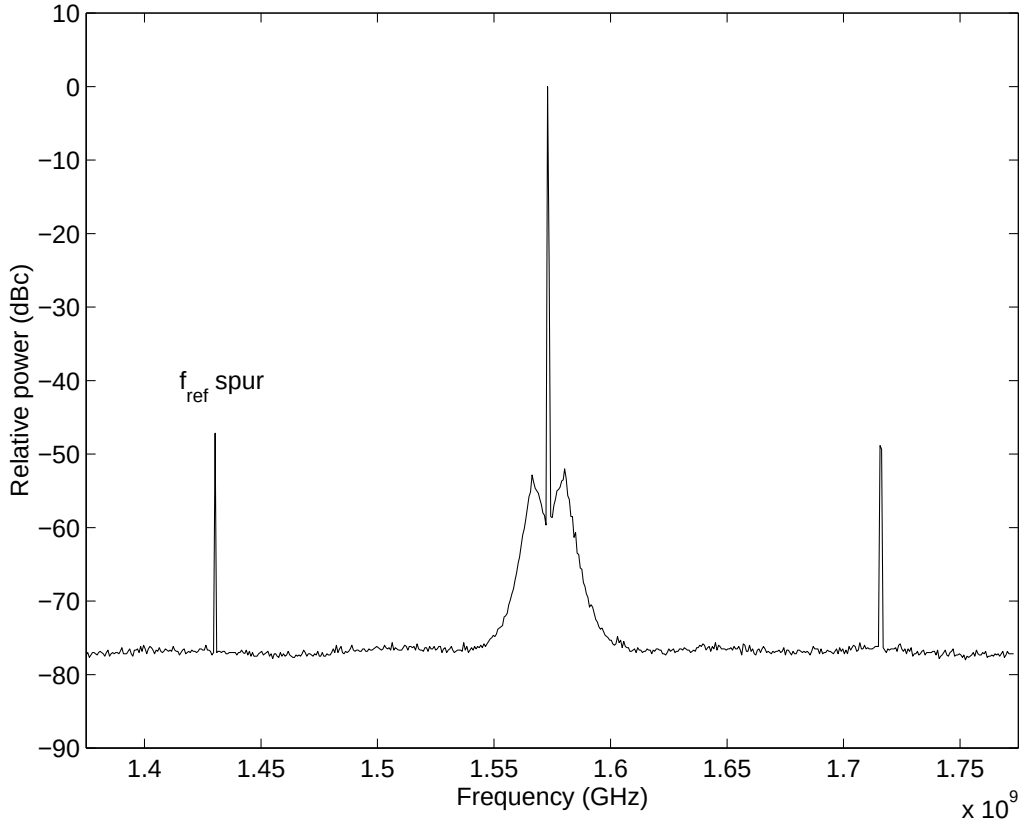


Figure 7.20: LO spectrum

because if we start with an APD PLL that consumes less power (e.g., if the VCO design is more power efficient) then the ratio of powers with and without a divider can be greater than the roughly 1.5 factor that we have here.

Table 7.2: Measured APD-PLL performance

APD PLL	
Synthesized frequency	1.573GHz
Reference frequency	143MHz
Loop bandwidth	6MHz
f_{ref} spur	≤ -45dBc
$2 * f_{ref}$ spur	≤ -55dBc
VCO power consumption	26mW
Total power consumption *	36mW
LO leakage @ LNA input	≈ -53dBm
Die area	3.1mm^2
Technology	0.5μm CMOS
VCO	
Gain constant, K_o	$2\pi(1.2 * 10^9)$rad/s/V
Tuning range †	240MHz
Phase noise @ 35MHz	≤ -140dBc/Hz
Power consumption:	
preamplifier	5.5mW
core	5.0mW
postbuffer	2.5mW
postbuffer with AGC	12.5mW

*2.5V supply.

†14.8% with center frequency of 1.625GHz.

7.3 Summary

This chapter covered an implementation of a dividerless frequency synthesizer. CMOS is a good technology for the synthesizer, because the aperture phase detector can be designed with precharged gates, easing aperture timing requirements. By using an APD in place of the divider and phase/frequency detector, we see that the charge pump's design requirements change. Specifically, the time difference between the APD's E and L output signals, i.e., the pump duration, is very small (on the order of the VCO period). This duration is in contrast to that of a phase/frequency detector which can have pump times on the order of the reference period. Therefore, static mismatch current in the charge pump is a more significant issue for APD based loops. On the other hand, the requirement of minimum PFD U and D pulse widths, to prevent a deadzone in the phase detector characteristic, is not an issue for the APD's E and L signals. Finally, we showed that measured results on the actual implementation agree with the theoretical model from Chapter 6.

Chapter 8

Conclusion

WE have discussed two promising new architectures for frequency conversion and synthesis in the context of the first single-chip CMOS GPS radio. These techniques, presented in Chapters 4 through 7, enable the radio to consume only 115mW in a 0.5μm technology. We now conclude with a summary of the voltage mixer and dividerless frequency synthesizer.

8.1 Summary

The double-balanced CMOS voltage mixer is constructed from four transistors and is ideally suited for CMOS processes since MOS transistors function as excellent voltage switches. In contrast, bipolar junction transistors are better suited to steering currents. Recognizing the inherent difference between a MOS and bipolar junction transistor, we can see why the Gilbert-type mixer, based on commutating current, is ubiquitous in bipolar radios but not in CMOS radios. In fact, the voltage mixer exceeds the performance of the Gilbert-type mixer when implemented in CMOS in all metrics except conversion gain.

Although conversion gain is a concern for the voltage mixer, our analysis shows that improved conversion gain is achieved with capacitive loading. In fact, it is interesting to note that the commutating mixer can achieve a conversion gain of

one. This result contradicts a prior understanding that limited the maximum conversion gain to $\frac{2}{\pi}$. While its low conversion gain compared to other architectures may be perceived as a disadvantage, the fact that the voltage mixer consumes very little power and requires no static bias current for the quad, makes the gain sacrifice worthwhile. It also achieves SSB noise figures on the order of 6dB with good linearity, determined by the LO drive amplitude. Thus, this architecture delivers wide dynamic range with minimal power consumption.

