UC Berkeley

Title
Design of Reconfigurable Radio Front-Ends

Permalink
https://escholarship.org/uc/item/0xg9b3v4

Author
Xiao, Xiao

Publication Date
2016

Peer reviewed|Thesis/dissertation

Design of Reconfigurable Radio Front-Ends

by

Xiao Xiao

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Borivoje Nikolic, Chair
Professor Ali Niknejad
Professor Paul Wright

Spring 2016

Design of Reconfigurable Radio Front-Ends

Abstract

Design of Reconfigurable Radio Front-Ends

by

Xiao Xiao

Doctor of Philosophy in Engineering - Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Borivoje Nikolic, Chair

Modern and future mobile devices must support increasingly more wireless standards and bands. Currently, multi-band coexistence is enabled by a network of discrete, off-chip components that are bulky, expensive, and narrowband. As transceivers are required to accommodate an increasing number of wireless bands, the required number of discrete components increases accordingly, resulting in greater bill of materials (BoM) cost and front-end module (FEM) area. This work focuses on design techniques to enable front-end integration and reconfigurability in multi-band radios.

In the first part of this work, we present a wideband spectrum sensing receiver with high sensitivity, wide dynamic range, and low power overhead. Reconfigurability in multi-band radios requires environmental awareness, and spectrum sensing can be used for optimal channel selection and adaptive interference suppression. The 300MHz-700MHz spectrum sensing receiver uses subsampling downconversion and digital-analog hybrid correlation to achieve -104dBm sensitivity and 84dB dynamic range for a 6MHz channel while consuming only 28mW of power. In the second part of this work, we present a wideband time division duplex (TDD) front-end with an innovative transmit/receive (T/R) switching scheme. T/R switches conventionally are off-chip components, and existing integrated designs have been narrowband or high loss. We propose a wideband integrated T/R switching technique in which the PA is re-used as an LNA during receive mode, and we demonstrate this in a 20dBm polar transmitter that can be re-purposed into a 3.4GHz-5.4GHz LNA achieving -6.7dBm P1dB and 5.1dB NF. The two systems presented in this thesis contribute key innovations towards fundamental aspects of future reconfigurable radios - greater front-end integration, wideband transceiver design, and spectrum awareness.

Contents

List of Figures

List of Tables

Acknowledgments

First, I would like to thank my research advisor Prof. Borivoje Nikolic for his unyielding support and encouragement. He has been extremely accommodating and sensitive to my needs throughout my graduate school career, and has been tremendously helpful for my personal and professional development. It has been a pleasure to work with him.

I would also like to thank Prof. Ali Niknejad and Prof. Elad Alon for many helpful technical discussions, for providing research guidance, and for being great teachers in the classroom. I would further like to thank Prof. Dave Allstot, for valuable research discussions and feedback, and Prof. Paul Wright, for serving on my qualification and dissertation committees. Thanks also to industrial visitors Dr. Christopher Hull, Dr. Vason Srini, and Dr. Sudhir Aggarwal for being involved in my work and providing helpful feedback.

I am very grateful to all current and past BWRC staff for efficiently handling all the financial, administrative, and technical infrastructure aspects of academic research. Their efforts make BWRC an engaging, bright, and comfortable work environment. I would especially like to thank James Dunn for working closely with me on the TDD project and providing essential PCB and vendor support, even through all the frustrating delays and revisions. Much thanks also to Brian Richards, Fred Burghardt, Deirdre McAuliffe-Bauer, Leslie Nishiyama, Olivia Nolan, Sarah Jordan, and Candy Corpus.

I would like to acknowledge the funding sources that have supported my research as well as me personally, including NSF GRFP, C2S2, and DARPA RF-FPGA program (HR0011-12-9-0013). Special thanks to TSMC University Shuttle Program for chip fabrication and enabling us to perform meaningful integrated circuits research.

I would like to thank Amanda Pratt, who was my research partner on the TDD project. I am grateful to her not only as my project partner, but also for being a source of emotional support and understanding. Thanks to Angie Wang and Bonjern Yang for their direct contributions to the TDD project as well. I would also like to acknowledge all the RF/analog designers at the BWRC who have helped me with technical discussions over the years, and who were always open, friendly, and receptive to my questions and requests. Much thanks to Charles Wu, Will Biederman, Dan Yeager, Steven Callender, Nai-Chung Kuo, Sameet Ramakrishnan, Luke Calderin, Andrew Townley, Ashkan Borna, Yue Lu, Yida Duan, and Hanh-Phuc Le.

In addition to some already mentioned, I would like to thank the following BWRC students and alumni for their camaraderie and friendship. They made my graduate school experience cheerful and fun in joyful times, and they provided much needed empathy and support in more trying times. Thanks to Matt Weiner, Matt Spencer, Milos Jorgovanovic, Katerina Papadopoulou, Jaehwa Kwak, Angie Wang, Pi-Feng Chiu, Rachel Hochman, Mira Midenovic-Misic, Dusan Stepanovic, and Vinayak Nagpal.

I would like to thank my parents for their continuous support, encouragement, and guidance. They instilled in me the values of hard work, perseverance, and inner strength, and it is due to their efforts and their example that I was able to both start and complete this degree.

Lastly, I would like to thank my husband. Christopher, your support and patience have meant the world in my time of focus and stress. So much of my achievement rests upon your sacrifices, that I'd totally give you this Ph.D. if I could. You even wrote this paragraph, perhaps the best of the entire thesis. Your thoughtfulness knows no bounds.

Chapter 1

Introduction

Modern and future mobile devices must support increasingly more wireless standards and frequency bands. A current smartphone, for example, generally includes wireless functionality for not only cellular standards, but also Wi-Fi, Bluetooth, GPS, and depending on global region, possibly also FM radio and mobile television. Furthermore, there can be a myriad of different standards and bands even within each wireless application. Wi-Fi can include 802.11a/b/g/n standards and both 2.4GHz and 5GHz frequency bands. Cellular functionality must include the various 2G/3G/4G standards, and 4G LTE alone has more than 40 frequency bands globally spanning 700MHz to 6GHz.

Fig. 1.1 illustrates the typical front-end architecture of a modern smartphone [1]. As shown, multi-standard, multi-band co-existence is currently enabled by a multitude of discrete off-chip components. These include external low noise amplifiers (LNAs) and power amplifiers (PAs), transmit/receive (T/R) and band switches, duplexers, diplexers, and RF filters.

The discrete components use expensive materials, such as GaAs and SiGe processes or mechanical surface-acoustic wave (SAW) devices, to achieve the necessary loss, noise, and filtering performance specifications that could not be attained in bulk silicon CMOS. They are targeted towards specific bands and applications, generally narrowband, and not reconfigurable. Furthermore, they are bulky and expensive. As transceivers are required to accommodate an increasing number of wireless bands, the required number of discrete components increases accordingly, resulting in greater bill of materials (BoM) cost and printed circuit board (PCB) area. Thus, current commercial mobile devices are either regional, where only bands in their intended geographical region of operation are supported, or they support few bands from each region.

The concept of a reconfigurable radio refers to a wireless system that is multi-standard, multi-band, and self-adapting to utilize available bands in the most spectrum efficient and energy efficient manner. The ideal reconfigurable radio differs from traditional transceivers in two fundamental ways. First, it should be able to reconfigure itself to transmit and receive
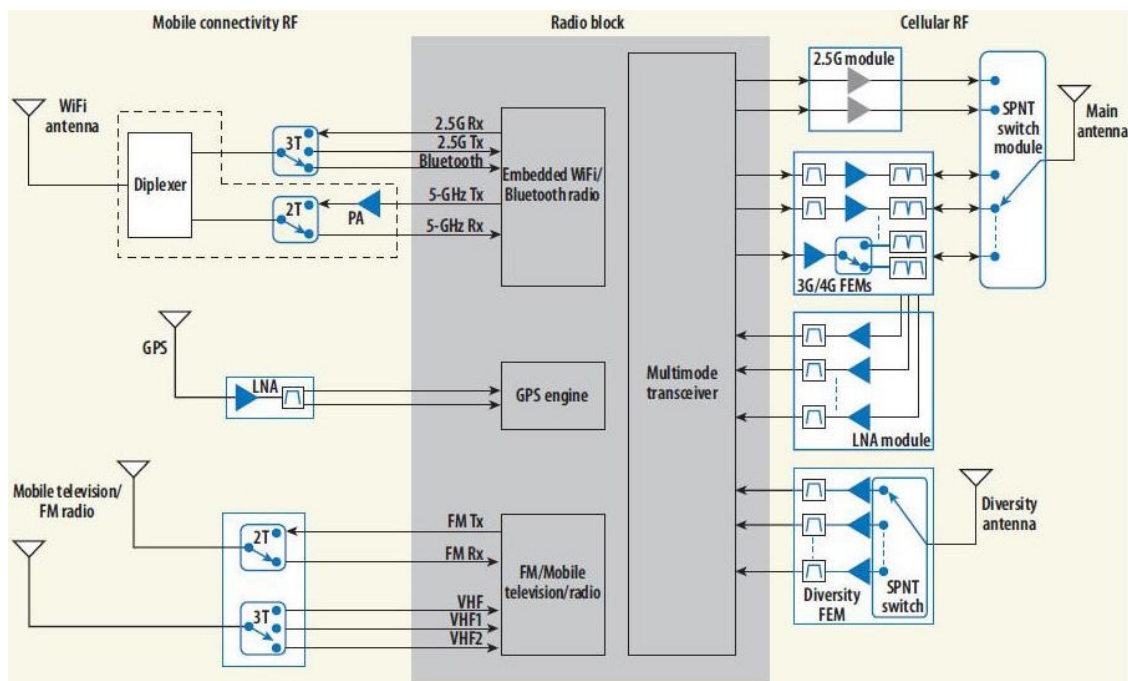
Figure 1.1: Front-end architecture of a typical smartphone.

on a variety of frequencies, bandwidths, and standards. This mean eliminating inflexible, bulky, and narrowband discrete front-end components. Instead, there should be a wideband and fully-integrated front-end, which allows for the flexibility and the efficient software or digital control that exemplifies reconfigurability.

Thus, a first major step towards the realization of a reconfigurable radio is to develop techniques to enable wideband front-end integration. With a truly integrated wideband transceiver, the system becomes more scalable and cost effective. Transceiver systems can be empowered to support the myriad of current and future wireless bands across the globe without the constraints of exponentially increasing cost and area from discrete front-end components.

The second fundamental aspect of a reconfigurable radio is its cognitive ability and spectrum awareness. It must have knowledge of its spectrum environment in order to select the most efficient frequency bands to operate on, and the sensing of the spectrum must be fast and responsive to rapidly varying environmental factors. Spectrum sensing can be used for optimal channel selection and dynamic blocker suppression, and spectrum awareness enables energy efficiency by liberating the transceiver from having to always operate assuming worst case conditions. Since spectrum sensing is an auxiliary function separate from the core transceiver, it should, in addition to being robust and fast, add low complexity and low overhead to the system.

Moreover, spectrum sensing and cognitive capabilities have also been proposed for spectrum re-use, where unlicensed users re-use licensed bands when their primary users are not present [2]. This scheme takes advantage of the fact that no band is being used everywhere at all times, and spectrum utilization can be significantly increased by allowing unlicensed users onto licensed but sparsely used bands. If every band can be filled to near capacity in this way, that could potentially results in an order of magnitude increase in spectrum efficiency. In a spectrum re-use scenario, however, non-interference with primary licensed users must be guaranteed. Thus, robust spectrum detection with high sensitivity would also be required in addition to rapid sensing time and low overhead.
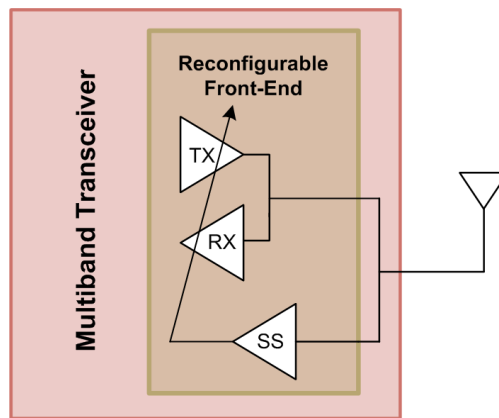
Figure 1.2: Conceptual diagram of reconfigurable radio front-end.

Fig. 1.2 illustrates a reconfigurable radio front-end conceptually. A multi-band transceiver with wideband, fully-integrated RF front-end connects directly to the antenna with no discrete, fixed-band components, while an efficient, low power spectrum sensing receiver provides real-time environmental knowledge to enable self-adaptation and dynamic reconfigurability.

1.1 Scope of Work

This work focuses on two specific aspects of the broad goal towards reconfigurable radios outlined above. First, we investigate optimal circuit and architectural techniques to perform high sensitivity spectrum sensing with minimal complexity and overhead. Our analysis of spectrum sensing specifically targets the ultra-high frequency (UHF) TV band, as it is one of the few bands in which the operation of unlicensed cognitive radio devices has been approved by the Federal Communications Commission (FCC) [3]. We design and implement a high-sensitivity, low power spectrum sensing receiver for UHF TV band applications as a demonstration of the feasibility of robust integrated sensing for mobile applications.

In the second part of this work, we develop a novel integrated T/R switching technique. T/R switches make up a prominent segment of discrete front-end components, and integrated wideband solutions with comparable performance have not yet been demonstrated. We propose an integrated T/R switching method in which the PA can be transformed into and re-used as an LNA, and as proof-of-concept for the proposed technique, we design and implement a wideband integrated time division duplex (TDD) front-end utilizing the PA re-use switching technique. We demonstrate the feasibility of wideband, integrated T/R switching as a stepping stone towards greater front-end integration and reconfigurability.

1.2 Thesis Organization

The following chapters of this thesis are organized as follows:

Chapter 2 describes the development of a spectrum sensing receiver for TV band applications. This chapter first reviews the characteristics of the UHF TV band and the FCC requirements for cognitive sensing in the band. Next, an analysis of detection methods and past works are presented. Finally, we describe the design and present the measured results of the implemented spectrum sensing receiver.

Chapter 3 presents the fundamental concepts of T/R switching and proposes an innovative technique for integrated wideband T/R switching. This chapter analyzes the constraints and challenges of T/R switches and presents a survey on current state-of-the-art. We then describe our proposed T/R switching scheme, in which the PA is re-used as an LNA, and how this switching method alleviates the drawbacks of existing techniques.

Chapter 4 describes the design and implementation of a wideband TDD front-end utilizing the proposed PA re-use T/R switching technique. Particular focus is put on the design of a shared PA/LNA front-end transformer, the method of PA/LNA transformation in the shared PA/LNA core, and the power and mode switches used to enable PA/LNA transformation. We then describe the architecture of the full implemented transceiver, some considerations for PCB design, and finally, measured results from the implemented chip.

Chapter 5 summarizes this work and discusses possible future directions.

Chapter 2

Low-Power Spectrum Sensing for Cognitive Radio Applications

With the growth of mobile data usage and as users crowd existing cellular bands, the FCC has been experimenting with cognitive radio operations in an effort to improve spectrum utilization and efficiency. Fig. 2.1 illustrates conceptually how a cognitive radio would operate in time and frequency domains [4]. At fixed time intervals, shown in yellow, the cognitive radio would scan the spectrum and identify idle bands, shown in white. It would then operate on those bands, shown in blue, until the return of a primary user has been detected. When this happens, the cognitive radio would vacate the re-occupied band for other bands that are now idle. Since cognitive radios are unlicensed devices operating in licensed bands, spectrum sensing is essential in guaranteeing non-interference with the activities of the bands' primary users.
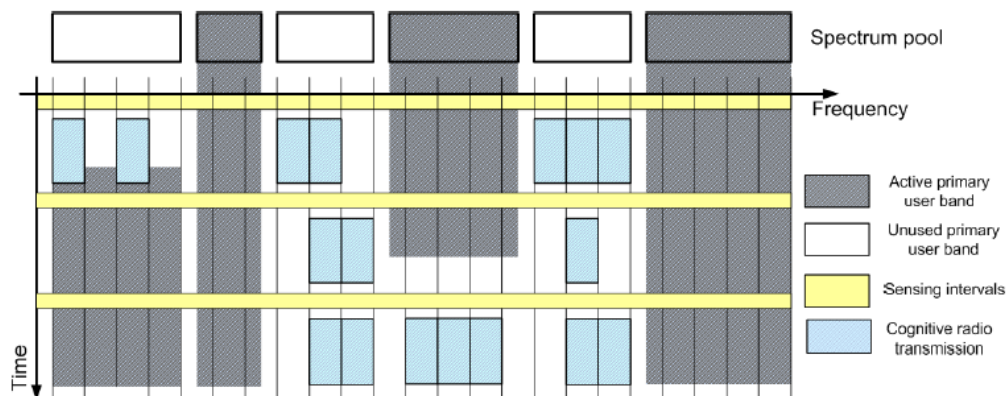
Figure 2.1: Cognitive radio operating in time and frequency domains.

In 2008, the FCC approved the operation of unlicensed cognitive radio devices in the TV bands, with unlicensed mobile devices approved in the UHF TV band [3]. However, two

years later, the FCC dropped the spectrum sensing requirement for unlicensed TV band devices (TVBDs) due to concerns over its viability, and instead ordered TVBDs to identify idle bands by accessing a geolocation database [5]. One major contribution to the FCC's decision to drop spectrum sensing requirements for TVBDs is the failure thus far of any system to meet the detection specifications set by the FCC.

During a demonstration of industry prototypes conducted in 2008, no system achieved the necessary detection sensitivity of -114 dBm in real world scenarios [6]. In academic research, on the other hand, while there has been much work done on theory topics like detection algorithms and cooperative sensing, there has been a lack of work on physical front-end designs for spectrum sensing applications. Some works may embed sensing functionality into receiver designs, but since the sensitivity required for sensing is much higher than that of conventional receivers, these works all fail to achieve the necessary detection sensitivity as well. Thus, a system targeted specifically for spectrum sensing achieving high detection sensitivity with good dynamic range and low power remains elusive.