The voltage mixer addresses the need for a low-power frequency conversion architecture and similarly, the aperture phase detector provides a method for low-power frequency synthesis by eliminating the need for a power consuming divider. It is interesting to discover that phase lock can be maintained between two signals with dissimilar frequencies using a windowing technique to measure phase errors. As anticipated, an APD PLL has the same LTI model in lock as a loop containing a divide-by-N block. The implications to a loop based on this phase detector are reduced power consumption and on-chip interference from a digital divider.

In the course of this work, other important characteristics of the APD were discovered. Its unique out-of-lock behavior indicates a desire to sweep to a frequency related to window timings rather than the desired integer reference multiple. Furthermore, investigation of an APD PLL shows that the loop can lock to subharmonic integer multiples or reference subharmonics. These insights explain a fundamental need for a frequency acquisition aid to initially lock the loop to the proper output frequency.

8.2 Further Research

For robustness, the frequency acquisition aid (FAA) currently needs to be periodically activated to verify that the loop is still in lock or if it has lost lock, to re-lock the loop. An advance to this approach is enabled if an out-of-lock detector can be designed for the APD PLL. It would signal when the FAA needs to be powered on to bring the loop back to lock.

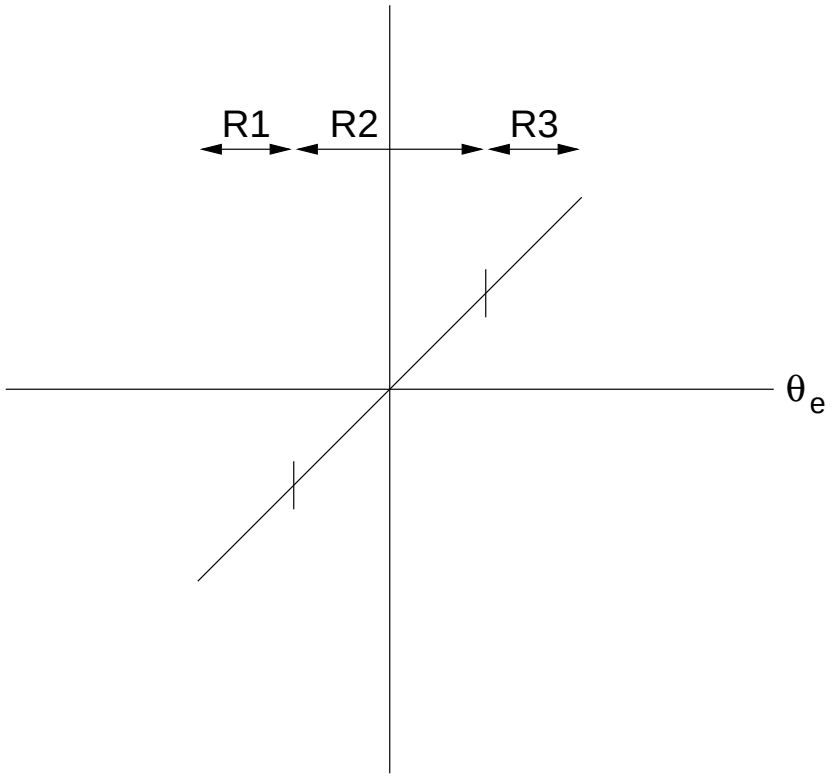


Figure 8.1: APD characteristic for one VCO period divided into regions

One difficulty in designing the out-of-lock detector is differentiating between an integer and subharmonic lock mode. If the APD PLL is in a subharmonic mode, the out-of-lock detector should signal the FAA to turn on. As a starting point, we return to an observation that was made in §6.5. The VCO edge position within the window changes from cycle to cycle if the loop is locked in a subharmonic mode. It also moves when the APD PLL is out-of-lock. Thus, if the out-of-lock detector can detect this motion, it can signal that the loop is either not locked or stuck in a subharmonic mode, both of which require the FAA to be started.

One strategy to detect this movement is to divide the APD characteristic into regions, as shown in Figure 8.1, using the APD's L and E output signals. The out-of-lock detector would be a state machine using this region knowledge to implement its out-of-lock determining algorithm. Since the state machine is clocked at the reference

frequency, its additional power consumption is minimal and its design complexity is eased compared to that of a much higher frequency divider.

Appendix A

Mixer Core's Impulse Response and Superposition Integral

An impulse is applied to the circuit in Figure 4.23 at time τ, $v_T(t) = \delta(t-\tau)$. To determine the initial voltage produced on C_L, the Thévenin equivalent circuit is transformed into a Norton equivalent circuit with the following short circuit current:

$$i_N(t) = g_T(t)v_T(t) = g_T(\tau)\delta(t-\tau) \tag{A.1}$$

The total charge delivered to the capacitor, as a result of the impulse in voltage, is $g_T(\tau)$ coulombs. This charge produces an initial voltage of $g_T(\tau)/C_L$ volts on C_L at time τ. Then, the following differential equation describes the circuit's response to this initial condition:

$$C_L\frac{dv_{if}(t)}{dt} = -g_T(t)v_{if}(t) \tag{A.2}$$

The solution has the form $h(t) = Ae^{-f(t)}$. Combining the initial condition with this solution, and noting that the system is causal, yields

$$h(t,\tau) = \frac{g_T(\tau)}{C_L}e^{-\int_\tau^t \frac{g_T(s)}{C_L}ds}u(t-\tau) \tag{A.3}$$

where $u(t)$ is the unit step function. Finally, using (A.3) in the superposition integral produces

$$v_{if}(t) = \int_{-\infty}^{t} \frac{g_T(\tau)}{C_L} e^{-\int_{\tau}^{t} \frac{g_T(s)}{C_L} ds} m(\tau) v_{rf}(\tau) d\tau \tag{A.4}$$

Some useful manipulations are enabled if $g_T(t)$ is written as

$$g_T(t) = \overline{g_T} + \sum_{n=1}^{\infty} a_n cos(n2\omega_{LO}t + \phi_n) = \overline{g_T} + \widetilde{g_T}(t) \tag{A.5}$$

where $\overline{g_T}$ is the DC level of $g_T(t)$. Furthermore, the integral of $\widetilde{g_T}(t)$ will be called $\widetilde{f_T}(t)$:

$$\widetilde{f_T}(t) = \frac{\overline{g_T}}{2\omega_{LO}} \sum_{n=1}^{\infty} \frac{a_n sin(n2\omega_{LO}t + \phi_n)}{n\overline{g_T}} + K \tag{A.6}$$

where K is an arbitrary constant. These modifications allow us to write

$$\begin{aligned} v_{if}(t) = e^{\frac{\widetilde{f_T}(t)}{C_L}} \int_{-\infty}^{t} & \frac{\overline{g_T}}{C_L} e^{-\frac{\overline{g_T}}{C_L}(t-\tau)} e^{-\frac{\widetilde{f_T}(\tau)}{C_L}} \\ & \frac{g_T(\tau)}{\overline{g_T}} m(\tau) v_{rf}(\tau) d\tau \end{aligned} \tag{A.7}$$

This last result warrants close attention. The exponentials involving $\widetilde{f_T}$ have a coefficient that multiplies a series of normalized sinusoids. This coefficient is equal to

$$\frac{\overline{g_T}}{2\omega_{LO}C_L} \tag{A.8}$$

and gives rise to three cases: if it is much less than 1, the exponentials involving $\widetilde{f_T}$ reduce to 1; if it is much greater than 1, the result for $C_L = 0$ should be used; or if it is between these two extremes, the impact of the two exponential terms involving $\widetilde{f_T}$ is ambiguous.

Appendix B

Break-before-make Conversion Gain

Figure 4.25(a) shows the modified mixing function for a break-before-make drive. Since this waveform has quarter-wave symmetry, we can focus on a quarter period to determine its Fourier series, as shown in Figure B.1. Notice that the modified mixing function becomes zero at a time prior to $\frac{T_{LO}}{4}$. We can solve for this time by evaluating

$$A_{LO}\cos(2\pi f_{LO}t) + B_{LO} = V_{th} \tag{B.1}$$

for t:

$$t = \frac{T_{LO}}{2\pi}\cos^{-1}\left(\frac{V_{th} - B_{LO}}{A_{LO}}\right) = \frac{T_{LO}}{2\pi}\cos^{-1}(r). \tag{B.2}$$

From Figure B.1, the magnitude of the modified mixing function's Fourier transform evaluated at the LO frequency is

$$|M'(f_{LO})| = \frac{4}{T_{LO}}\int_0^{\frac{T_{LO}}{2\pi}\cos^{-1}(r)} \frac{A_{LO}\cos(2\pi f_{LO}t) + B_{LO} - V_{th}}{A_{LO} + B_{LO} - V_{th}}\cos(2\pi f_{LO}t)dt. \tag{B.3}$$

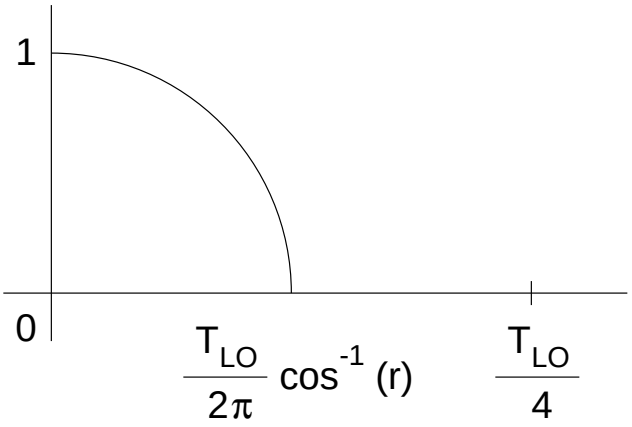


Figure B.1: Quarter period of modified mixing function

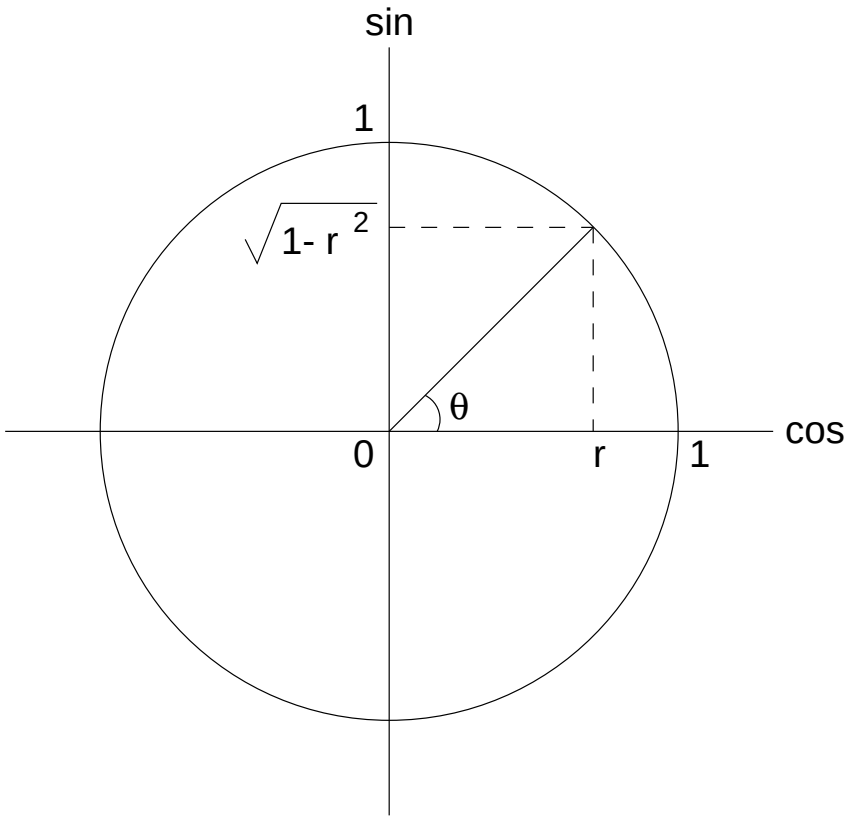


Figure B.2: Trigonometric relationships

The two integral forms in (B.3) evaluate as:

$$\int_0^{\frac{T_{LO}}{2\pi}\cos^{-1}(r)} \cos(2\pi f_{LO}t)dt = \frac{T_{LO}}{2\pi}\sin(\cos^{-1}(r)) \tag{B.4}$$

and

$$\int_0^{\frac{T_{LO}}{2\pi}\cos^{-1}(r)} \cos^2(2\pi f_{LO}t)dt = \frac{T_{LO}}{4\pi}\cos^{-1}(r) + \frac{T_{LO}}{8\pi}\sin(2\cos^{-1}(r)). \tag{B.5}$$

Two useful trigonometric equalities are apparent from Figure B.2:

$$\sin(\cos^{-1}(r)) = \sqrt{1-r^2} \tag{B.6}$$

and

$$\sin(2\cos^{-1}(r)) = 2r\sqrt{1-r^2}. \tag{B.7}$$

Using (B.6) with (B.4) and (B.7) with (B.5) in (B.3) gives

$$\begin{aligned} |M'(f_{LO})| =& \frac{A_{LO}}{\pi(A_{LO}+B_{LO}-V_{th})}\left[\cos^{-1}(r) + \frac{2r\sqrt{1-r^2}}{2}\right] + \\ & \frac{2(B_{LO}-V_{th})}{\pi(A_{LO}+B_{LO}-V_{th})}\sqrt{1-r^2}. \end{aligned} \tag{B.8}$$

Next, by definition

$$A = \frac{A_{LO}+B_{LO}-V_{th}}{\frac{4}{T_{LO}}\int_0^{\frac{T_{LO}}{2\pi}\cos^{-1}(r)} A_{LO}\cos(2\pi f_{LO}t) + B_{LO} - V_{th}dt}, \tag{B.9}$$

where the term in the numerator is the LO drive's peak value minus the switch threshold voltage, and the term in the denominator is the average value of a waveform with shape and peak value of the numerator as in Figure 4.21. Evaluating the integral in the denominator yields

$$A = \frac{\pi(A_{LO}+B_{LO}-V_{th})}{2[A_{LO}\sqrt{1-r^2} + (B_{LO}-V_{th})\cos^{-1}(r)]}. \tag{B.10}$$

The product of (B.8) and (B.10) is

$$A|M'(f_{LO})| = \frac{A_{LO}(\cos^{-1}(r) + r\sqrt{1-r^2}) + 2(B_{LO} - V_{th})\sqrt{1-r^2}}{2[A_{LO}\sqrt{1-r^2} + (B_{LO} - V_{th})\cos^{-1}(r)]}. \tag{B.11}$$

Dividing all terms by A_{LO} and recognizing that $\frac{B_{LO}-V_{th}}{A_{LO}} = r$ yields

$$A|M'(f_{LO})| = \frac{\cos^{-1}(r) - r\sqrt{1-r^2}}{2[\sqrt{1-r^2} - r\cos^{-1}(r)]}, \tag{B.12}$$

which is the conversion gain for a break-before-make LO drive when the average mixer bandwidth is much less than the LO frequency. One last note is that if the average mixer bandwidth is also less than or equal to the IF, then the conversion gain will be lower (see Figure 4.27).

Appendix C

De-Embedding Noise Figure for the Standalone Mixer

Figure C.1 partitions the experimental setup that was used to measure the noise figure of the standalone mixer into three distinct blocks. For each block, there is an associated available power gain G_a, noise factor F, RF source resistance R_s that matches the block's input, and RF output resistance R_o assuming R_s is present. The noise figure measured by the meter for this setup is

$$F_{meter} = F_1 + \frac{F_2 - 1}{G_{a1}} + \frac{F_3 - 1}{G_{a1} G_{a2}}. \tag{C.1}$$