Spectrum sensing offers much greater flexibility and lower overhead than a geolocation database or other cognitive radio management methods, It will play a role not only in the future of cognitive radios in the TV band, but also in the path towards general purpose reconfigurable radios. In fact, spectrum sensing functionality is already under consideration in the preliminary standards for 5G. To demonstrate the feasibility of high sensitivity, low power spectrum sensing in mobile devices going forward, we present in this chapter a spectrum sensing receiver for TV band cognitive radio applications.

2.1 Spectrum Sensing in the TV Band

In this section, we first look at the characteristic of the UHF TV band in Sec. 2.1.1 and discuss the IEEE 802.22 standard for TV white space sensing. Next, Sec. 2.1.2 surveys existing work that has been done on spectrum sensing for mobile applications. Finally, Sec. 2.1.3 analyzes an assortment of spectrum sensing techniques and their applicability to our application.

2.1.1 UHF TV Band and IEEE 802.22

In the United States, the UHF TV band currently spans from 470MHz to 698MHz, and it is composed of 38 6MHz-wide channels. After the full transition to digital TV (DTV) was completed in 2009, all channels became purely digital, adhering to the Advanced Television Systems Committee (ATSC) standard.

According to the ATSC standard, terrestrial DTV uses 8-level vestigial sideband (8VSB) modulation. Within each 6 MHz channel, data power is spread over the center 5.38MHz bandwidth, giving a 10.76MSymbol/s datarate. Each channel has two 610kHz-wide tran-

sition regions, one on each edge, from the response of the root raised cosine filters [7]. In
addition, each channel also has a pilot tone at the frequency of the buried carrier, 310kHz
away from the lower channel edge. The pilot, within its narrow 10kHz bandwidth, has a
power that is 11.3dB lower than the total data power of the 6MHz channel [7]. Fig. 2.2
illustrates the ideal and measured spectrums of a DTV channel.
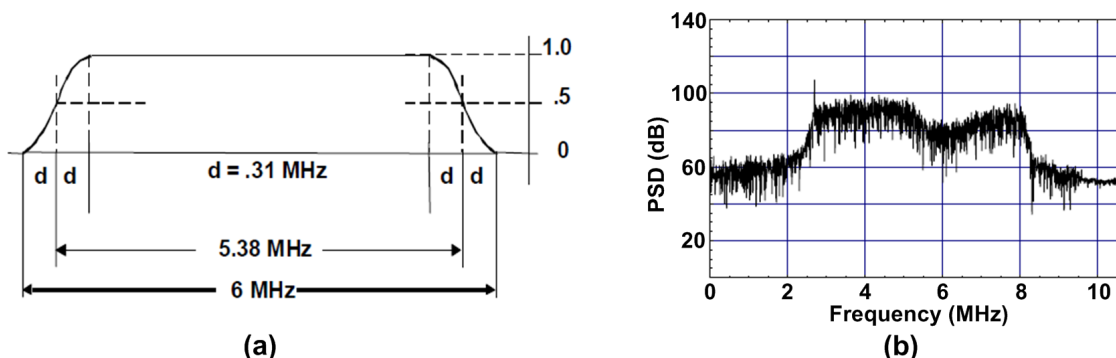
Figure 2.2: Spectrum of a DTV channel: (a) ideal (pilot not shown) [7], (b) measured [8].

The minimum signal-to-noise ratio (SNR) required for decoding an ATSC signal is set at
15.2dB. With a maximum adjacent channel splatter of -46.5dBc on the transmit side and
a margin of about 4 dB, the adjacent channel blocker power is limited to a maximum of
27dB higher than the wanted signal power in order to maintain sufficient SNR in the wanted
channel. The maximum blocker power grows to 44dB higher for the N+/-2 channel, and
increases by 4dB with each additional channel spacing until the N+/-6 channel [7]. The
overall ATSC blocker profile for weak signals is illustrated below in Fig. 2.3.
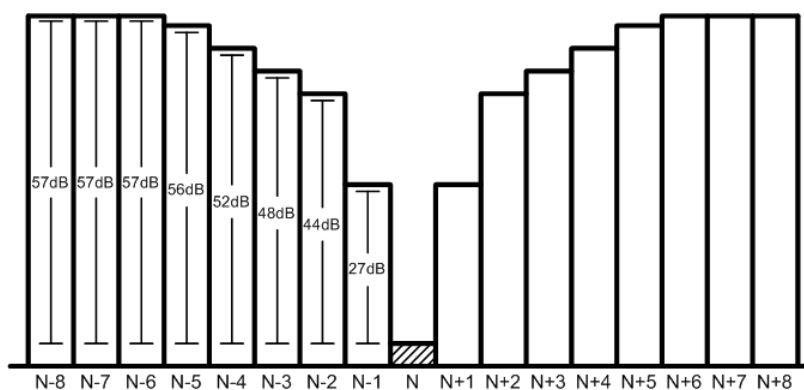
Figure 2.3: ATSC blocker profile.

The thermal noise in a 6MHz channel is about -105dBm. Given the required decode SNR of
15.2dB and a margin for receiver noise figure, the minimum sensitivity for DTV receivers,

or the minimum input energy of a decodable signal, is set at -83dBm [9]. However, since the sensitivity requirement must be met under worst case test conditions, many TV tuners in practice have higher sensitivity under common, realistic operating conditions.

In addition to DTV itself, wireless microphones also operate in the UHF band in vacant TV channels. There is no homogenized standard for wireless microphones, but they are usually narrowband, frequency-modulated (FM) signals with a bandwidth of 200kHz [8]. Although wireless microphones are also unlicensed devices, they have priority over TVBDs and are considered primary users.

The IEEE 802.22 standard governs Wireless Regional Area Networks (WRAN), which were developed to operate cognitively in the TV bands. The 802.22 work group, tasked with developing the standard, studied the geographical characteristics of base stations and primary receivers, and the interference and fading characteristics of wireless TV signals. The group then set sensing requirements for TVBDs to ensure an acceptably low level of interference with primary receivers.

In its preliminary standard published in 2008, the 802.22 work group set required sensing sensitivity level to -114 dBm for DTV in its 6 MHz channel bandwidth, for analog TV in its 100kHz carrier bandwidth, and for wireless microphone in a 200kHz bandwidth [3]. The sensitivity levels assume 0dBi antenna gain, and that the receiver antenna is outside and at least 10m above ground, relatively free of obstructions. Since analog TV transmissions no longer exist in the U.S. and wireless microphones are under consideration to be moved into its own dedicated channels, we will focus only on DTV sensing in this work.

The challenge of spectrum sensing arises mainly from the -114 dBm sensitivity requirement, which is much lower than that of a standard receiver and is significantly below the -105dBm thermal noise floor in a 6MHz channel. Furthermore, sensing must occur with a low power overhead for mobile applications, and within a reasonably short sensing time. IEEE 802.22 set the sensing interval to 2 seconds [10]; while this is a lengthy period, the actual sensing time should only be a small fraction since any time spent sensing cannot be spent transmitting and receiving on the detected idle channels. Additionally, IEEE 802.22 set both the probabilities of missed detection and false alarm to 0.1 [10]. Again, while these are relatively lax standards, it is in the interest of a cognitive radio system to minimize these probabilities for operational efficiency.

2.1.2   State of the Art and Related Works

During the development of its WRAN policy, the FCC solicited TVBD prototypes for laboratory and field testing. Five such devices, submitted by Adaptrum, Institute for Infocomm Research, Microsoft, Motorola, and Philips Electronics, were tested in 2008. All prototypes reliably detected -114dBm clean DTV signals in laboratory conditions; however, the sensitivities of the prototypes severely degraded, some up to 60-70dB and others to the point of

malfunction, when given real world conditions and the presence of interference [6]. The failure of the prototypes contributed to the FCC's ultimate decision to drop spectrum sensing requirements for TVBDs.

Aside from their less than desirable performance, the prototypes were designed as bulky and high-powered stationary devices analogous to traditional TV tuners. Detection was performed with power-hungry and time-consuming digital signal processing that put the shortest detection time at 0.1 seconds per channel, and the longer ones at tens of seconds per channel [6]. All of these factors make it evident that alternative approaches are necessary.

In the space of spectrum sensing for mobile devices, the first major work eliminated the need for complex digital signal processing as well as the preceding high-bandwidth, high-resolution analog-to-digital converter (ADC) by using analog correlation [11]. The work cross-correlated the downconverted receive signal with a window waveform, which effectively acted as an easily-tunable narrowband filter in the analog domain. However, since only energy detection was performed on the filtered signal, the sensitivity was limited by the noise floor to -74dBm, which is no better than the sensitivity of standard TV band receivers. Furthermore, the receiver in [11] consumed 180mW during both sensing and receive modes, whereas ideally, the sensing overhead should be much less than the power used for standard receiver operations.

The receiver front-end in a more recent work [12] was able to improve upon the performance of [11], achieving -84dBm detection sensitivity while consuming only 30-44mW. However, while [12] had a state-of-the-art, well-designed RF front-end that achieved good linearity, noise figure, and harmonic rejection with low power, nothing separate or different was dedicated to spectrum sensing functionality save the design of the received signal strength indicator (RSSI) block itself. Thus, the detection sensitivity, while an improvement, was still constrained by the noise floor and no better than that of standard receivers.

The work in [13] proposed an architecture that cross-correlates the received signal from two identical, parallel front-end paths. This results in uncorrelated front-end electronic noise that could ideally average out to zero, effectively lowering the receiver noise figure without harming linearity or dynamic range. The receiver achieved a 4dB noise figure (corresponding to an estimated detection sensitivity of -110dBm) and 89dB of spurious-free dynamic range (SFDR) in a 1MHz bandwidth. This performance, however, came at the cost of high power, due to the use of two full analog front-ends as well as complex digital signal processing (DSP) performing fast Fourier transforms (FFTs) and correlation. Not even including the power of two 14-bit ADCs, the receiver consumed 191mW, which is much too high for mobile applications.

In addition, the work in [13] relied on off-chip discrete components and signal processing in software. Only the RF front-ends of the two receivers were integrated [14]. The analog baseband circuitry was implemented with discrete PCB components, while the ADC and

DSP functions were performed in software.[1] A fully integrated spectrum sensing receiver with high sensitivity and good dynamic range has therefore not yet been demonstrated.

The work in [15], in contrast to previous works, sought to demonstrate that spectrum sensing using digital signal processing could be performed in a fast and low power manner. The digital baseband processor in [15] achieved a detection sensitivity of -115dBm with a detection period of 50ms while consuming only 7.4mW. However, in order to obtain this sensitivity, the processor required digital inputs with 20 bits of precision. Thus, while the processing power itself is low, the power that would be consumed by the RF front-end to provide this 20-bit high dynamic range input would certainly be prohibitively high.

2.1.3 Detection Techniques

There have been several proposed techniques to detect the presence of a signal or lack thereof. These methods have been analyzed extensively in theoretical literature, but we now evaluate them for their feasibility and applicability to our specific goal of sensing in the TV band.

The fastest and simplest detection method is energy detection, where we merely measure the received power. This can be done non-coherently, with knowledge of only the center frequency and the bandwidth of the target signal, and without any complex algorithms or processing. Simple energy detection, however, is incapable of distinguishing signal power from noise power, and hence it fails to detect in highly negative SNR regimes where signal is buried far under noise. For example, [16] evaluated the effectiveness of energy detection on a 4-MHz-wide QPSK signal, and the results are shown in Fig. 2.4.
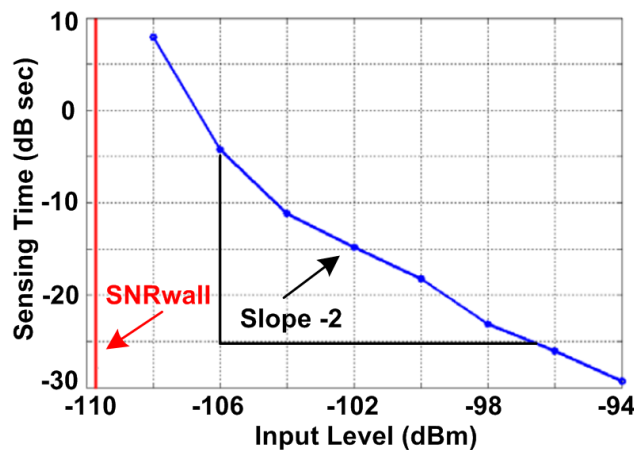
Figure 2.4: Energy detection on a 4 MHz-wide QPSK signal [16].

---

[1]The power consumption reported in [13] included off-chip components and estimated power for DSP processes. The integrated RF front-ends alone consumed 41-66mW [14].

The authors in [16] looked at sensing time versus input signal power with a constant false
alarm rate of 5% and a detection rate of 60%. We see in Fig. 2.4 that even using this
extremely lax probability of detection, there exists an SNR wall below which detection is
impossible given infinite sensing time. Furthermore, the sensing time remains prohibitively
long for at least 10dB above the SNR wall. Thus, alternative detection methods are necessary
if we seek to sense down to the -114dBm level.

Since we are targeting sensing specifically at ATSC signals, we can take advantage of certain
characteristics specific to ATSC signals. One such characteristic is the sinusoidal pilot tone.
As shown in Fig. 2.5, the pilot in its narrow bandwidth has a higher SNR than the signal of
the entire channel. Since the pilot is at a known position within the channel, we could simply
narrow the bandwidth of detection to just that of the pilot, and gain higher sensitivity with
energy detection.
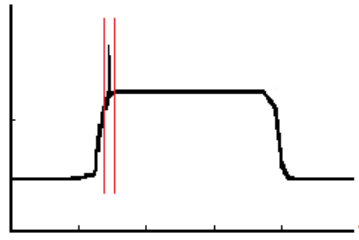
Figure 2.5: ATSC pilot tone.

However, while pilot energy detection can achieve a higher sensitivity, it still has the same
fundamental noise-floor limitations as channel energy detection. In addition, when dealing
with such a narrow bandwidth, frequency offsets can be problematic and sharp filtering
that approximates a brick-wall response is difficult to realize. These issues will all degrade
the sensitivity further from the theoretical limit, making other, more robust approaches
necessary.

A more robust detection method, cyclostationary feature detection, takes advantage of the
fact that human-made, modulated signals are fundamentally different from noise in that they
usually exhibit some kind of periodic behavior, either from carrier tones, cyclic prefixes, or
other features. This periodicity can be extracted using a spectral correlation function (SCF),
which measures the density of correlation between all the spectral components in a signal [4].
The SCFs for white noise and for a QPSK signal are show below.

As shown in Fig. 2.6, white noise is only correlated at identical frequencies, while the spectral
components at different frequencies of a modulated signal are also correlated, creating a
distinguishable pattern in its SCF. However, feature detection presents many implementation
challenges, the foremost of which is the complex processing required to generate the two-
dimensional FFT and calculate the correlation functions. Complex processing indicates high
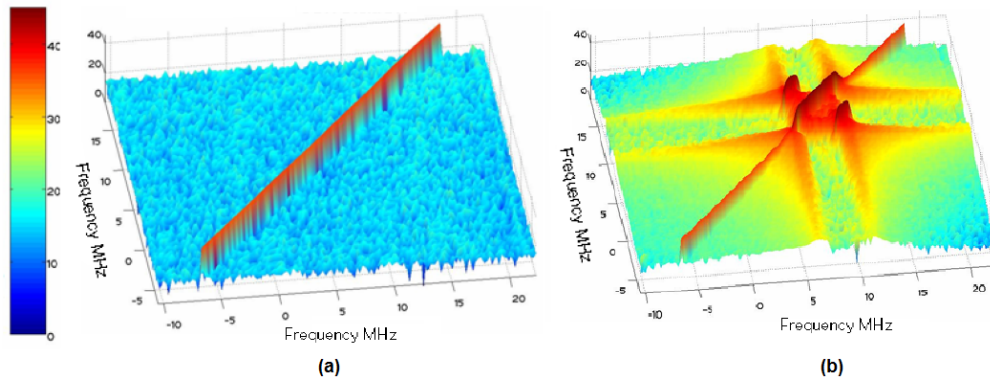
Figure 2.6: SCF of (a) white noise, and (b) a QPSK signal [4].

power and long sensing time, both of which makes this technique unattractive for spectrum
sensing in the mobile space.

Another simpler approach that can be used to extract periodicity is to autocorrelate the
signal. Autocorrelation, correlating a signal with a time-shifted version of itself, is a method
that has long been used in spectrum analyzers, radio astronomy, and other applications that
require high-sensitivity signal detection. Eq. 2.1 below presents the autocorrelation function
$R(\tau)$, where $X(t)$ is the signal being correlated and τ is the time shift.

$$R(\tau) = E\left[X(t) * X(t + \tau)\right] \tag{2.1}$$

Since white noise is a stationary process that is uncorrelated with itself at different points
in time, $R(\tau)$ for white noise would be 0 everywhere except for when $\tau = 0$. For modulated
signals which are cyclostationary processes, the signal time-shifted by an integer number of
periods should have the same statistical properties as the non-time-shifted version of the
signal. Therefore, $R(\tau)$ should exhibit periodic behavior, with peaks occurring at τ equaling
0 as well as each integer period. Fig. 2.7 illustrates this property.

If we look at $R(\tau)$ with τ equaling a non-zero integer period, the additive noise power would
ideally average out to zero given enough time, and the resulting SNR of the autocorrelation
function would be an improvement over the SNR of the signal itself. However, a digital
implementation of autocorrelation still requires a relatively high amount of processing due
to the multiplication operations needed to calculate the correlation function. As mentioned in
Sec. 2.1.2, [13] demonstrated the power inefficiency of this approach. On the other hand, [11]
demonstrated that correlation can be done in a lower power manner in the analog domain.
While the system in [11] was high powered, much of it was used for generating the window
waveforms, and the actual correlation operation itself consumed only 8mA. We thus see
that autocorrelation in the analog domain could potentially be a low-power, high-sensitivity
solution to spectrum sensing.
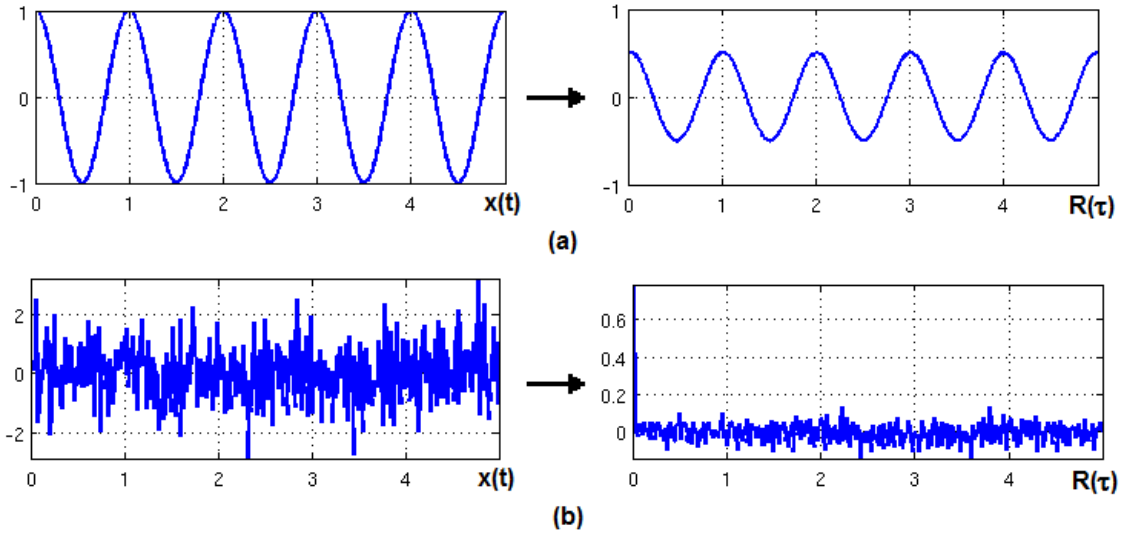
Figure 2.7: Time-domain waveform and autocorrelation function of (a) a sinusoid and (b) white noise.

2.2 Evaluation of a Multi-Mode Detection System

An optimal spectrum sensing system should use a combination of channel energy detection, pilot energy detection, and autocorrelation. The pilot tone, as a pure sinusoid, provides an ideal candidate to extract the periodicity from using autocorrelation. The channel energy detection mode would be used for a fast, coarse scan of all the channels to eliminate ones that have strong signals residing within. Pilot detection and autocorrelation, being more time consuming, would only be used on a few likely-vacant channels where the signal, if present, is too weak to be sensed with energy detection.

To evaluate achievable performance of the various sensing methods in a practical implementation, we constructed in simulation the multi-mode detection system shown in Fig. 2.8. We adopt the equivalent-time sampling technique from [17] to implement autocorrelation in the time domain. The input f_{RF} is captured in parallel by two samplers, with the sampling clocks offset from each other by a time delay τ. By correlating the two sampled signals, and by varying τ in discrete steps, the autocorrelation of the input is captured as a function of τ.

Using the equivalent-time sampling autocorrelation method, the resolution of the sampled signal is determined not by the actual sampling frequency but by the minimum τ-step, with effective sampling frequency equaling $1/\tau_{step}$. Thus, this method enables power savings by using a sampling frequency below the Nyquist rate. Since subsampling also acts as a mixer and downconverts the RF signal to baseband, the signal is then immediately channel-
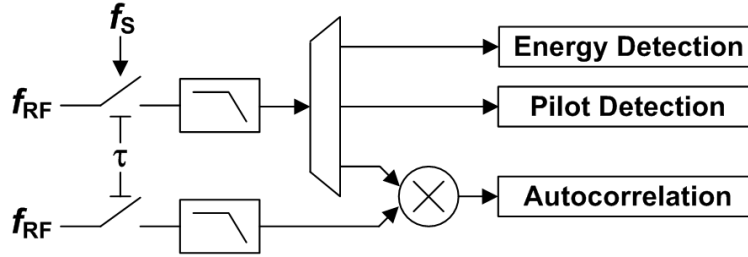
Figure 2.8: High level block diagram of proposed system.

filtered following the sampler.[2] In the top path, the filtered channel signal undergoes energy detection and pilot detection. For autocorrelation mode, the signals from the two paths are multiplied together in the analog domain as τ, the time delay between the two paths, is varied.

Sec. 2.2.1 delves into the detailed mechanisms of subsampling and autocorrelation, and how these techniques apply to our spectrum sensing system. Then, Sec. 2.2.2 describes a Matlab-based simulation platform built to evaluate the proposed multi-mode detection system. Finally, Sec. 2.2.3 presents the theoretical performance of the proposed system in each of its detection modes.

2.2.1 Subsampling and Autocorrelation

In traditional Nyquist sampling, the number of points per cycle that are sampled scales linearly with the ratio of sampling frequency to signal frequency. Thus if we want to sample many points per cycle in order to not lose signal fidelity through the sampling process, we must oversample the signal with a much higher sampling frequency. If the input frequencies are in the hundreds of megahertz, oversampling becomes very power-inefficient and unattractive. Subsampling, on the other hand, maintains high signal fidelity while using lower sampling frequencies.

Fig. 2.9 illustrates the subsampling operation conceptually. The signal is sampled once per period for subsampling ratio (SR) of 1, and once every two periods for SR of 2. Each sample occurs at a different temporal location within the signal period, and if given enough signal cycles, the entire characteristic of a periodic signal will be sampled. The result is a high-fidelity signal with numerous sampled points per cycle, albeit translated to a a lower frequency.

Based on the number of samples per cycle, the equivalent oversampling frequency f_{S-EQ} of

---

[2] "Subsampling" refers to sub-Nyquist sampling with $f_S < 2 * f_{RF}$. It is used interchangeably with "equivalent-time sampling" in this work."
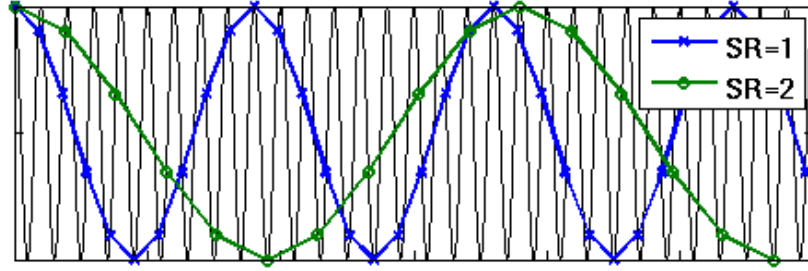
Figure 2.9: Illustration of subsampling with subsampling ratios of 1 and 2.

a subsampled signal is:

$$f_{S-EQ} = \frac{f_{RF} * f_S}{f_{RF} - N * f_S}, \qquad 0.5 * f_{RF} < N * f_S < f_{RF} \tag{2.2}$$

where f_S is the subsampling frequency, f_{RF} is the signal frequency, and N is the subsampling ratio. An extremely high f_{S-EQ} can be obtained with an extremely low f_S if N is large. More importantly, the linear relationship between samples per cycle and sampling frequency has been broken: a higher f_{S-EQ} can be achieved by simply minimizing $(f_{RF} - N * f_S)$.

The translated lower frequency after subsampling is:

$$f_{sampled} = f_{RF} - N * f_S, \qquad 0.5 * f_{RF} < N * f_S < f_{RF} \tag{2.3}$$

We can easily choose N and f_S so that $f_{sampled}$ falls in the baseband, providing a convenient mixing step as well. This frequency aliasing effect, however, also causes noise folding, in which the noise around every harmonic of f_S is translated to baseband as well. The mechanism of noise folding is illustrated in Fig. 2.10. Thermal noise is represented in green, and we assume a low-pass filter has attenuated high-frequency noise above the desired signal (in red in Fig. 2.10(a)).
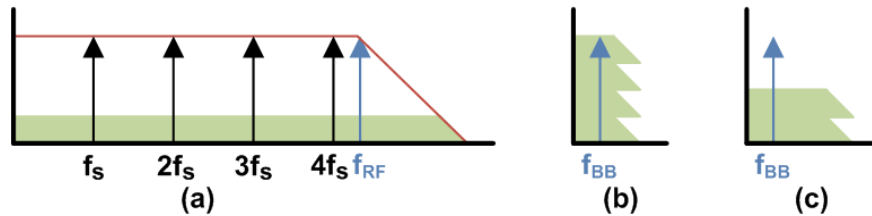
Figure 2.10: Noise folding from subsampling: (a) unsampled spectrum, (b) sampled with f_S, (c) sampled with $2f_S$.

Fig. 2.10(b) demonstrates the sampled signal (f_{BB}) using sampling frequency f_S and $SR = 4$. Noise from each harmonic accumulates in the same frequency band, and the signal at baseband has severely degraded SNR. Bandpass anti-alias filters can be used at RF to mitigate

noise folding, but the noise generated from the sampler itself still has a low-pass charac-
teristic and will not be filtered. The folding of sampler noise has resulted in subsampling
mixers generally having prohibitively high noise figures, making them unattractive for low
noise systems [18]. Alternatively, baseband SNR can be improved with a lower subsampling
ratio, such as $SR = 2$ in Fig. 2.10(c). The tradeoff is higher power consumption associated
with higher-frequency clock generation and distribution networks.

Spectrum sensing systems, in order to achieve the required high sensitivities, must be low
noise. We therefore choose for our system the low subsampling ratio of 1, so that noise
is folded only once. The noise is also spread over the full spectrum up to the sampling
frequency and is not concentrated purely in the baseband, which further improves noise
figure. While the power benefits of subsampling are lessened with $SR = 1$, it is still a factor
of 2 improvement over the minimum Nyquist sampling frequency, and more critically, we
still gain the benefit of higher f_{S-EQ}.

Next, we consider the options for choosing f_S. Standard receivers would set the mixing
frequency to the carrier frequency, and use I and Q paths to reject aliased images. A Hilbert
filter would then reconstruct the real baseband signal from the downconverted vestigial
sideband in order to allow signal decoding [19]. However, since the pilot also falls on the
carrier frequency, this provides several problems for our purposes. Firstly, the pilot power, as
a DC component, becomes vulnerable to DC offset and low frequency flicker noise. Moreover,
the pilot is in a 10kHz bandwidth, indicating that a lengthy time will be required to collect
enough samples for an accurate measurement.
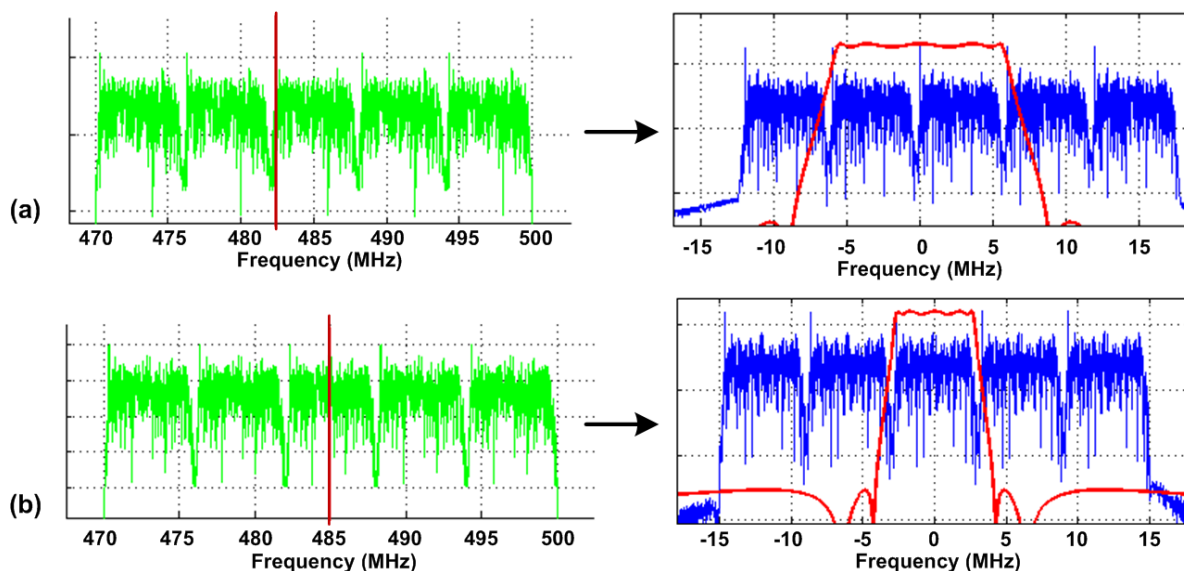
Figure 2.11: Conversion to baseband with sampling frequency set to (a) carrier frequency,
(b) channel center frequency.

A better solution, shown in Fig. 2.11, is to sample at the center frequency of each channel instead of at the carrier. This results in true direct conversion where image rejection and I and Q paths are no longer necessary, as no out-of-channel signals would fall in-band. Without image rejection and Hilbert filters, the signal would fold onto itself to create an un-decodable baseband signal. However, since we only seek to detect the presence of a signal and not to receive it, a non-real signal is acceptable for our system. In addition, the pilot falls at the low IF frequency of 2.69MHz. This alleviates the flicker noise problem and eliminates the DC offset problem altogether. Pilot power can also be measured more rapidly, and the pilot retains its periodicity to allow autocorrelation detection.

To perform autocorrelation in analog time domain, we step the time shift τ between the two samplers. Fig. 2.12(a) shows an example sinusoid signal, and Fig. 2.12(b) shows 5 τ-steps encompassing 1 cycle of its autocorrelation in time domain. In post-processing, the mean at each τ-step would be calculated to obtain autocorrelation as a function of τ. Fig.2.12(b) plots $R(\tau)$ in blue.
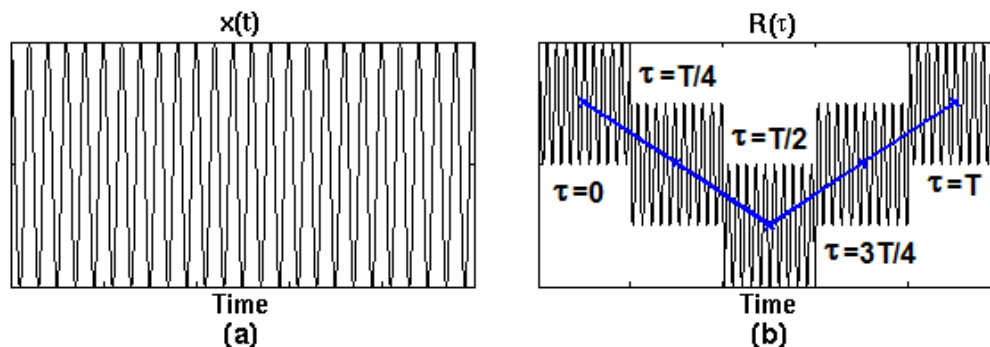
Figure 2.12: Time-domain waveform of (a) a sinusoid, (b) its autocorrelation.

Fig. 2.13 demonstrates the extraction of periodic autocorrelation behavior from a noisy signal. Fig. 2.13(a) plots the time domain signal, in which a sinusoid has been buried in noise and cannot be distinguished by energy detection. Its autocorrelation in Fig. 2.13(b), on the other hand, extracts the periodicity of the signal and demonstrates clear SNR improvement.

Stepping through multiple autocorrelation cycles as in Fig. 2.13(b) is useful if we eventually want to extract frequency information through a FFT; however, it is unnecessary for our purposes. Since the pilot is a pure sinusoid, one autocorrelation cycle is adequate to gain its entire characteristic and to make a detection decision. There should be no difference (statistically) between the autocorrelated results of, for example, $\tau = T/4$ and $\tau = 5T/4$. Thus, performance could be better improved by averaging at each τ-step for a longer period of time rather than incorporating more τ-steps and multiple autocorrelation cycles. This eliminates the need for long, variable delay lines, and phase-shifting the sampling clock is sufficient to create all necessary all τ-steps where $\tau < T$. Due to the properties of

subsampling, a phase shift in the sampling clock translates directly to an identical phase
shift in the subsampled baseband signal.
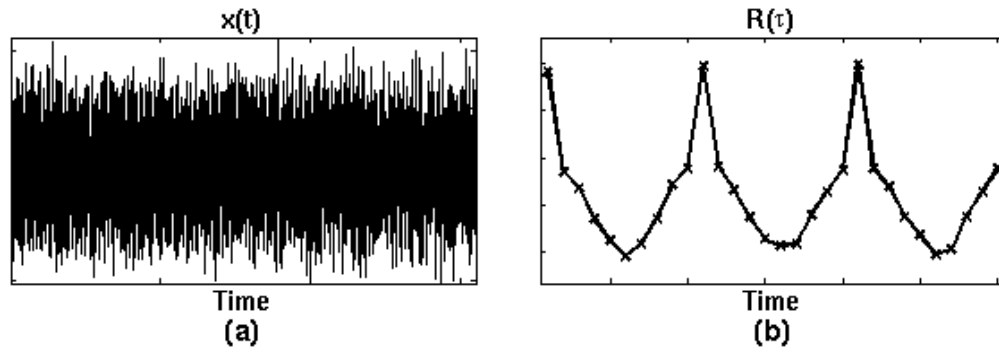
Figure 2.13: Time-domain waveform of (a) a noisy sinusoid, (b) its autocorrelation.