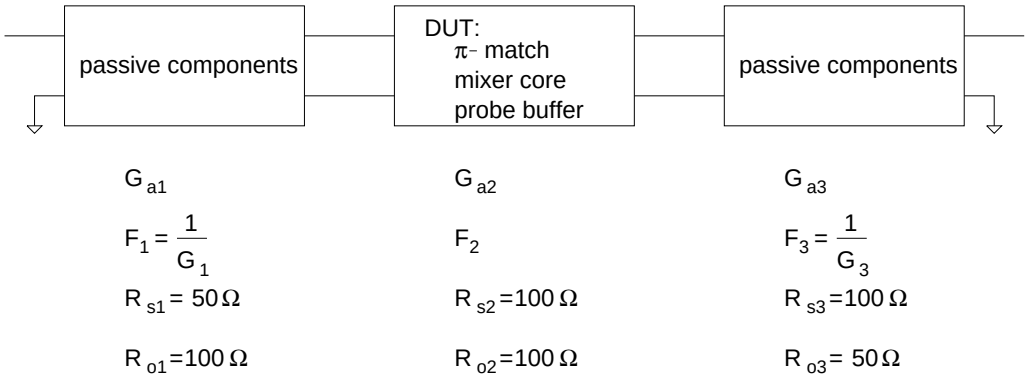


Figure C.1: Partition of experimental setup

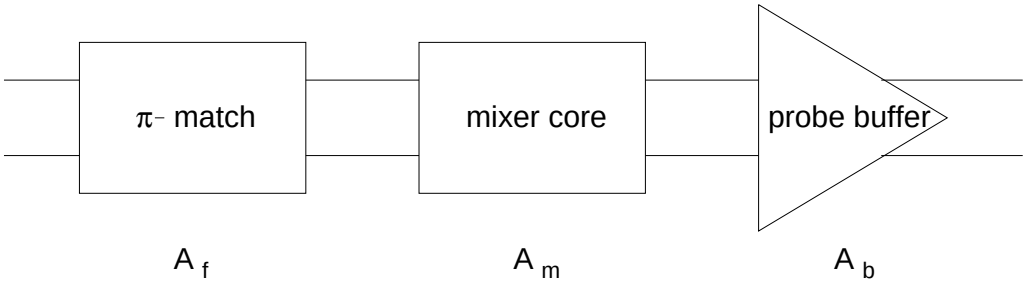


Figure C.2: Partition of DUT

The DUT's noise figure, F_2, is the unknown to be solved for, and is isolated such that

$$F_2 = G_{a1}\left[F_{meter} - F_1 - \frac{F_3 - 1}{G_{a1}G_{a2}}\right] + 1. \tag{C.2}$$

G_{a1} and G_{a3} are measured during testing, and hence F_1 and F_3 are known. But G_{a2} needs to be calculated before F_2 can be determined.

By definition,

$$G_{a2} = \frac{P_{ao2}}{P_{ai2}}, \tag{C.3}$$

where in Figure C.2 P_{ai2} is the available input power and P_{ao2} is the available output power. We can write

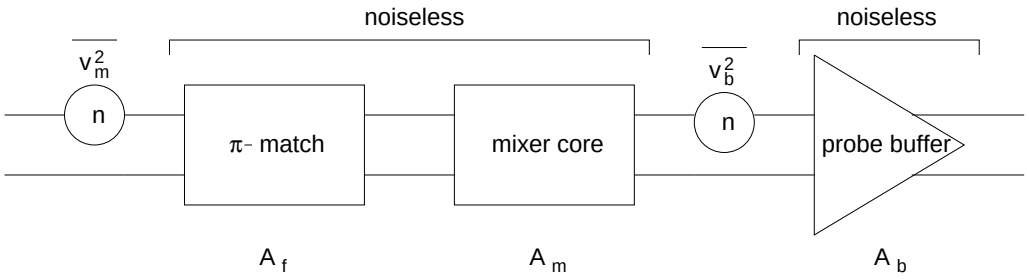
$$P_{ai2} = \frac{V^2}{4(100\Omega)} \tag{C.4}$$

and

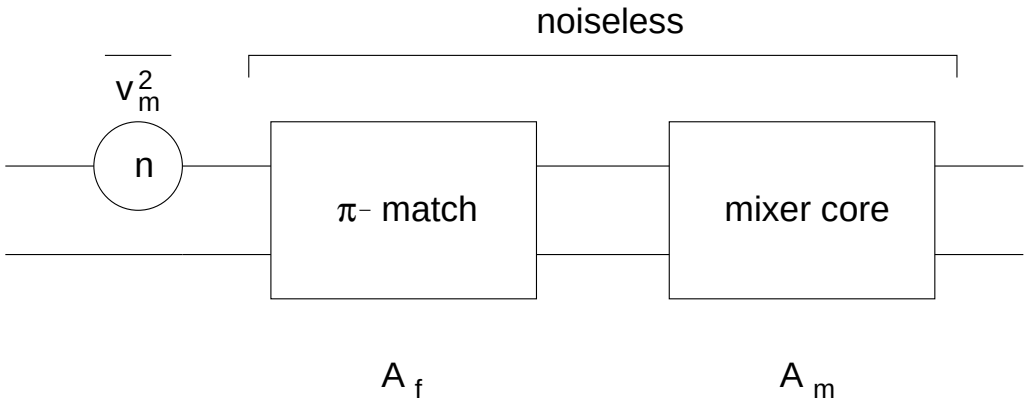
$$P_{ao2} = \frac{(A_f A_m A_b V)^2}{16(100\Omega)}, \tag{C.5}$$

which yields

$$G_{a2} = \frac{(A_f A_m A_b)^2}{4}. \tag{C.6}$$



(a) Input-referred voltage noise sources in DUT



(b) Input-referred voltage noise of mixer

Figure C.3: DUT noise sources

The product $A_f A_m A_b$ can be found from measurement, in which the As are the voltage gains of blocks within the DUT.

Now that F_2 can be calculated from the meter's reading, the noise from the probe buffer must be removed to arrive at the mixer's noise figure. Using Figure C.3(a), we can express the DUT's noise figure as

$$F_2 = 1 + \frac{\overline{v_m^2}}{4kT100\Omega} + \frac{\overline{v_b^2}}{kT100\Omega(A_f A_m)^2}, \tag{C.7}$$

where v_m and v_b are the input-referred voltage noises for the mixer and probe buffer, respectively. The product $A_f A_m$ can be calculated, since the product $A_f A_m A_b$ is

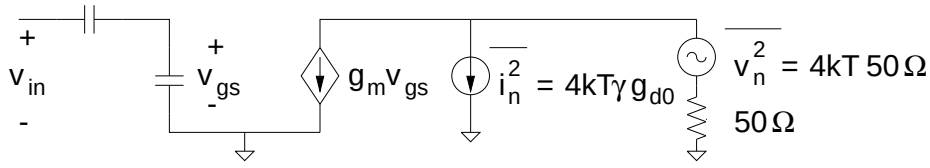


Figure C.4: Small-signal model of probe buffer and noise sources

measured and A_b is known from simulations. Figure C.3(b) shows that the mixer's noise figure is

$$F_m = 1 + \frac{\overline{v_m^2}}{4kT100\Omega}. \tag{C.8}$$

Noticing that the expression for F_m appears in (C.7) gives

$$F_2 = F_m + \frac{\overline{v_b^2}}{kT100\Omega(A_f A_m)^2}. \tag{C.9}$$

This time, F_m is the unknown to be solved for and is isolated such that

$$F_m = F_2 - \frac{\overline{v_b^2}}{kT100\Omega(A_f A_m)^2}. \tag{C.10}$$

All that is left is to determine is the probe buffer's input-referred voltage noise.