2.2.2   Simulation Setup

Realistic spectrum sensing scenarios using the multi-mode detection system were modeled
in Matlab. Fig. 2.14 illustrates how transmit signals were generated in the modeling envi-
ronment. A random stream of 8-level symbols in baseband are filtered to obtain the VSB
frequency response, and a pilot is then added as a DC level. Next, for every channel where we
desire a signal to exist, we scale the signal power as desired and up-mix it to its appropriate
RF channel position. Finally, additive white Gaussian noise (AWGN) is applied to model
environmental thermal noise.
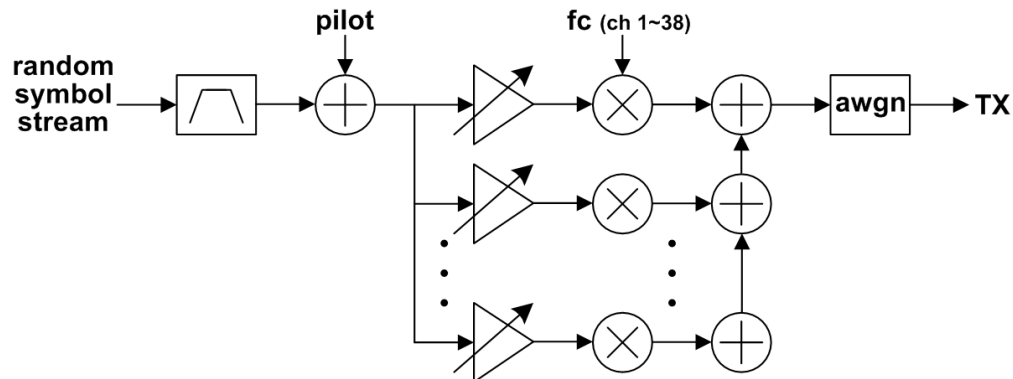
Figure 2.14: Transmit signal generation in simulation.

For example, a simulated model of the worst case sensing scenario is shown in Fig. 2.15.
The signal power (with additive thermal noise) is shown in blue and pure noise is shown in
green. Fig. 2.15(a) shows a 7-channel worst case blocker profile with a weak -114dBm signal

in the center channel. On the other hand, when a channel is idle like the center channel of
Fig. 2.15(b) is, the power of its surrounding channels can be arbitrarily high. We seek to
differentiate between the two scenarios by detecting the presence or lack thereof of the weak
signal in the center channel.
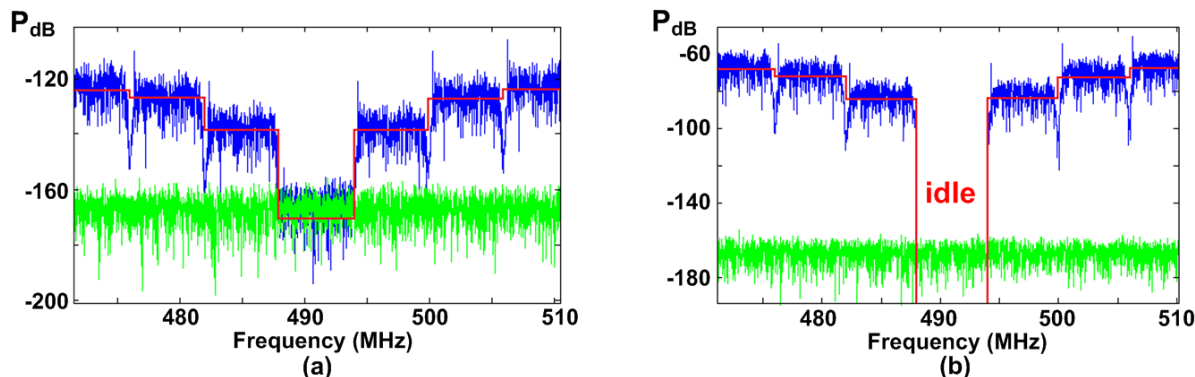
Figure 2.15: Worst case sensing scenario when target channel is (a) occupied, (b) idle.

Fig. 2.16 illustrates how the receiver is modeled in simulation for the measurement of detected
power. First, a coarse bandpass filter is applied to the entire UHF band to attenuate out-
of-band noise and thus reduce noise-folding during subsampling. We ignore out-of-band
interference and assume they have been sufficiently attenuated at this point in the receive
path. We then subsample the signal at the desired channel with a sampling frequency equal
to the channel frequency. The downconverted signal is filtered by either a channel filter or an
ideal narrowband pilot filter, and the resulting power of the signal is measured using square
law energy detection. For autocorrelation, we apply ideal mathematical autocorrelation to
the filtered pilot, and then again apply square law energy detection to measure the power of
the autocorrelated signal.
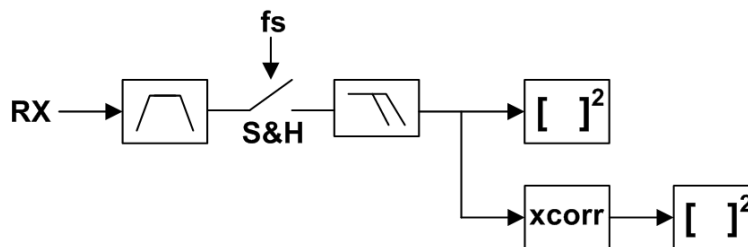
Figure 2.16: Receiver modeling in simulation.

To establish a metric to evaluate the performance of each detection strategy, we look at the
two scenarios in Fig. 2.17. In the first scenario, when no signal is present, the receiver will
detect only the power of the noise from the environment; we denote this quantity N_{DET}. For

simplicity, we assume pure thermal noise and ignore non-white interference at this junction. In the second scenario, when a signal is present, the receiver picks up that signal along with the environmental noise, and we denote the total detected power under this condition S_{DET}.
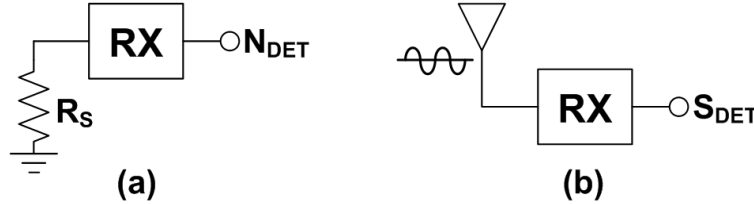
Figure 2.17: Equivalent block diagram when (a) only noise is present, (b) signal and noise are present.

If we denote the power of the input signal S_{IN}, power of the environmental thermal noise N_{THERM}, power of the receiver electronic noise at its output N_{RX-OUT}, and the total gain through the receiver A_{RX}, then for naive energy detection:

$$\begin{aligned}
N_{DET} &= A_{RX} * N_{THERM} + N_{RX-OUT} \\
S_{DET} &= A_{RX} * (S_{IN} + N_{THERM}) + N_{RX-OUT}
\end{aligned} \tag{2.4}$$

Given a receiver noise figure, we can define N_{RX-OUT} in terms of the input thermal noise and the noise factor F_{RX}:

$$N_{RX-OUT} = A_{RX} * N_{THERM} * (F_{RX} - 1) \tag{2.5}$$

The SNR of an input signal is the ratio of the signal power to the thermal noise power. We can similarly define a term "detected SNR" to be the ratio of S_{DET} to N_{DET}, or the ratio of power detected when a signal is present to power detected when only noise is present. Using equations Eq. 2.4 and Eq. 2.5, we find:

$$Detected\ SNR = \frac{S_{DET}}{N_{DET}} = \frac{S_{IN}/N_{THERM} + F_{RX}}{F_{RX}} = \frac{Input\ SNR + F_{RX}}{F_{RX}} \tag{2.6}$$

From Eq. 2.6, when the receiver adds infinite noise, the detected SNR reaches its lower bound of 1, or 0 dB. This indicates that inputs with the signal present and without the signal present are indistinguishable from each other at the point of detection. Conversely, when the receiver noise figure is very small, the detected SNR tracks with the input SNR.

2.2.3 Evaluation of Detection Modes

Using the simulation framework described in Sec. 2.2.2, we evaluate the theoretical performance of our spectrum sensing system in each of its detection modes. Simulated energy
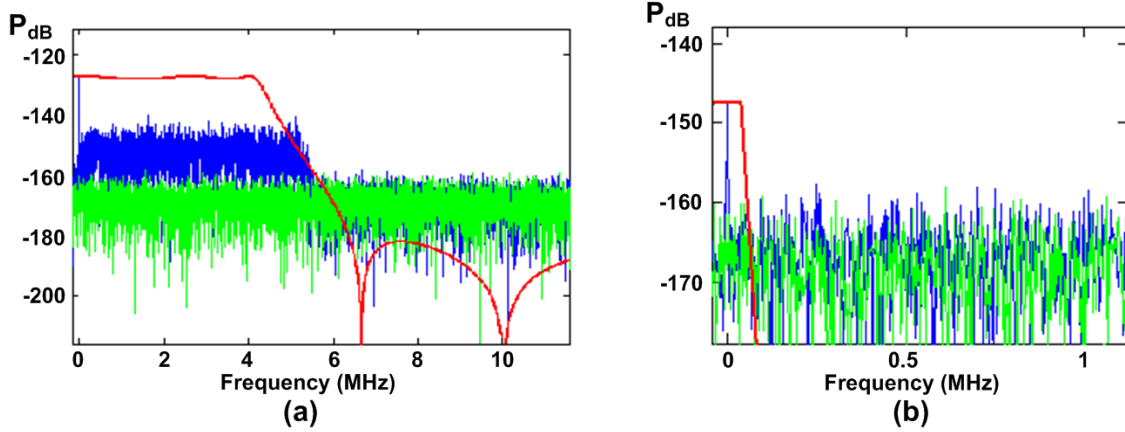
Figure 2.18: Simulated (a) energy and (b) pilot detection at baseband.

and pilot detection are shown in Fig. 2.18, where the power spectrum densities of signal and noise are shown in blue and green respectively. The red outlines illustrate example baseband filter responses.

In Fig. 2.18(a), where signal power is distinctively higher than noise power, we can perform energy detection by filtering the entire channel and measuring the resultant power. On the other hand, in Fig. 2.18(b), we see that for weak inputs, the signal power is completely buried in noise and indistinguishable from it. However, the narrowband pilot tone still rises above the noise floor. In this case, we can perform pilot detection by narrowing the filter response to just the bandwidth of the pilot, and then, again, measuring the resultant power.

To evaluate the performance and limitations of energy and pilot detection, we first note that thermal noise power is given by:

$$N_{THERM} = k * T * \Delta f \tag{2.7}$$

where k is Boltzmann's constant equaling 1.38×10^{-23}JK^{-1}, T is temperature in Kelvin, and Δf is the noise bandwidth under consideration.

At room temperature, kT gives -204dB/Hz, or -174dBm/Hz, of noise power. In a 6MHz bandwidth, the noise power becomes -105dBm, and an input signal at -114dBm has a -9dB input SNR. In a 10kHz bandwidth, the noise power lowers to -133dBm. However, the power of the pilot in its narrow bandwidth is 11.3dB below the total channel power. Thus, the minimum pilot power is -125.3dBm, making the minimum input SNR of the pilot 7.7dB.

Using Eq. 2.6, we can plot detected SNR as a function of input SNR for various receiver noise figures. Shown in Fig. 2.19, for an input SNR of -9dB for channel energy detection, the detected SNR in the ideal case (with no added receiver noise) is about 0.5dB. Similarly, the detected SNR in the ideal case for pilot energy detection is about 8.4dB.
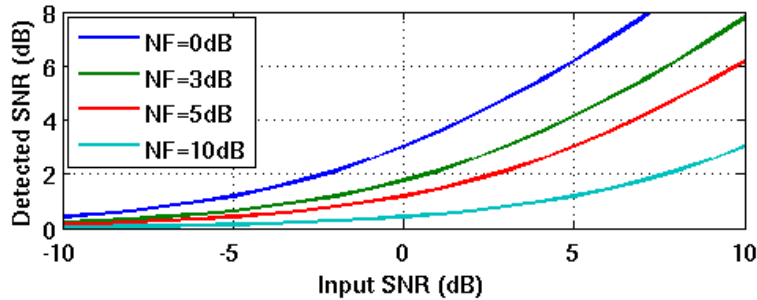
Figure 2.19: Detected SNR as a function of input SNR for energy detection.

Decision statistics and algorithms for spectrum sensing are outside the scope of this work;
however, we will briefly discuss at this juncture the factors that determine the detected SNR
threshold for sensing and their implications on system design. In theory, the probabilities
of false alarm (P_{FA}) and detection (P_D) are functions of where the decision threshold is set
and how many samples are used. More samples allow more averaging, resulting in a more
accurate measurement that is less sensitive to instantaneous variances in noise and signal
power. To find P_{FA} and P_D, we can model the probability density function (PDF) of noise
with a Gaussian distribution and use Neyman-Pearson hypothesis testing. This process is
illustrated in Fig. 2.20 [16].
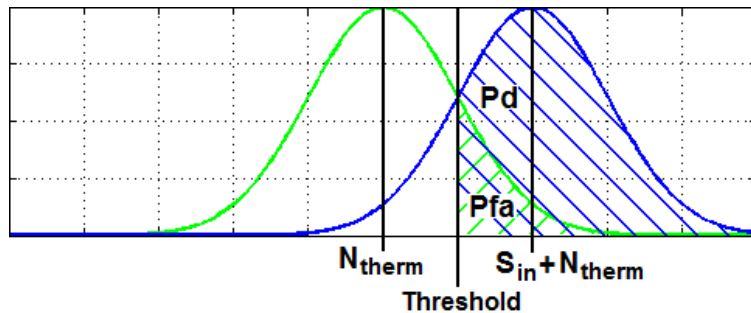
Figure 2.20: Probability density functions for signal and noise.

In Fig. 2.20, the green shows the PDF for pure noise, and the blue shows the PDF for signal
with noise, where the mean power has shifted but the shape remains Gaussian. The decision
threshold should be set somewhere between the two means, and anything above threshold in
the noise distribution results in a false alarm, while anything above threshold in the signal
distribution indicates a correct, desired detection. The shaded areas of P_{FA} and P_D can be
quantified as:

$$P_{FA} = Q\left(\frac{N*(SNR_{THR}-1)}{\sqrt{2*N}}\right) \tag{2.8}$$

$$P_D = Q\left(\frac{N*(SNR_{THR} - SNR_{DET})}{\sqrt{2*N*SNR_{DET}^2}}\right) \tag{2.9}$$

where SNR_{DET} is detected SNR, SNR_{THR} is the decision threshold detected SNR, N is the number of samples, and $Q(x)$ is the tail function for a normal distribution. For an ideal channel detection scenario, where the detected SNR is about 0.5dB, we gain the following P_{FA} and P_D as a function of N shown in Fig. 2.21. As expected, a higher threshold results in a longer detection time, but a lower false alarm rate. And when the decision threshold (in the form of detected SNR) is set to 0.7dB>0.5dB, the system fails to detect completely.
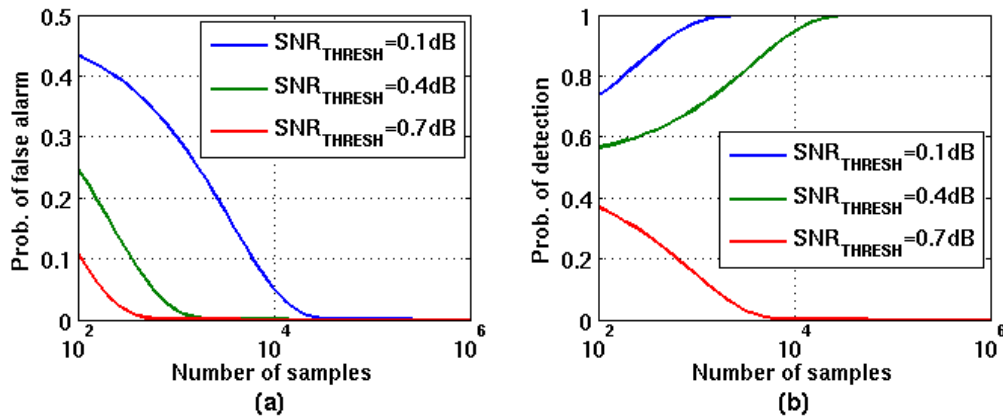
Figure 2.21: Ideal channel energy detection: (a) P_{FA}, (b) P_D.

While theory seems to have indicated that we can set a detection threshold that will sense the 0.5dB detected SNR with $P_{FA} < 0.1$ and $P_D > 0.9$, there are numerous problems in a realistic implementation that precludes this. The main limitation is the problem of noise uncertainty, where the actual noise power can differ greatly, more than several dB, from the expected, theoretical value due to process, temperature, and environmental variations. Periodically estimating noise using the detection system such as in Fig. 2.17 can alleviate the problem somewhat, ensuring that the system's own noise figure variations as well as gradually changing environmental conditions are accounted for. However, there would always be some residual estimation error, and other factors, such as interference profiles, can be rapidly time-varying. These variations in noise, and consequently in detected SNR, make pure energy detection impossible at -114dBm.

On the other hand, going back to Fig. 2.19, we see for pilot detection's 7.7dB input SNR, the detected SNR is about 8.4dB, which should provide enough margin for noise uncertainty. However, there are additional factors to take into account. Firstly, up to this point, we have only looked at the ideal case without added receiver noise. Furthermore, a real system has imperfect filtering, and all residual out-of-channel noise and signals become additive noise from the perspective of our sensing system. So for example, if we add in a standard receiver noise figure of 5dB, and make the conservative estimate that effective noise bandwidth is 2

times the desired signal bandwidth, then the input SNR lowers to 4.4dB and the detected SNR to 2.9dB. The detection margin has now become severely eroded. With possible additional erosions from non-linearities, interference, and wider effective noise bandwidths, a more robust detection method such as autocorrelation becomes necessary.

To understand why autocorrelation results in more robust detection, we consider the PDF of autocorrelated white noise. Since every value in the autocorrelation function is an average of numerous samples of noise, the variance of autocorrelated noise power should ideally be zero. This, in effect, makes the PDF of noise power approximate a Dirac delta function rather than a Gaussian. This behavior is demonstrated in Fig. 2.22(a). The PDF of white noise ("wn") and its autocorrelation ("corr") are shown, respectively, in green and red.
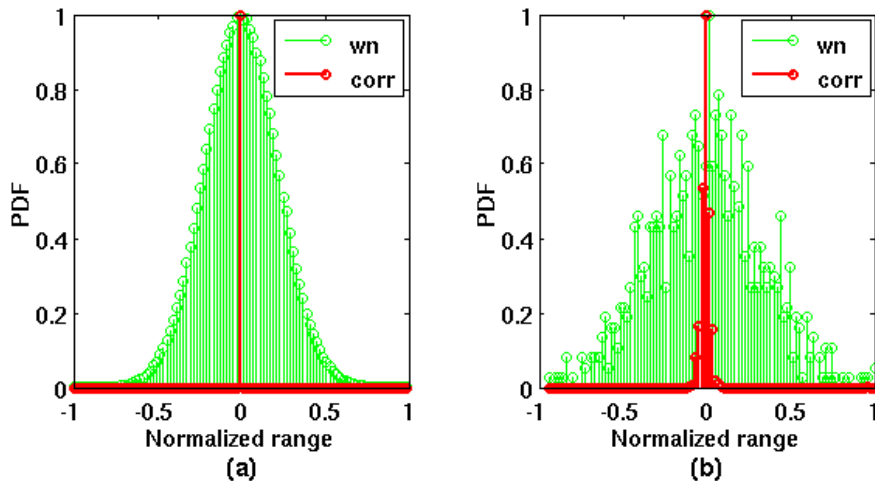
Figure 2.22: Normalized PDF of white noise and its autocorrelation with (a) 10^6 samples, (b) 10^3 samples.