Figure C.4 shows the probe buffer's half-circuit from which we can calculate its input-referred voltage noise. The probe buffer's input-referred voltage noise is double this quantity, which is

$$\overline{v_b^2} = 2\left[\frac{4kT\gamma g_{d0}}{(ag_m)^2} + \frac{4kT50\Omega}{A_b^2}\right]. \tag{C.11}$$

Here, a is the attenuation due to the capacitive voltage divider at the input, g_m is the transconductance, and g_{d0} is the peak value of the drain to source conductance.

These three quantities are readily found from simulations of the probe buffer. Using (C.11) in (C.10) gives

$$F_m = G_{a1}\left[F_{meter} - F_1 - \frac{F_3 - 1}{G_{a1}G_{a2}}\right] + 1 - \frac{\overline{v_b^2}}{kT100\Omega(A_f A_m)^2}, \tag{C.12}$$

which shows how to calculate the mixer's noise figure from the noise figure meter's reading.

Appendix D

A Formal Treatment of PLL Fundamentals

We begin this appendix with Figure 6.1, and we end with Figure 6.2. Our goal is to show explicitly the mathematics relating these two figures. In §6.1 we expressed the PLL's input signal, $v_i(t)$. Now, we do the same for the PLL's phase-locked signal:

$$v_o(t) = A_o cos(\int_{-\infty}^{t} X(v_c(u))du + \phi_o) + B_o. \tag{D.1}$$

It too is a sinusoid with amplitude A_o and DC level B_o, but the angle of the sinusoid is described by the VCO's oscillation frequency-versus-control voltage behavior, $X(v_c)$, and an initial phase ϕ_o at $t = -\infty$. Because of the integral in (D.1), the angle depends on the history of the control voltage applied to the VCO. An alternate way to describe (D.1) is that it represents frequency modulation (FM) by $v_c(t)$, as it should, since a VCO can be used as an FM modulator [40].

The phase detector pictured in Figure 6.1 can be continuous time or sampled in nature. For the case of a sampled phase detector, we must invoke an assumption that the loop's bandwidth is small compared to the input frequency. Figure 6.1 is redrawn in Figure D.1 with Figure 6.1's phase detector broken into a phase detector and a filter. The filter in Figure D.1 will ensure that the narrow loop bandwidth criterion is met. The end result is a phase detector characteristic, $Y(\theta_e)$, whose

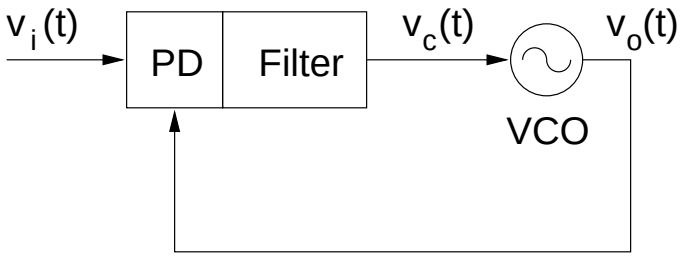


Figure D.1: PLL loop filter

output voltage or current, in the case of a charge-pump PLL, is a function of phase error, the difference in phase between the two inputs.

At this point, we have a function, $X(v_c)$, describing the VCO and a function, $Y(\theta_e)$, describing the phase detector. In general, these two blocks, and hence X and Y, can be nonlinear, increasing the PLL's analytic complexity. However, when the loop is in lock (i.e., the output phase tracks the input phase), the phase errors are small. Small phase errors allow the linearization of the VCO and phase detector about the lock point. The VCO's locked control voltage is designated by v_l and the phase detector's locked phase error is designated by θ_l. Linearizing $X(v_c)$ about v_l gives

$$X(v_c) \approx X(v_l) + K_o(v_c - v_l) = \Omega_o + K_o \Delta v_c, \tag{D.2}$$

where $X(v_l)$ is referred to as the VCO's free-running frequency, Ω_o, and K_o is $X'(v_l)$, which is the VCO's gain constant. The prime indicates a derivative (e.g., $' = \frac{d}{dv_c}$ in this case). Equation (D.1) can now be written as

$$v_o = A_o cos(\Omega_o t + K_o \int_{-\infty}^{t} \Delta v_c(u) du + \phi_o') + B_o, \tag{D.3}$$

where ϕ_o' is a new arbitrary constant. Similarly, linearizing $Y(\theta_e)$ about θ_l gives

$$Y(\theta_e) \approx Y(\theta_l) + K_d(\theta_e - \theta_l), \tag{D.4}$$

where K_d is $Y'(\theta_l)$, which is the phase detector's gain constant.

We are now in a position to add an input phase signal, $\Phi_i(t)$, to $v_i(t)$, and examine the loop in lock:

$$v_i(t) = A_i cos(\omega_i t + \phi_i + \Phi_i(t)) + B_i. \tag{D.5}$$

The first condition to note is that if the loop is locked, $\Omega_o = \omega_i$. Next, let us designate the remaining time-varying phase, $K_o \int_{-\infty}^{t} \Delta v_c(u)du$, with $\Phi_o(t)$ so that

$$v_o(t) = A_o cos(\Omega_o t + \Phi_o(t) + \phi'_o) + B_o. \tag{D.6}$$

The effect of $\Phi_i(t)$ on the VCO's control voltage is to perturb it from its locked value, v_l. Using (D.4), we can quantify the voltage perturbation as

$$\Delta v_c(t) = K_d(\theta_e(t) - \theta_l). \tag{D.7}$$

To make this discussion explicitly clear, $\theta_e(t) = \theta_i(t) - \theta_o(t)$, where $\theta_i(t) = \omega_i t + \phi_i + \Phi_i(t)$ and $\theta_o(t) = \Omega_o t + \Phi_o(t) + \phi'_o$.