When a signal is present, its additive thermal noise should also average out to zero when autocorrelated, while the signal itself retains its mean power as its periodicity is captured by and translates to the autocorrelated response. Thus, the PDF of the signal power should also approximate a delta function, but centered at the mean power of the signal. With two delta functions rather than two Gaussians, P_{FA} and P_D are obviously drastically decreased and increased, respectively, for the same threshold and mean power separation between S_{IN} and N_{THERM}. More robust sensing could therefore be achieved with lower detection margins.

In Fig. 2.22(b), with fewer number of samples, the PDF of the autocorrelation deviates from its ideal behavior and widens. However, since each autocorrelation point is already the average of many noise samples, this creates a layer of buffering against instantaneous variances in noise power. Therefore, the difference in the autocorrelation PDF is slight while the PDF of pure energy detection differs much more drastically from its ideal distribution, and the relative robustness of autocorrelation is maintained as the number of samples scale.

Using the multi-mode detection system and simulation framework, we evaluated the achievable sensitivity of the three detection modes. The results are plotted in Fig. 2.23. The system has been assumed to be ideal without additive electronic noise. Noise-folding effects from subsampling, however, are included in the simulation.
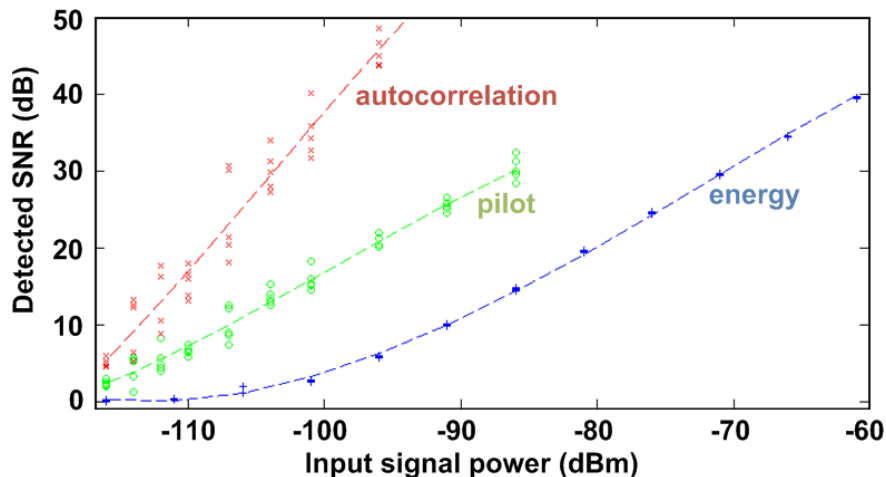
Figure 2.23: Detected SNR vs. input power for all detection modes.

As expected, for channel energy detection, the detected SNR tracks input power until the noise floor is reached at around -105dBm. Then, as signal power becomes much weaker than noise power, the detected SNR flattens to 0dB since a signal buried in noise has approximately the same power as pure noise, and energy detection can no longer differentiate between the two scenarios. Pilot detection never reaches its own noise floor at -133dBm, but at weak input signals, instantaneous variations bring detected SNR very close to 0dB. Finally, autocorrelation demonstrates improvements in detected SNR over the two other methods, and remain relatively robust at the weakest signal levels.

2.3 Design of a Dual-Mode, Correlation-Based Spectrum Sensing Receiver

Building off of the multi-mode detection system framework from Sec. 2.2, we designed a dual-mode, correlation-based spectrum sensing receiver for the UHF TV band. Fig. 2.24 illustrates a top-level block diagram of the system.

After the antenna, we first use an off-chip bandpass SAW filter to attenuate potential out-of-band blockers, especially from neighboring GSM bands at 800MHz to 900MHz. Next, a low noise amplifier (LNA) provides wideband input impedance matching as well as high gain, which is critical for low system noise figure. Following the LNA, another RF bandpass
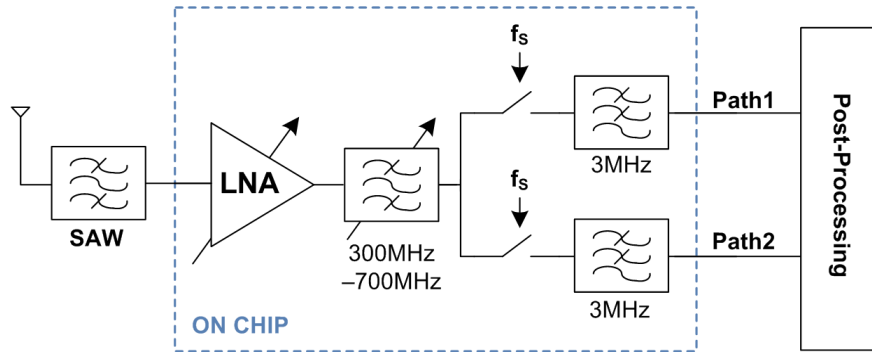
Figure 2.24: Block diagram of spectrum sensing system.

filter attenuates in-band blockers prior to sampling. The on-chip RF filter is frequency-tunable across the band of interest, tracking the signal frequency and mitigating harmonic folding from subsequent sampling. Both the LNA and RF tracking filter are gain-tunable to accommodate a wide dynamic range in input signal power levels.

The RF signal is then split into two identical paths for correlation function. The samplers downconvert the signal to baseband using sampling frequency f_S equal to the center frequency of the desired channel. Finally, baseband low-pass filters provide channel selection, and the filtered baseband outputs are processed off-chip in software for signal detection.

The additive noise from the samplers and all baseband blocks are uncorrelated between the two paths, and can be eliminated in post-processing using correlation detection. Path-splitting from the antenna and having two fully separate receivers, as in [13], would reduce system noise even further. However, having two RF front-ends would incur significant power overhead. Thus, our architecture of a single RF front-end with parallel downconversion provides an optimum compromise between detection sensitivity and system power consumption.

2.3.1 Dual-Mode Detection and Hybrid Correlation

Fig. 2.25 illustrates the detection methods used in the post-processing of the chip's analog baseband outputs. Post-processing is performed in software with Matlab. Coarse and fine detection modes, implemented respectively as energy detection and correlation, allow efficient sensing of a wide dynamic range of input signals. Energy detection simply measures the received power. This mode in practice would be used for a fast, coarse scan of all the channels to eliminate ones that have strong signals residing within. Correlation, being more time consuming, would then be used on a few likely-vacant channels where the signal, if present, is too weak to be sensed with energy detection.

In correlation mode, the signal from the two baseband paths are multiplied and then averaged. The correlation operation eliminates uncorrelated sampling and baseband noise, drastically reducing the effective noise contribution of the system. We use an analog-digital
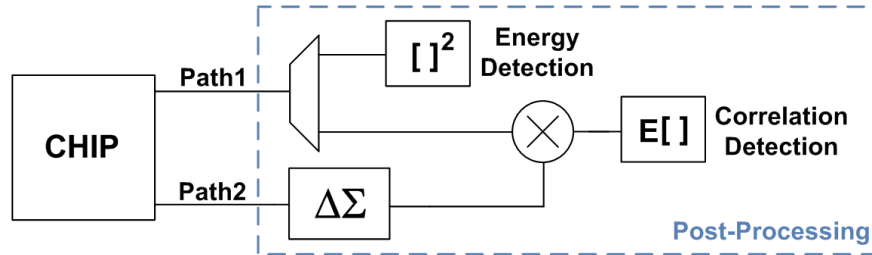
Figure 2.25: Block diagram of detection methods in software post-processing.

hybrid correlation scheme in which one path is first converted into a digital bitstream with a sigma-delta modulator prior to multiplication. In the remainder of this section, we will evaluate various techniques to implement correlation function, and then describe the mechanism and advantages of this hybrid correlation scheme.[3]

The high-resolution, high-linearity multiplication operations required for correlation are complex and power-intensive in both analog and digital implementations. This complexity can be significantly reduced if signals were instead converted to low-resolution, single-bit digital signals prior to correlation. A sigma-delta ($\Sigma\Delta$) modulator is a well known method for converting high-resolution analog signals into digital bitstreams while maintaining signal fidelity. If $\Sigma\Delta$ modulation were used, multiplication would become a simple XOR operation on two bitstream outputs.

The structure of a $\Sigma\Delta$ modulator is shown in Fig. 2.26(a). Each output digital bit V_O is converted into an analog high or low V_{FB} through a 1-bit DAC. Then, V_{FB} is fed back to and subtracted from the analog input V_I. Next, the error $V_E = V_I - V_{FB}$ is integrated in order to estimate the next bit output. The comparator frequency F_S is much greater than the highest frequency content of the input. This results in an oversampled bitstream that approximates the input signal, as shown in Fig. 2.26(b). The ratio and density of high and low output bits (in black) correspond to the instantaneous analog signal level (in blue). Thus, with enough oversampling, fidelity of high-resolution analog signals can be retained in digital bitstreams.

The noise transfer function of a $\Sigma\Delta$ loop has a high-pass characteristic, where quantization noise is pushed to higher frequencies outside of baseband. This results in the noise-shaping behavior shown in Fig. 2.26(c). In a standard $\Sigma\Delta$ ADC, the noise-shaping property is beneficial. The quantization noise at higher frequencies would be filtered out through digital decimation while a desired, high-resolution digital baseband signal remains.

---

[3]Autocorrelation can be performed by varying the relative phase of the sampling clock for the two downconversion paths. However, in measured results, standard cross-correlation of the two paths achieved similar performance as autocorrelation at significantly shorter sensing time. Thus, in this section, we refer only to standard correlation where the two paths use the same sampling clock with no phase shift.
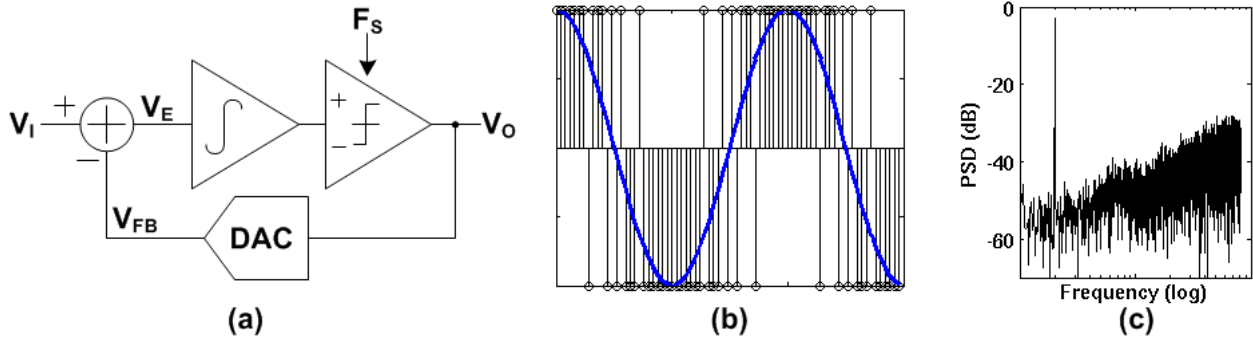
Figure 2.26: Sigma-delta modulation: (a) block diagram, (b) time-domain conversion of a sinusoid, (c) $\Sigma\Delta$ noise-shaping.

However, for our detection application, if we were to correlate two bitstreams directly, the out-of-band quantization noise is correlated, will mix to DC, and will severely degrade baseband SNR. On the other hand, if we decimate the bitstreams to filter out high frequency noise, we are left with two high-resolution digital signals, the multiplication of which then require the complex digital processing we sought to avoid. We thus lose the advantage of bitwise multiplication that using $\Sigma\Delta$ modulators was originally intended to provide.

An alternative to decimation is to correlate one $\Sigma\Delta$ bitstream with the initial analog signal, a technique originally proposed in [20] for digital audio applications. Since the quantization noise in the $\Sigma\Delta$ output is completely uncorrelated with the analog thermal noise, it should not affect the final correlated result. While multiplying a bitstream with an analog signal is more complicated than a mere XOR operation, it is still trivial compared to full analog or full digital multiplication. The bitstream would simply act as a sign bit that selects between the inverted and non-inverted forms of the analog signal.

Fig. 2.27 compares the simulated detected SNR of various correlation methods.[4] Both ideal analog and digital correlation have similarly good performance, while bitstream correlation has severely degraded SNR. The hybrid correlation scheme, however, achieves a performance comparable with and in fact mildly better than the analog baseline. We thus adopt the hybrid scheme for our spectrum sensing system. Finally, as expected, with more samples and more averaging, the detected SNR in all correlation schemes improve.

2.3.2 LNA

The LNA, as the first block in the receiver chain, needs to have high gain and low noise figure to meet the high sensitivity requirement. Simultaneously, the LNA must also have

---

[4]"Analog" correlates two analog signals directly. "Bitstream" converts both analog signals into $\Sigma\Delta$ bitstreams prior to multiplication. "Digital" decimates both bitstreams into high-resolution digital signals prior to multiplication. "Hybrid" correlates one original analog signal with one $\Sigma\Delta$ bitstream.
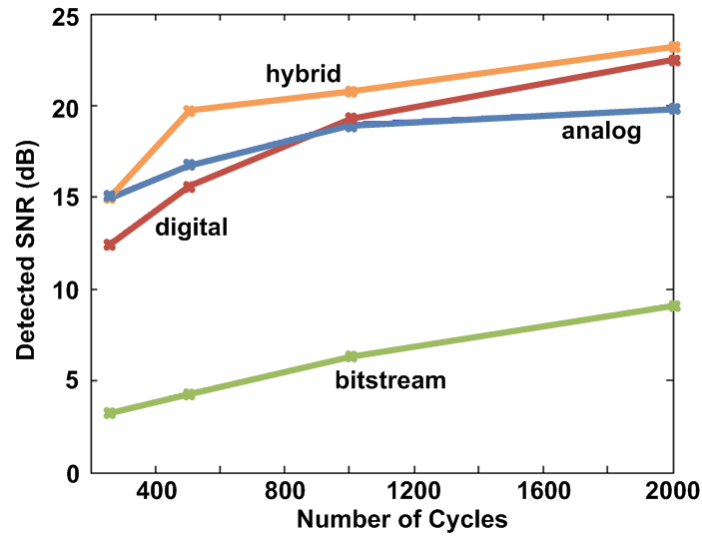
Figure 2.27: Comparison of various correlation implementations.

adequate linearity and tunable gain for good dynamic range. Fig. 2.28(a) shows a circuit
schematic of the implemented LNA. We adopted a capacitively cross-coupled common gate
(CG) architecture to present a wideband impedance match to a 50Ω antenna. This topology
also provides an optimal trade-off between gain, linearity, and noise figure. The LNA has
one bit of tuning D_{IN} in its resistive load to enable high and low gain modes.
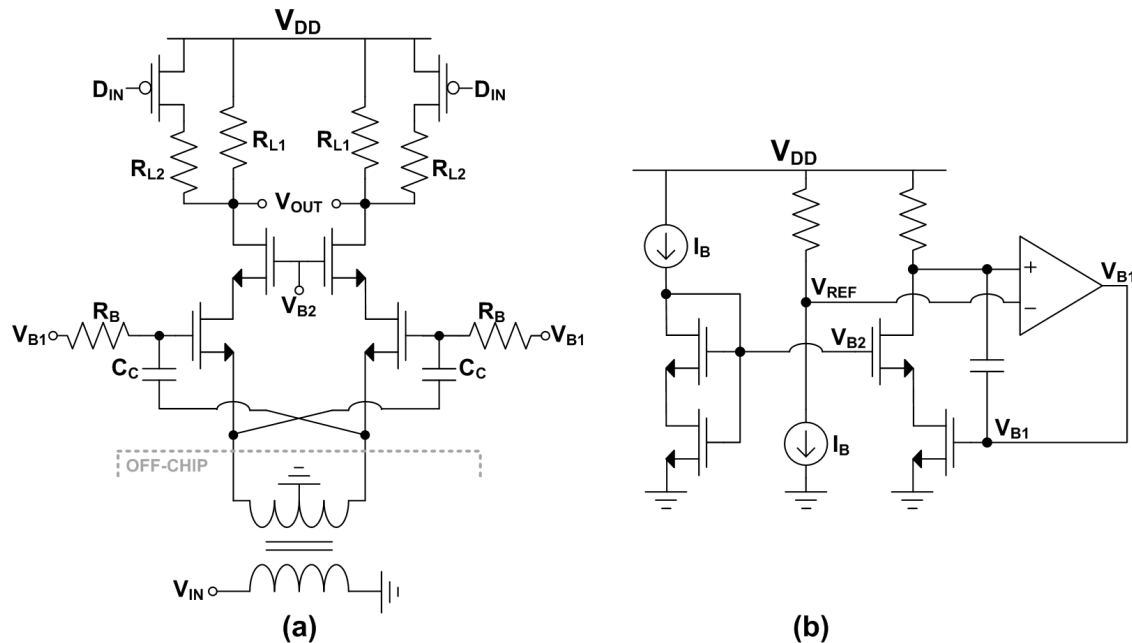
Figure 2.28: Circuit schematic of LNA: (a) core, (b) bias network.

The LNA is made fully-differential for good IIP2 performance, and an off-chip balun converts
the incoming single-ended signal to differential. The cross-coupling capacitors C_C act as gain-
boosting amplifiers with a gain of 1, which effectively doubles the g_m of the input devices.
Resistive loads are used for lower noise, and the cascode devices provide a high output
impedance to not degrade the load resistance. The set of switched load resistors R_{L2} have
lower resistance than R_{L1} and provide a low gain mode with improved linearity.

A replica bias scheme is used for the LNA, and a schematic of the bias circuitry is shown in
Fig. 2.28(b). The rightmost branch is a replica of one single-ended branch of the LNA at 1/6
its size. The middle branch generates a reference voltage for the output common mode, and
the feedback loop generates bias voltage V_{B1} to keep the output common mode constant.
The leftmost branch generates the cascode bias V_{B2}.

Assuming LNA output impedance $R_{OUT} \gg R_L$, the theoretical gain and noise factor of the
LNA are:

$$A_V = 2 * g_m * R_L \tag{2.10}$$

$$F = 1 + \frac{\gamma}{2} + \frac{2}{g_m R_L} \tag{2.11}$$

In practice, the gain and noise factor deviate from Eq. 2.10 and Eq. 2.11 due to finite
cross-coupling capacitance. The capacitors should ideally feedthrough the inverted signal
un-changed, but in actuality, the feedthrough ratio is less than 1 due to the capacitive
divider formed by C_C and the input device's gate capacitance C_G. The resulting effective
G_M, rather than being twice g_m, is given in Eq. 2.12. To not degrade performance, C_C much
be made sufficiently large.

$$G_{M-EFF} = g_m \left(1 + \frac{C_C}{C_C + C_G} \right) \tag{2.12}$$

The input impedance of a CG LNA is $1/g_m$, or in our case, $1/G_{M-EFF}$. The standard
antenna impedance of 50Ω therefore limits g_m to 20 mS.[5] As seen in Eq. 2.11, noise figure
(NF) depends solely on $g_m R_L$. Since g_m is limited by the input impedance, and R_L is limited
by headroom, there is a fundamental lower limit to the noise figure achievable in a CG LNA
architecture.

However, Eq. 2.11 assumes a perfectly impedance-matched condition. If we intentionally
mismatch the LNA input impedance R_{IN} with antenna impedance R_S, the noise factor
becomes:

$$F = 1 + \alpha * \frac{\gamma}{2} + \frac{(\alpha + 1)^2}{2\alpha} * \frac{2}{g_m R_L}, \qquad \alpha = \frac{R_{IN}}{R_S} \tag{2.13}$$

Assuming $g_m R_L$ is sufficiently large so that $\gamma/2$ is the dominant term, a smaller R_{IN} (corre-
sponding to a larger g_m) results in a smaller noise figure. On the other hand, an impedance

---

[5]The factor of 2 gm-boost is cancelled out by a factor of 2 impedance transformation through the balun.

mismatch causes more power to be reflected back to the antenna rather than delivered to
the LNA. This reflection is measured with S11, defined in Eq. 2.14.

$$S11 = \left| \frac{Z_{IN} - R_S}{Z_{IN} + R_S} \right| \tag{2.14}$$

A perfect match results in an S11 of 0, meaning all power is delivered and none reflected.
However, an S11 of -10 dB has traditionally been considered acceptable, so some mismatch
can be tolerated. Fig. 2.29 plots NF and S11 as a function of R_{IN}.[6]
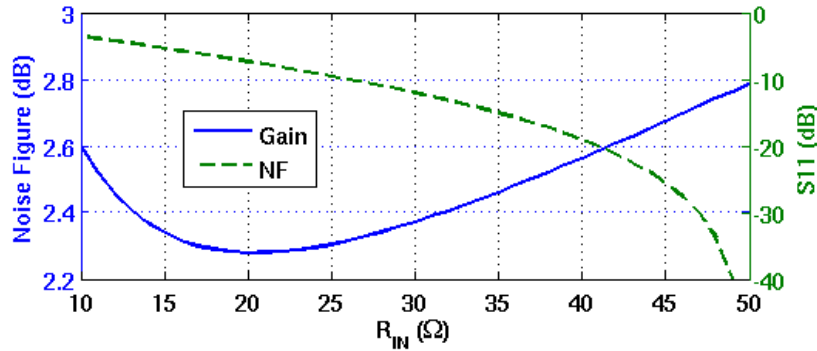
Figure 2.29: LNA performance as a function of R_{IN}: (a) NF, (b) S11.

At the cost of spending power to make g_m bigger, we can achieve lower noise figure while
still having sufficient matching. The implemented LNA targets $\alpha \approx 0.7$ and $R_{IN} \approx 35\Omega$. For
its high and low gain modes, the LNA has simulated gains of 24dB and 14dB respectively,
and simulated NF of 2.2dB and 3.3dB respectively. Simulated S11 is below -15dB for the
entire band of interest. The complete LNA block, including bias circuitry, consumes 6mW
from 1.2V supply.

2.3.3 RF Tracking Filter

The RF filter is a 4th order bandpass filter implemented with a cascade of two Gm-C biquads.
The biquad structure presents a high impedance to the previous stage LNA and does not have
to be impedance matched, and the Gm-C implementation gives the lowest power. Fig. 2.30
illustrates the RF filter as well as the structure of a biquad.

The Gm-C biquad architecture imitates a parallel RLC tank to achieve a bandpass response.
The components G_{M3}, G_{M4}, and C_2 form a gyrator with a capacitive load, which is well-
known to have an inductive response. This acting inductor is placed in parallel with capacitor
C_1 and diode-connected G_{M2} (which acts as a resistance) to form an RLC tank. The block

---

[6]In NF plot, $g_m R_L$ has been set to 10. This accounts for increased NF at lower limits of R_{IN}.

Figure 2.30: Block diagram of RF tracking filter.

G_{M1} is a transconductor driving the RLC load, and, like the LNA, G_{M1} has one bit of gain tuning D_{IN} for improved dynamic range.

The transfer function of the biquad is given by:

$$H(s) = \frac{sG_{M1}C_2}{s^2C_1C_2 + sG_{M2}C_2 + G_{M3}G_{M4}} \tag{2.15}$$

From the transfer function, we can derive characteristics for peak gain A_V, peak frequency ω_o, Q of the bandpass response, and noise factor at the peak frequency:

$$A_V = G_{M1}/G_{M2} \tag{2.16}$$

$$\omega_o = \sqrt{\frac{G_{M3}G_{M4}}{C_1C_2}} \tag{2.17}$$

$$Q = \sqrt{\left(\frac{G_{M3}}{G_{M2}}\right)\left(\frac{G_{M4}}{G_{M2}}\right)\left(\frac{C_1}{C_2}\right)} \tag{2.18}$$

$$F|_{\omega_o} \approx 1 + \frac{\gamma}{A_V G_{M2} R_S}\left(1 + \frac{1}{A_V} + \frac{G_{M4}}{A_V G_{M2}}\left(1 + \frac{C_1}{C_2}\right)\right) \tag{2.19}$$

Varying C_1 and C_2 changes the peak frequency of the bandpass filter, and the implemented capacitors are made tunable in order to track the desired channel. Varying G_{M1}, on the other hand, changes only the gain and not the frequency response. Thus, high and low gain modes are implemented by switching an extra section of G_{M1} as shown in Fig. 2.30. For lower noise, we choose $G_{M4} = G_{M2}$, so that we are left with $Q = G_{M3}/G_{M2} = C_1/C_2$.

Any finite output resistance at the output node combines in parallel with $1/G_{M2}$ resistance, which translates to a larger effective G_{M2}. This, in turn, degrades both Q and gain. This degradation can be significant since the output node sees the output impedances of 3 Gm-cells ($G_{M1,2,4}$) in parallel, with G_{M1} being especially large if high gain is desired. Since there is not enough headroom for cascode devices, we boost the Gm-cell output impedances by using longer channel devices. We also add a Q-boosting negative-gm component to further compensate for the loss.

The schematic of one Gm-cell is shown in Fig. 2.31. The Gm-cells employ a resistively-degenerated common source architecture, where R_X is the degeneration resistor and $g_m R_X$ is set to 1 for balance between lower power and greater linearity. The transistors $M_{1,2}$ are input devices, $M_{3,4,5,6}$ are current sources, and $M_{7,8}$ are a negative-gm load for implementing Q-boost. The output common mode is set by the common mode feedback (CMFB) loop controlling the load PMOS current sources. While both the CMFB and Q-boost are shown in the individual Gm-cell schematic in Fig. 2.31, they are implemented only once on each shared biquad node.
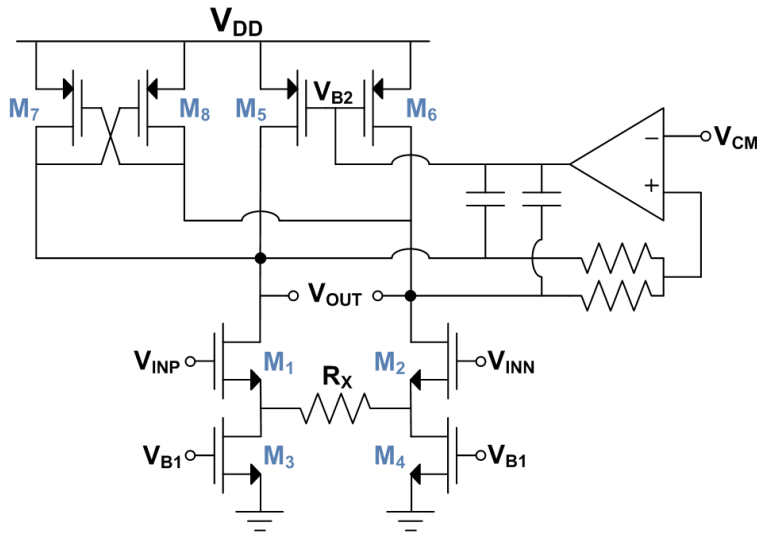
Figure 2.31: Circuit schematic of a Gm-cell.

The gyrator loop involving G_{M3} and G_{M4} has two poles at the same frequency, and thus needs to be carefully designed for stability. The phase margin of the loop, however, also corresponds with Q, where a higher Q translates to a loop closer to instability. To have a safety margin where there is at least 30° phase margin across all corners and process variations, we are limited to $Q = 2$, which translates to about 14dB of rejection at the third harmonic frequency for one biquad. Since more rejection is needed before the sampler, we cascaded two Gm-C biquads to form a 4th order bandpass filter.

The peak gain of each biquad is about 10dB and 0dB respectively for high and low gain modes. Rejection at the 3rd harmonic is 12dB to 14dB for each biquad across all gain and frequency settings. The tunable capacitors are implemented as binary-coded capacitor banks with 3-bit digital inputs and 8 frequency settings. Since each biquad can contribute additional gain, we can tolerate a higher noise figure from the second biquad. The second biquad is thus designed to be half the size of the first one, consuming roughly half of the first's power at a cost of higher noise. The two cascaded biquads, including bias and peripheral circuitry, consume 9mW in total.

The complete RF front end, composed of the LNA and the full RF tracking filter, has programmable gain from 14dB to 44dB. The simulated RF front end noise figure in its highest gain setting is 3.5dB.

2.3.4 Sampler

As mentioned in Sec. 2.1.1, the UHF TV band in the U.S. consists of 6MHz channels with vestigial sideband modulation. A sinusoidal pilot tone exists at 310kHz above the lower channel edge, which is also the frequency of the buried carrier. To perform downconversion, the receiver uses an LO frequency equal to the center, not carrier, frequency of the desired channel, as described in Sec. 2.2.1 and shown here in Fig. 2.32.
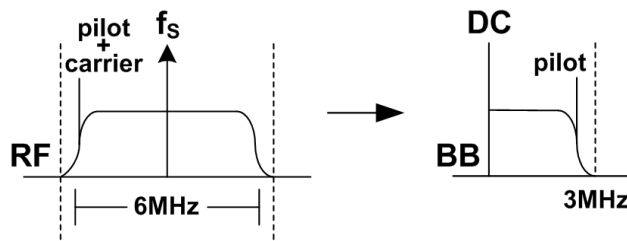
Figure 2.32: Downconversion of a 6MHz UHF TV channel.

This approach has several advantages. First, it enables direct conversion of the channel and eliminates any need for image rejection. Second, the pilot tone falls to the low-IF frequency of 2.69MHz and can be detected within its narrow bandwidth as an alternate detection mode. Third, the channel at baseband folds onto itself and becomes only 3MHz wide. Due to this folding, the channel data becomes corrupted and undecodable, but this is acceptable for our application as we only wish to sense the presence of the signal in the channel.

Fig. 2.33 illustrates the structure of the sampler used to enable downconversion. The sampler consists of two sample-and-hold (S&H) stages clocked on the opposite phase. Buffers before and after each S&H stage act as drivers and provide input-output isolation. Gate bootstrapping is used on all sampling switches to enable constant V_{GS} sampling and maximize linearity.

The gate-bootstrap circuit is shown in Fig. 2.33 [21]. When $\phi=0$ and $\overline{\phi}=1$, the bootstrap capacitor C_{BOOT} is charged up to full $V_{DD}=1.2$V. A clock multiplier creates the $2V_{DD}$ gate voltage for the pull-up device that connects C_{BOOT} to the supply. During this phase, C_{BOOT} is disconnected from the sampling switch, and the gate of the sampling switch is pulled low to ground. When $\phi=1$ and $\overline{\phi}=0$, the charging transistors for C_{BOOT} turn off, and C_{BOOT} is applied between the source and gate nodes of the sampling switch. While the sampling switch is on, C_{BOOT} ensures that the sampling switch gate swings with the input signal at
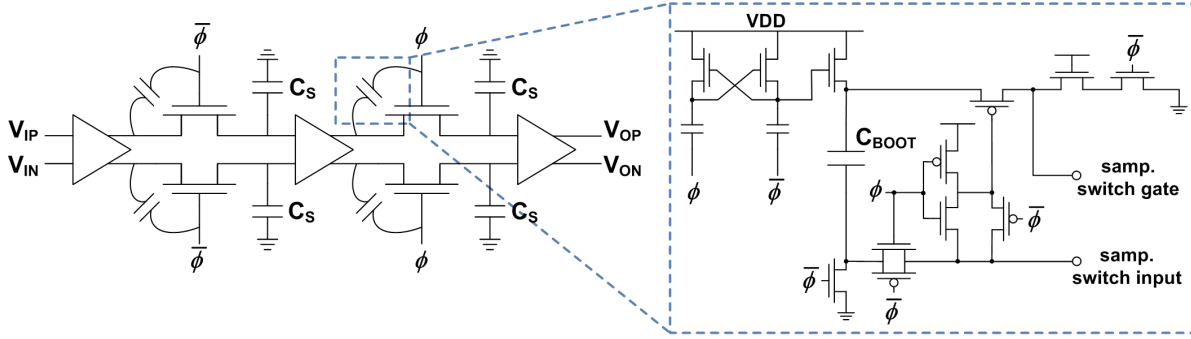
Figure 2.33: Schematic of downconversion sampler.

a constant V_{GS}=1.2V. The remaining devices in the bootstrap circuit exist to prevent any transistor from experiencing voltages greater than V_{DD}.
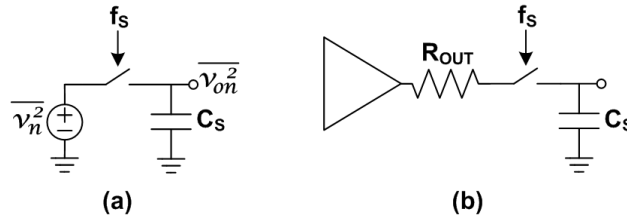
Figure 2.34: Sample-and-hold noise analysis: (a) sampling switch and capacitor, (b) with buffer.

Next, we discuss some design considerations that drove the implementation of the sampling switch, sampling capacitor, and buffers. First, we consider the noise contribution of the simple S&H circuit shown in Fig. 2.34. During the sample phase, the output noise consists of the switch resistance R_{ON}'s noise power shaped by the RC frequency response of R_{ON} and C_S. Given 50% duty cycle, the output sample noise spectrum is [22]:[7]

$$\overline{v_{on,SAMP}^2}(f) = \frac{1}{2}\left(\frac{4kTR_{ON}}{1 + (2\pi f R_{ON}C_S)^2}\right) \tag{2.20}$$

When the switch turns off, the total integrated noise power sampled onto the capacitor is kT/C_S. Due to the characteristics of sampling and noise-folding, the accumulated noise power distributes evenly from DC to $f_S/2$, where f_S is the sampling frequency. Zero-order hold response then applies a sinc characteristic to the noise frequency spectrum. The resulting output hold noise spectrum is [22]:

$$\overline{v_{on,HOLD}^2}(f) = \left[\frac{1}{2} * sinc\left(\frac{1}{2} * \frac{f}{f_S}\right)\right]^2 * \left(\frac{2kT}{Cf_S}\right) \tag{2.21}$$

[7]The total noise is halved because the sample phase occurs for half of a clock period.

The output frequency of interest is at baseband and much lower than both f_S and the sampler's RC bandwidth. Given this, the total output noise of a S&H circuit is:

$$\overline{v^2_{on,TOT}}|_{f \ll fs} = 2kTR_{ON} + \frac{kT}{2Cf_S} \qquad (2.22)$$

For low sampler noise, small switch resistance and large sampling capacitance are desired. In a practical sampler implementation, however, the output impedance of the preceding driver dwarfs R_{ON} and is the design bottleneck. For the circuit shown in Fig. 2.34(b), hold phase noise remains the same while sample phase must include the output noise of the buffer. If we assume buffer noise power $\propto g_m$ and output resistance $R_{OUT} \propto 1/g_m$, the modified S&H noise becomes:

$$v_{on,TOT} = 2kT \left(R_{ON} + \frac{\gamma}{g_m} \right) + \frac{kT}{2Cf_S} \qquad (2.23)$$

Because our target application is low power, buffer output impedance $1/g_m$ is much greater than R_{ON} for any reasonable implementation of the sampling switch, and buffer noise during sample phase dominates. The considerations of buffer power consumption versus sampler dynamic range, as constrained by buffer noise, was the primary design trade-off for the sampler. Consequently, both sampling switch and capacitor could be minimized with little effect on noise performance. A smaller sampling capacitor results in increased bandwidth and decreased switching power. A smaller sampling switch results in a more relaxed, compact, and power-efficient design of the bootstrap circuit.