At the time of lock, t_l, $\Omega_o = \omega_i$ and $\theta_e(t_l) = \theta_l$. Using these two conditions at t_l gives

$$\theta_l = \phi_i + \Phi_i(t_l) - (\Phi_o(t_l) + \phi'_o). \tag{D.8}$$

Equation (D.8) indicates that the PLL imposes a certain phase relationship between the input signal and phase-locked signal at lock. Once this relationship is established

$$\Delta v_c(t) = K_d(\Phi_i(t) - \Phi_i(t_l) - (\Phi_o(t) - \Phi_o(t_l))). \tag{D.9}$$

In this expression, the arbitrary constants ϕ_i and ϕ'_o have disappeared. By defining two new variables that are valid in lock,

$$\Phi'_i(t) = \Phi_i(t) - \Phi_i(t_l) \tag{D.10}$$

and

$$\Phi'_o(t) = \Phi_o(t) - \Phi_o(t_l) = K_o \int_{t_l}^{t} \Delta v_c(u) du, \tag{D.11}$$

(D.9) can be written as

$$\Delta v_c(t) = K_d(\Phi'_i(t) - \Phi'_o(t)). \tag{D.12}$$

Finally, we are at a point to construct the LTI block diagram for a locked PLL. By combining the Laplace transforms of (D.9) and (D.11), Figure 6.2 is produced with (6.2) as its closed-loop phase transfer function.

Appendix E

Static-current Mismatch in the Charge Pump

There is a current imbalance between the loop filter nodes due to random mismatches in the bias circuitry of Figure 7.6. In this appendix, we quantify the magnitude of the static-current mismatch, using 3σ statistics.

For the purposes of this calculation, Figure 7.6 can be partitioned into Figure E.1 and Figure E.2. Figure E.1 assumes that the loop filter nodes have the same voltage[1]. In Figure E.1, I_{1P} and I_{2P} are random variables with a mean of $\frac{I_T}{2}$ and a variance of σ_P^2. The difference between I_{1P} and I_{2P}, I_{MP}, is a random variable that captures the current imbalance due to the PMOS current-source loads. It has zero mean and a variance equal to $2\sigma_P^2$. The situation is slightly different in Figure E.2, because instead of having two independent current sources we have a correlated differential pair. In this case, our two random variables, I_{1N} and I_{2N}, always add to I_T. Therefore, to reflect this dependence,

$$I_{1N} = \frac{I_T}{2} + X \tag{E.1}$$

[1]This assumption is justified by the large capacitance at these nodes due to the loop filter, and allows us to focus on the current imbalance between these nodes.

and

$$I_{2N} = \frac{I_T}{2} - X, \tag{E.2}$$

where X is a random variable with zero mean and a variance of σ_N^2. The difference current, I_{MN}, is $2X$, thus it also has zero mean, but its variance is $4\sigma_N^2$. The net mismatch, I_M, is the difference of I_{MP} and I_{MN}, with variance

$$\sigma_M^2 = 2\sigma_P^2 + 4\sigma_N^2. \tag{E.3}$$

Using 3σ design rules, the maximum static current, I_s, that must be tolerated in the charge pump is

$$I_s = 3\sigma_M. \tag{E.4}$$

Using equations from [41], we can write

$$\sigma_P^2 = \frac{I_T^2}{4W_P L_P}\left[\frac{4A_{V_{T0}}^2}{V_{ODP}^2} + A_k^2\right] \tag{E.5}$$

and

$$\sigma_N^2 = \frac{I_T^2}{4W_N L_N}\left[\frac{4A_{V_{T0}}^2}{V_{ODN}^2} + A_k^2\right], \tag{E.6}$$

where A_{VT0} is the area proportionality constant for the threshold voltage, V_{ODP} and V_{ODN} are the respective overdrive voltages on the devices, and A_k is the area proportionality constant for the current factor. A_{VT0} and A_k are determined by the process's nominal gate oxide. Substituting (E.5) and (E.6) into (E.3), and combining this new expression with (E.4) yields

$$I_s = \frac{I_T}{4}\left[\frac{6}{W_P L_P}\left(\frac{4A_{V_{T0}}^2}{V_{ODP}^2} + A_k^2\right) + \frac{12}{W_N L_N}\left(\frac{4A_{V_{T0}}^2}{V_{ODN}^2} + A_k^2\right)\right]. \tag{E.7}$$

Table E.1 shows the parameters that influence the mismatch current and their values for this design. The 3σ static current, I_s, is also evaluated in it.

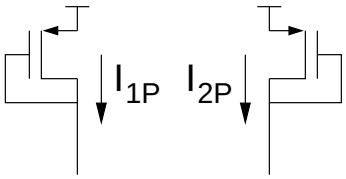


Figure E.1: Bias circuitry's current-source loads

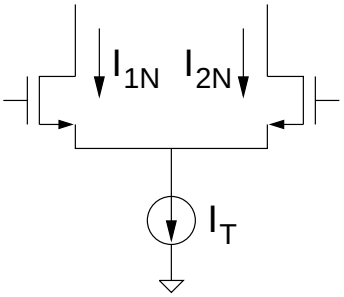


Figure E.2: Bias circuitry's differential pair

Table E.1: Static-current mismatch parameters

I_T	8μA
W_P	10μm
L_P	5μm
W_N	5μm
L_N	20μm
V_{ODP}	350mV
V_{ODN}	450mV
A_{VT0}	10mVμm
A_k	3%μm
I_s	220nA

Bibliography

[1] Theodore S. Rappaport, *Wireless communications: principles and practice*, Prentice Hall, 1996.

[2] Donald C. Cox, "Universal digital portable radio communications," *Proceedings of the IEEE*, vol. 75, no. 4, pp. 436–477, Apr. 1987.

[3] L. M. Burns, "Applications for GaAs and silicon ICs in next generation wireless communications systems," in *Proceedings of 1994 IEEE GaAs IC Symposium*, 1994, pp. 155–158.

[4] Barrie Glibert, "Why bipolar?," Internet: the Electronics Design, Technology & News network's analog avenue, Apr. 1998, http://www.edtn.com/analog/barrie2.htm.

[5] Derek K. Shaeffer et al., "A 115mW, 0.5μm CMOS GPS receiver with wide dynamic-range active filters," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 12, Dec. 1998.

[6] B. Bjerede et al., "An intermediate frequency modulator using direct digital synthesis techniques for japanese personal handy phone (PHP) and digital european cordless telecommunications (DECT)," in *1994 IEEE 44th Vehicular Technology Conference*, 1994, vol. 1, pp. 467–471.

[7] Lawrence E. Larson, "Integrated circuit technology options for RFIC's – present status and future directions," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 3, Mar. 1998.

[8] D. B. Leeson, "A simple model of feedback oscillator noise spectrum," *Proceedings of the IEEE*, vol. 54, no. 2, pp. 329–330, Feb. 1966.

[9] Paul R. Gray, "Architectures and technologies for CMOS RF transceivers," ISSCC 97 Short Course.

[10] Derek K. Shaeffer and Thomas H. Lee, "A 1.5V, 1.5GHz CMOS low-noise amplifier," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 5, pp. 745–759, May 1997.

[11] Ali Hajimiri and Thomas H. Lee, "A general theory of phase noise in electrical oscillators," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, Feb. 1998.

[12] Donald C. Cox, "Wireless personal communications: what is it?," *IEEE Personal Communications*, vol. 2, no. 2, pp. 20–35, Apr. 1995.

[13] Bradford W. Parkinson et al., *Global Positioning System: Theory and Applications Volume I*, American Institute of Aeronautics and Astronautics, Inc., 1996.

[14] P. W. Ward, "GPS receiver RF interference monitoring, mitigation, and analysis techniques," *Journal of the Institute of Navigation*, vol. 41, no. 4, pp. 367–391, 1994-1995, Winter.

[15] Derek K. Shaeffer, *The Design and Implementation of Low-Power CMOS Radio Receivers*, Ph.D. thesis, Stanford University, Dec. 1998.

[16] Edwin H. Armstrong, "The super-heterodyne – its origin, development, and some recent improvements," *Proceedings of the IRE*, vol. 12, pp. 539–552, Oct. 1924.

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

[18] M. McWhorter et al., "High frequency laboratory," Course Notes, 1996, Spring.

[19] David H. Shen et al., "A 900MHz RF front-end with integrated discrete-time filtering," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 12, pp. 1945–1954, Dec. 1996.

[20] M. Shinagawa et al., "Jitter analysis of high speed sampling systems," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 1, pp. 220–224, Feb. 1990.

[21] Alan V. Oppenheim and Ronald W. Schafer, *Discrete-Time Signal Processing*, Pentice Hall, 1989.

[22] Jan Crols and Michel S. J. Steyaert, "A 1.5GHz highly linear CMOS downconversion mixer," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 7, pp. 736–742, July 1995.

[23] Andrew N. Karanicolas, "A 2.7V 900MHz CMOS LNA and mixer," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 12, pp. 1939–1944, Dec. 1996.

[24] L. G. Kazovsky, "The fourier transform and its applications," Course Notes, 1994, Spring.

[25] Christopher D. Hull and Robert G. Meyer, "A systematic approach to the analysis of noise in mixers," *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications*, vol. 40, no. 12, pp. 909–919, Dec. 1993.

[26] H. P. Walker, "Sources of intermodulation in diode-ring mixers," *The Radio and Electronic Engineer*, vol. 46, no. 5, pp. 247–255, May 1976.

[27] J. G. Gardiner, "The relationship between cross-modulation and intermodulation distortions in the double-balanced modulator," *Proceedings of the IEEE*, vol. 56, no. 11, pp. 2069–2071, Nov. 1968.

[28] David M. Pozar, *Microwave Engineering*, Addison-Wesley Publishing Company, 1990.

[29] *M-tron Engineering Notes*, Dec. 1997.

[30] Gene F. Franklin et al., *Feedback Control of Dynamic Systems*, Addison-Wesley Publishing Company, second edition, 1991.

[31] Floyd M. Gardner, *Phaselock Techniques*, John Wiley & Sons, second edition, 1979.

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

[33] David A. Hodges and Horace G. Jackson, *Analysis and Design of Digital Integrated Circuits*, McGraw-Hill, Inc., 1988.

[34] J.F. Parker and D. Ray, "A 1.6GHz CMOS PLL with on-chip loop filter," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 337–343, Mar. 1988.

[35] D. Richman, "Color carrier reference phase synchronization accuracy in ntsc color television," *Proceedings of the IRE*, vol. 42, pp. 106–133, Jan. 1954.

[36] Jiren Yuan and Christer Svensson, "New TSPC latches and flipflops minimizing delay and power," in *1996 Symposium on VLSI Circuits Digest of Technical Papers*, 1996, pp. 160–161.

[37] C. Patrick Yue and S. Simon Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF ICs," in *1997 Symposium on VLSI Circuits Digest of Technical Papers*, 1997, pp. 85–86.

[38] C. Patrick Yue and S. Simon Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF ICs," *IEEE Journal of Solid-State Circuits*, vol. 33, pp. 743–752, May 1998.

[39] T. Soorapanth et al., "Analysis and optimization of accumulation-mode varactors for RF ICs," in *1998 Symposium on VLSI Circuits Digest of Technical Papers*, 1998, pp. 32–33.

[40] Thomas S. Laverghetta, *Analog Communications for Technology*, Saunders College Publishing, 1991.

[41] Marcel J.M. Pelgrom et al., "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, pp. 1433–1439, Oct. 1989.