The implemented buffer uses $g_m \approx 2.4$mS, resulting in $R_{OUT} \approx 420\Omega$. In comparison, the implemented sampling switch has $R_{ON} \approx 13\Omega$. The designed sampling capacitor is approximately 140fF, although actual C_S is somewhat bigger due to parasitics.
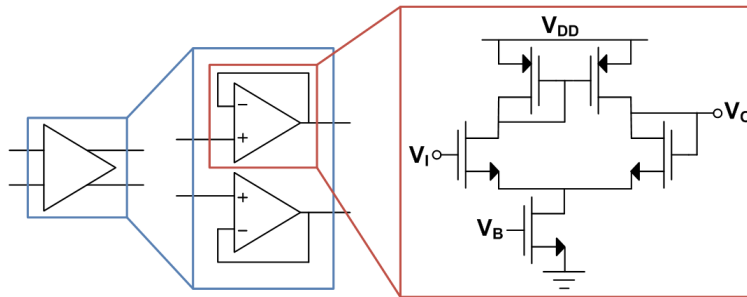
Figure 2.35: Schematic of a sampler buffer.

Fig. 2.35 shows a schematic diagram of the buffer. We use a unity-feedback operational amplifier (OpAmp) topology for its optimal linearity performance and insensitivity to input common mode. While this topology consumes more power than alternative structures due to the required differential pair for each single-ended signal path, the linearity improvement was significant and worth the cost. The implemented downconversion sampler in its entirety

achieves in simulation 0.3dBm output P1dB and 46μV integrated RMS output noise, result-
ing in 74dB dynamic range.[8] Combined with 30dB of variable gain from the RF front-end,
the target receiver achieves sufficient linearity to robustly detect the full range of input signal
powers.

2.3.5 Baseband Filter

The downconversion sampler is followed by a 4th order lowpass Chebyshev baseband filter
to provide channel selection at 3MHz. Fig. 2.36 shows a schematic of the baseband filter.
The filter is implemented as a cascade of two Tow-Thomas biquads with a DC servo loop to
attenuate flicker noise near DC frequencies.
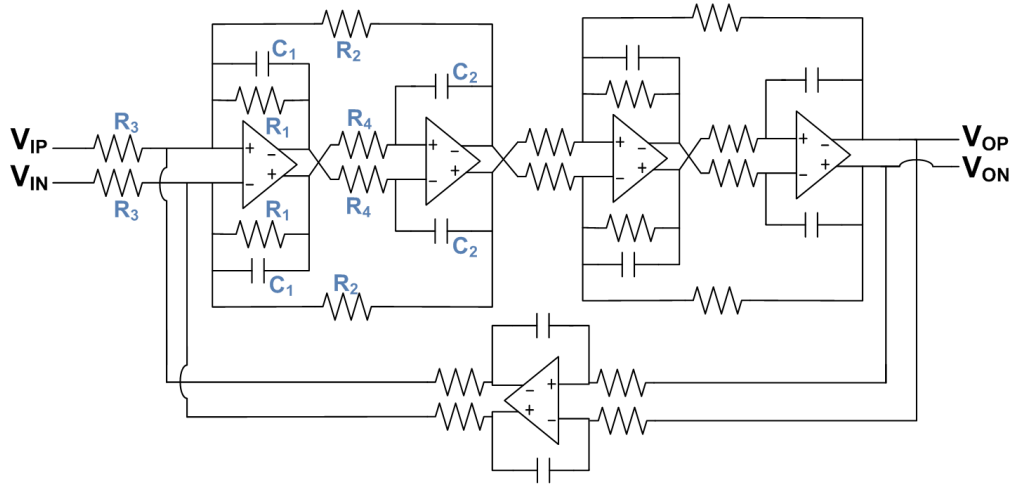
Figure 2.36: Schematic of 4th order low-pass baseband filter with DC servo.

Each biquad constructs a set of complex poles, and the transfer function for one Tow-Thomas
biquad is:

$$H(s) = \frac{R_2}{R_3}\left(\frac{1}{s^2 R_2 R_4 C_1 C_2 + s R_2 R_4 C_2 / R_1 + 1}\right) \tag{2.24}$$

From Eq. 2.24, we can extract the filter parameters DC gain A_V, peak frequency ω_o, and Q:

$$A_V = R_2/R_3 \tag{2.25}$$

$$\omega_o = \sqrt{\frac{1}{R_2 R_4 C_1 C_2}} \tag{2.26}$$

$$Q = \frac{R_1}{\sqrt{R_2 R_4}}\sqrt{C_1 C_2} \tag{2.27}$$

---

[8]Noise is integrated from 100Hz to BB bandwidth of 3MHz.

The ω_o and Q coefficients for each biquad are listed in Eq. 2.28. They combine to generate a 4th order Chebyshev response with the desired passband frequency (3MHz), passband ripple, and stopband attenuation. The target frequency response of each biquad, as well as their combined 4th order response, is plotted in Fig. 2.37.[9]

$$\omega_1 = 0.996 * 3\text{MHz} \qquad Q_1 = 3.503$$
$$\omega_2 = 0.533 * 3\text{MHz} \qquad Q_2 = 0.777 \tag{2.28}$$

Additional gain in baseband was not needed for our receiver, so we set $R_2=R_3$ in both biquads for 0dB gain. For simplicity of implementation, we also set $R_2=R_4$ and $C_1=C_2$. This results in $Q = R_1/R_2$ and $\omega_o = 1/(R_2C)$, where $C=C_1=C_2$. To mitigate resistance mismatch from layout, we approximate the target Q coefficients to $Q_1=3.5$ and $Q_2=0.75$ to enable integer ratios of resistor segments. All capacitors, including the feedback capacitor used for DC servo, are 10.4pF, and the value was chosen as a balance between opposing constraints of layout area and lower noise. The servo loop generates a high-pass corner at approximately 100kHz. The baseband filter is a fixed 3MHz filter; however, all resistors are implemented as 3-bit resistor banks with moderate tuning range to compensate for process corners and variations.
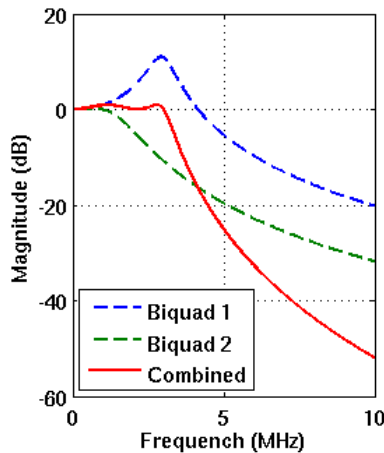
Figure 2.37: Frequency response of each biquad along with their combined response.

Fig. 2.38 shows a circuit diagram of the OpAmp used in the baseband filter. The OpAmp consists of a folded cascode topology with an output swing stage. The input stage uses PMOS devices for their lower flicker noise, and the telescopic and output stages each have separate CMFB loops to set their DC operating points. RC compensation is used for robust OpAmp stability, and phase margin is more than 60° across all process corners. The OpAmp has unity-gain frequency of approximately 55MHz, and it has negligible effect on the overall baseband filter's frequency response.

---

[9]The plot shows the theoretical ideal and does not include circuit non-idealities or DC servo response.
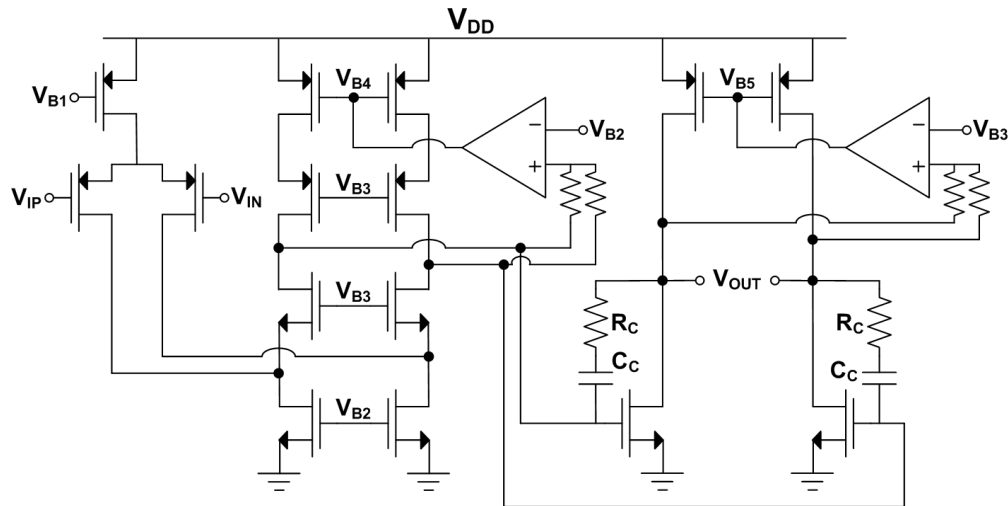
Figure 2.38: Schematic of baseband operational amplifier (OpAmp).

The differential output of the baseband filter forms the final analog baseband output from the implemented spectrum sensing receiver. Analog 50Ω buffers are inserted at the output to drive the signals off-chip. The baseband outputs are then measured and post-processed as described in Sec. 2.3.1 to determine detection performance.

2.4 Measurement Results

A chip prototype of the system, shown in Fig. 2.39, was fabricated in a 65nm CMOS process with 1.2V supply [23]. The chip measures 1mm x 1.2mm, and the die is wirebonded directly onto an FR4 PCB for measurement. The test PCB is 4 layers, consisting of top and bottom signal planes, a ground plane, and a split power plane in which the supply for the sensitive RF front-end has been separated from that of the sampling and baseband circuitry.

The test PCB has a single-ended antenna input port, and prior to reaching the differential input port of the chip, discrete on-board components consisting of an UHF band SAW filter, a balun, and a bias-T are applied. Clock inputs for the samplers are sourced directly from differential pulse generators. The insertion loss of all front-end discrete components as well as PCB traces are included in measurement results and not de-embedded.

Both the chip and test PCB have 3 sets of differential output ports. Two are the system baseband outputs from the two sampling paths; the third is a test port at the output of the RF tracking filter (prior to the samplers) which gives visibility to the performance of the standalone RF front-end. SMA connector power combiners are used to transform all differential outputs to single-ended for interfacing with measurement equipment. The power combiners as well as all cables are de-embedded in measurement results.
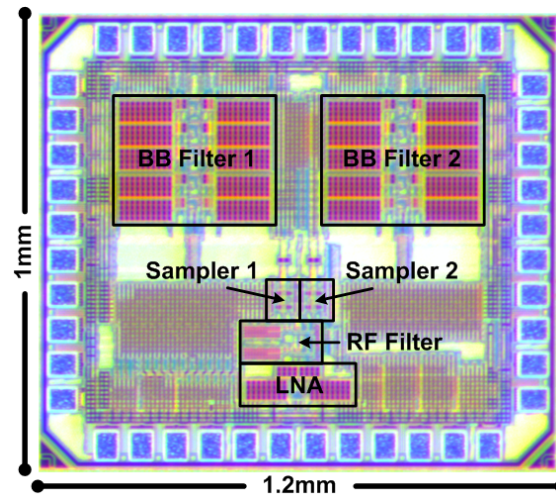
Figure 2.39: Chip micrograph of implemented spectrum sensing system.
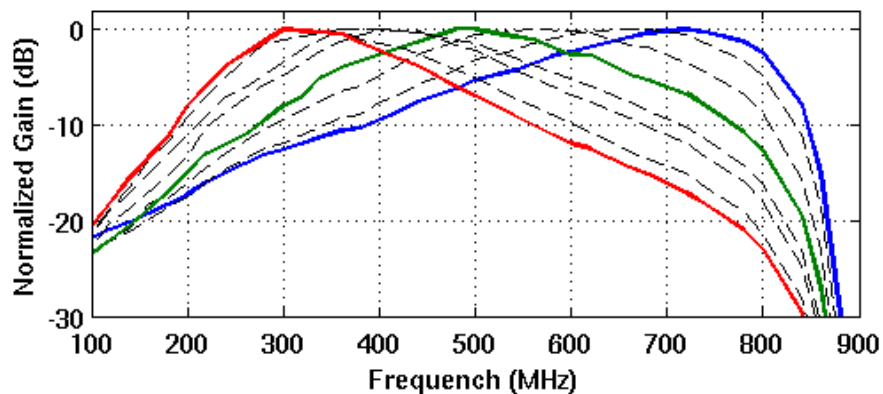
Figure 2.40: Normalized gain response of RF front-end for various RF tracking filter frequency settings.

Fig. 2.40 plots the measured response of the standalone RF front-end across the frequency band of interest. All 8 RF tracking filter frequency settings are plotted, and the lowest, the highest, and a mid-band setting are highlighted. The tracking filter covers the entire UHF TV band from 300MHz to 700MHz. Beyond 800MHz, all frequency settings experience sharp roll-offs due to the front-end SAW filter, which provides high attenuation of 900MHz GSM blockers.

Fig. 2.41(a) plots the measured baseband response. The baseband filter defines a 3MHz bandwidth with Chebyshev passband rippling and attenuation at near-DC frequencies. Fig. 2.41(b) plots adjacent channel rejection normalized to gain at 3MHz - the desired channel edge. The baseband filter is able to reject 37dB at the adjacent channel, 60dB at the N+2 channel, and greater than 70dB for N+3 and beyond. This is more than sufficient given the blocker
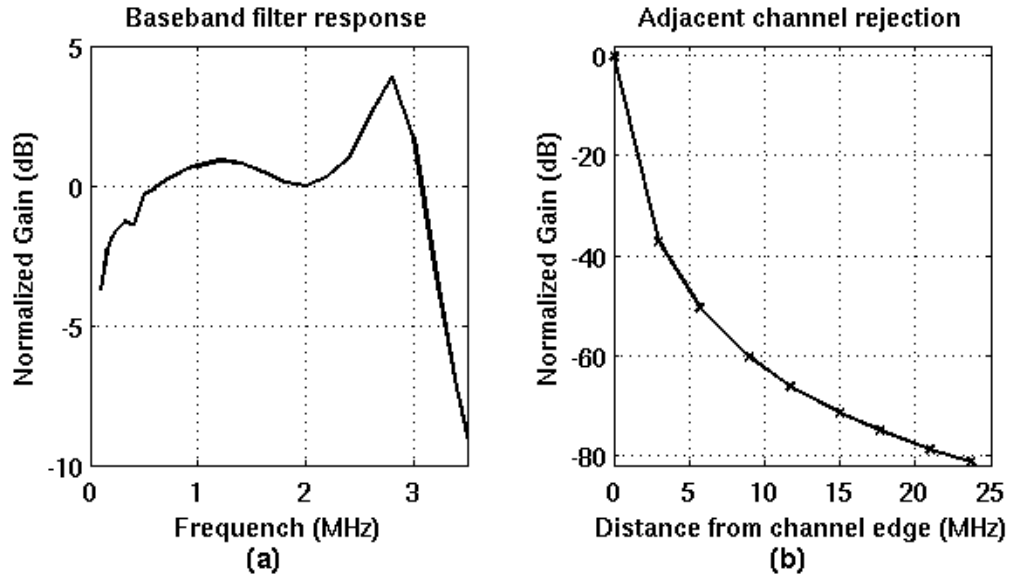
Figure 2.41: Baseband gain response: (a) passband, (b) adjacent channels.

profile of the TV band, and prevents strong adjacent channel blockers from desensitization our desired channel.

To determine receiver sensitivity, first, we measure the receiver's pure noise output. At the highest gain mode, the measured input-referred noise power within the 3MHz baseband channel is -96dBm to -91dBm across the 300MHz to 700MHz RF band at room temperature. When referred to the original 6MHz RF channel bandwidth, this corresponds to a sensitivity of -164dBm/Hz to -159dBm/Hz. In order to compare different detection modes, we also define receiver sensitivity as the input signal power at which the detected SNR reaches 3dB.[10] This threshold was chosen as the point at which signal power is equal to noise power; however, robust sensing is possible at lower detected SNR thresholds with accurate noise estimation algorithms.

Fig. 2.42 shows measured detected SNR as a function of input power at the highest gain mode.[11] For energy detection, the 3dB detected SNR point occurs at input powers of -96dBm to -91dBm, which, as expected, matches the channel noise power. With correlation, a sensitivity of -104dBm to -106dBm was achieved at room temperature.

Detection speed is determined primarily by the amount of time necessary to average enough noise samples for a robust estimation of noise power. Fig. 2.43 shows the convergence of noise power over time for energy and correlation detection. The total noise power is determined

---

[10]Detected SNR is defined as the ratio of detected channel power when a signal is present to detected channel power when no signal is present. It is described in more detail in Sec. 2.2.2.

[11]Error bars correspond to the maximum and minimum detected SNR over four signal and four noise samples.
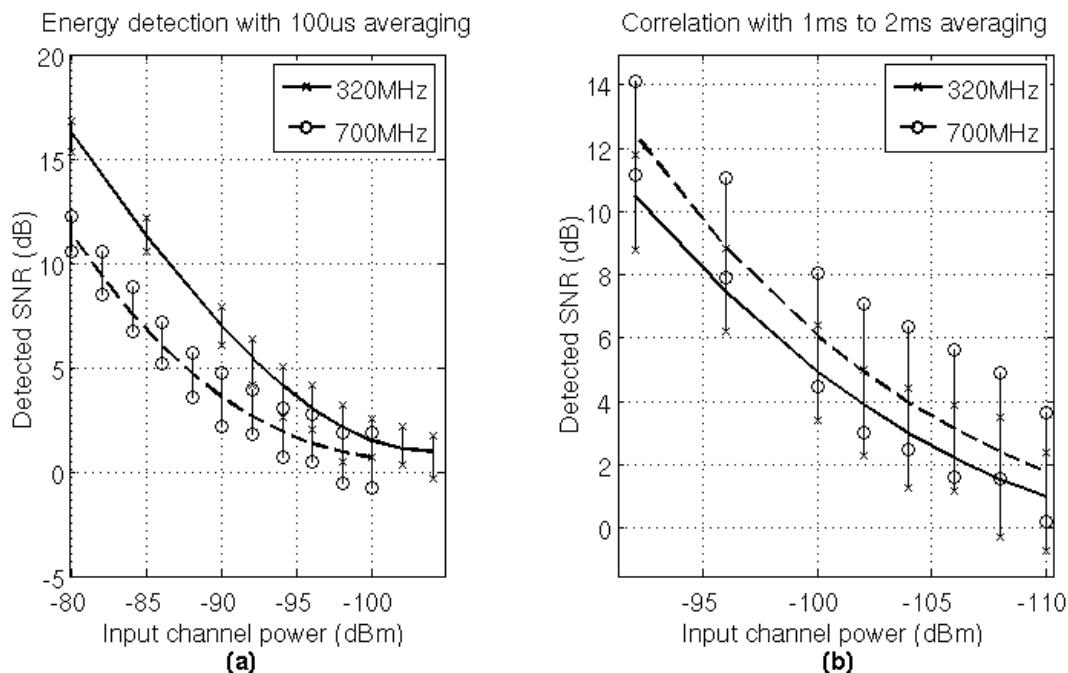
Figure 2.42: Measured detected SNR as a functional of input signal power: (a) energy detection, (b) correlation detection.

by averaging a 20ms sample, and noise error is defined as the maximum deviation from that value for each shorter averaging time. As expected, energy detection converges much faster, reaching an error of 1dB with an averaging time of about 100us. Correlation, on the other hand, needs more than 1ms to reach a similar level of noise error.[12]
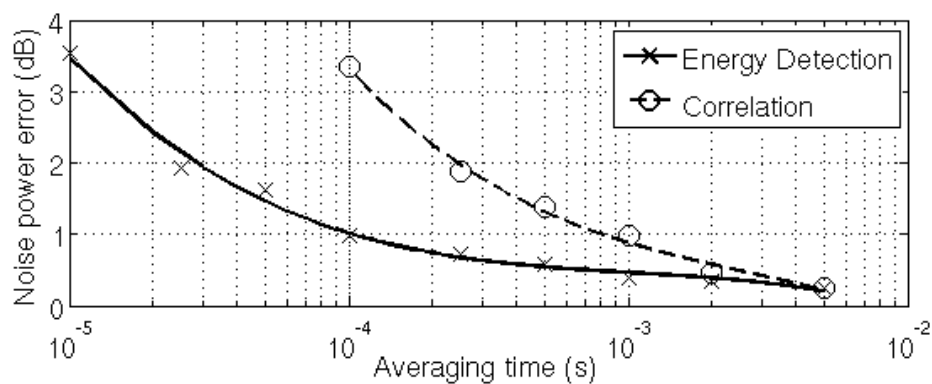
Figure 2.43: Convergence of noise power with averaging time.

---

[12]Pilot detection was also attempted in measurement. However, due to the narrow bandwidth of the pilot, detection was more time consuming than correlation detection while producing worse sensitivity. Therefore, pilot detection does not add value to this work and detailed results are not reported.

Fig. 2.44 shows P1dB linearity measurements for the highest, the lowest, and a medium gain
mode. The output P1dB for all modes are similar and in the ∼-27dBm to ∼-30dBm range.
Linearity was limited by the chip's output pad drivers and not any core receiver circuitry.
Regardless, for the lowest gain mode, an input P1dB of -14dBm to -20dBm was achieved
across the RF band, giving a dynamic range of 84dB across the full scale of gain settings.
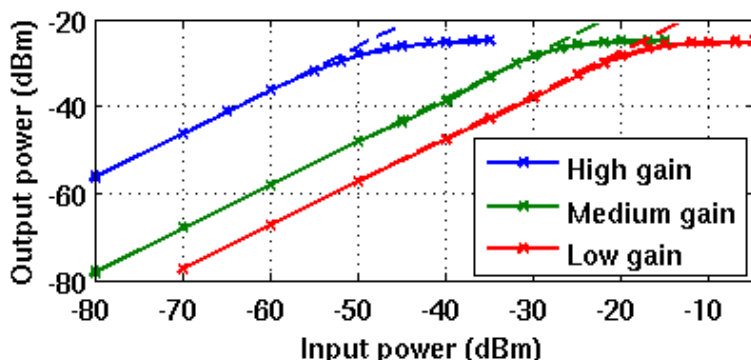
Figure 2.44: P1dB linearity measurements for high and low gain modes.

Fig. 2.45 shows in-band and out-of-band IIP3 for the lowest gain mode. The in-band IIP3
is measured with two tones in the N+1 and N+2 channels, and we achieve input IIP3 of
-12dBm. The out-of-band IIP3 is measured with two tones at 760MHz and 860MHz, and
IIP3 of +7dBm is achieved. The out-of-band IIP3 shows significant improvement due to the
effects of the front-end SAW filter.
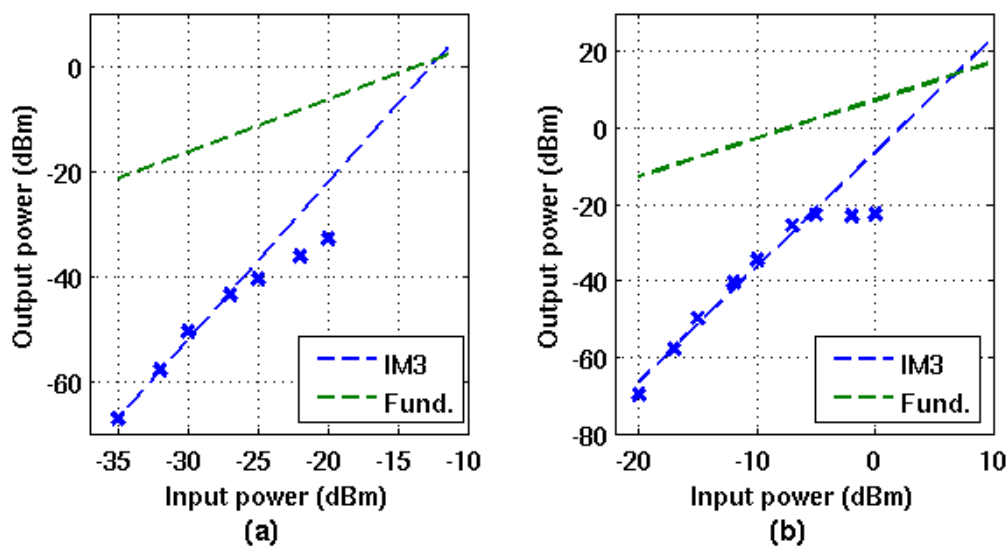
Figure 2.45: IIP3 linearity measurements: (a) in band, (b) out of band.

Linearity at the highest gain mode is also important as it defines when the desired channel
becomes desensitized by blockers, causing the sensitivity measurement to be no longer valid.
Fig. 2.46 shows these desensitization measurements. Due to high adjacent channel rejection
from the baseband filter, blocker-induced gain compression is limited by the linearity of the
RF chain. As shown in Fig. 2.46(a), the 1dB compression point in the desired channel occurs
at -42dBm blocker power for an N+1 blocker, and stays relatively constant at -40dBm for
an N+6 blocker.
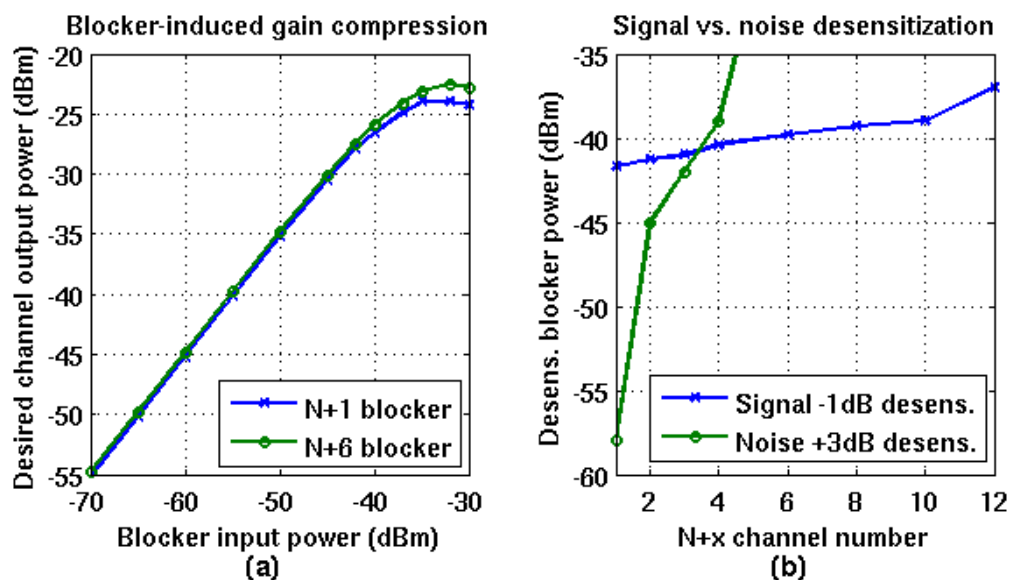
Figure 2.46: Desensitization measurements: (a) adjacent channel blocker-induced gain compression, (b) noise vs. gain desensitization.

In addition to gain compression, strong adjacent channel blockers also raise the noise floor
of the desired channel due to noisy LO. Fig. 2.46(b) shows blocker power at the point
of desensitization, for both gain and noise desensitization, as a function of N+x channel
number. As shown, noise desensitization effect dominates for close blockers below N+4.
Desensitization due to the N+1 blocker, defined as when the desired channel noise power
rises by 3dB, occurs at -58dBm blocker power. The desensitization blocker power rises to
-45dBm for the N+2 blocker. Since the highest sensitivity mode is only used to detect very
weak signals, and since adjacent channel power is limited to being a maximum of 27dB
greater, this desensitization result satisfies the intended application and does not degrade
the robustness of detection.

Table 2.1 presents a summary of measurement results. The spectrum sensing receiver in its
entirety consumes only 28mW.[13] The system uses a combination of energy detection and
correlation to provide both rapid and robust sensing, and is resilient against desensitization

---

[13]Reported power consumption is for midband frequency of 500MHz

from nearby blockers of its target application. This work is able to achieve the highest sensitivity per bandwidth (in dBm/Hz) for the targeted application with the lowest power [23], and demonstrates the feasibility of integrated low-power, high-sensitivity spectrum sensing for mobile receivers.

Table 2.1: Summary of spectrum sensing receiver measurements.

	High Gain	Medium Gain	Low Gain
Erg. Sensitivity (dBm)	-91 to -96	-69 to -75	-60 to -66
Corr. Sensitivity (dBm)	-103	n/a	n/a
P1dB (dBm)	-48 to -53	-22 to -28	-14 to -20
Dynamic Range (dB)	50	47	46
Full Scale DR (dB)	83	n/a	n/a
In Band IIP3 (dBm)	-31	-22	-12
Out of Band IIP3 (dBm)	-18	-1	+7

Chapter 3

Transmit/Receive Switching for TDD Co-existence

Time division duplex (TDD) systems are growing in popularity in 4G and 5G cellular standards because they are simpler, more energy efficient, and lower cost. Meanwhile, T/R switches to enable TDD co-existence still make up a significant portion of off-chip RF front-end modules in commercial solutions. Thus, as part of the broad goal towards reconfigurable radios and eliminations of discrete front-end components, we focus on enhancing integration and reconfigurability for TDD front-ends at cellular gigahertz frequencies. In this chapter, we describe the challenges of TDD co-existence, analyze the benefits and drawbacks of existing integrated T/R switches, and propose PA re-use as a T/R switching solution to overcome these obstacles.

3.1 Introduction to TDD

3.1.1 TDD vs. FDD

Frequency division duplexing (FDD) and time division duplexing (TDD) refer to the two methods of achieving bi-directional communications over a channel. In FDD, two separate frequency bands are allocated for uplink (UL) and downlink (DL). In other words, transmit and receive sides of an FDD system can operate simultaneously but on different carrier frequencies. In TDD, on the other hand, uplink and downlink use the same frequency band but take turns operating in time domain. Fig. 3.1 illustrations graphical conceptual representations of TDD and FDD.

Of the two duplexing methods, TDD has three main benefits over FDD. Firstly, it is considered more spectrum efficient due to the self-explanatory reason that it requires only one band as opposed to two. In fact, FDD requires additional spectrum to act as guard bands in between its UL and DL bands. Guard bands provide buffer for imperfect filtering and
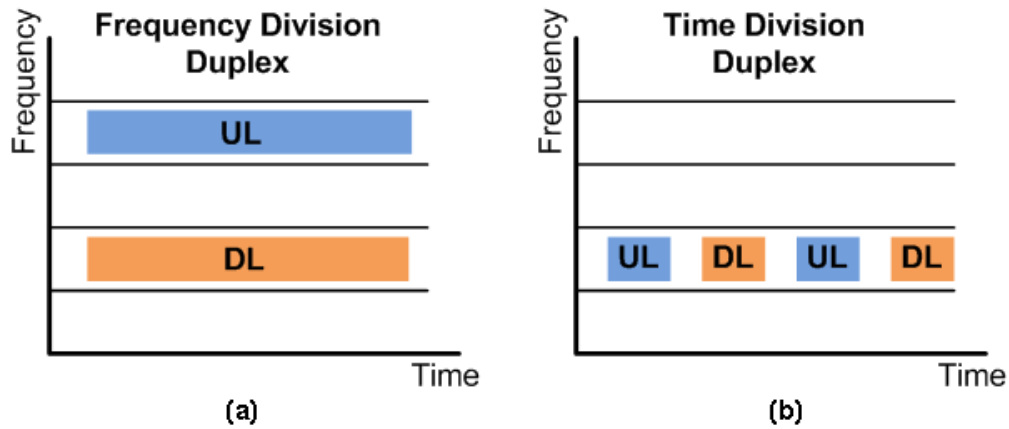
Figure 3.1: Conceptual representation of (a) FDD and (b) TDD in time and frequency domains.

improve isolation between UL and DL bands. Secondly, TDD systems can asymmetrically allocate bandwidth to UL and DL as needed. In FDD, UL and DL are each apportioned static and usually equal frequency bands. In TDD, UL and DL alternate based on need, and thus applications with asymmetric UL/DL traffic can be handled more efficiently. Lastly, because TDD systems only operate on one band, the system is less complex and lower cost.

The main drawback of TDD, on the other hand, is its higher latency. With FDD, UL and DL can operate more or less independently on their own dedicated bands. In TDD however, because UL/DL need to share the same band, handshaking protocols and guard times are required to prevent collision and interference, adding significant timing overhead. For reasons of high latency, TDD systems have also historically been limited to short-range applications.

Given its relative advantages and disadvantages, TDD has been primarily used by short-range data applications that have highly asymmetric UL/DL traffic, such as WiFi, WiMax, and Bluetooth. Meanwhile, 2G and 3G cellular standards such as GSM, UMTS, and CDMA have mostly used FDD since voice applications have symmetric UL/DL traffic and require long-range, low-latency communications. However, with the exponential growth of mobile data usage far out-stripping voice traffic in recent years, TDD has had a growing presence in 4G cellular standards and beyond. For example, Fig. 3.2 illustrates the spectrum of 3GPP LTE bands from 700MHz to 6GHz.[1] There is a mixture of both duplexing methods with LTE Bands 1-32 being FDD and Bands 33-46 TDD [24].

Going forward, TDD will maintain a significant presence in 5G cellular standards as well. Massive MIMO and beamforming, two of the main objectives of 5G systems, can be implemented in TDD with less complexity, lower cost, and more energy efficiency. Furthermore, TDD provides the potential for higher spectrum efficiency, especially as it becomes increas-

---

[1]Overlapping bands are staggered in the Y-axis for clarity of illustration. The Y-axis has no meaning.
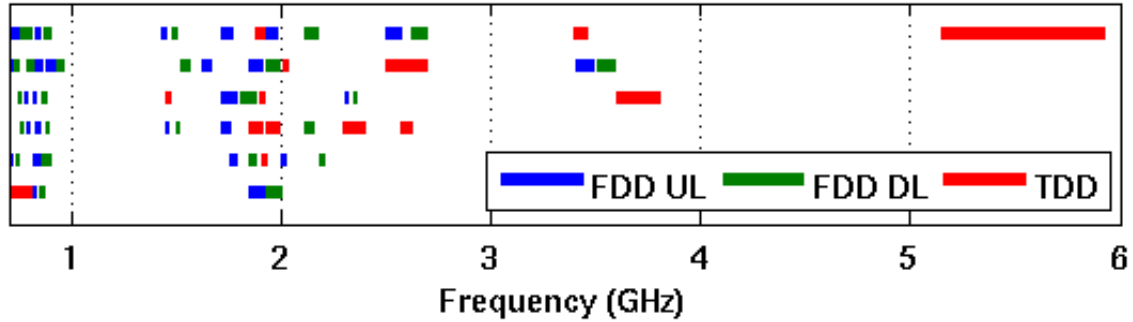
Figure 3.2: LTE FDD and TDD bands from 700MHz to 6GHz.

ingly harder to find wideband paired UL/DL bands in an already crowded spectrum.

3.1.2 T/R Switching for TDD Systems

In TDD systems, transmit and receive paths operate in the same frequency band while sharing an antenna. A switch is therefore required between the antenna and the transceiver to select between transmit and receive as well as isolate the two from each other.

Fig. 3.3 demonstrates the necessity and functions of a T/R switch (TRSW). Fig. 3.3(a) shows a conceptual diagram of a TDD transceiver front-end in receive mode. The transmitter is turned off, and all signal received by the antenna should appear at the receiver input. If RX and TX share an antenna interface as shown, the idle transmitter might provide a leakage path for the input signal, add noise, and corrupt the receiver's input impedance match. A switch is therefore required to isolate the transmitter from the RX side and present a low-noise high impedance to the sensitive RX input node.
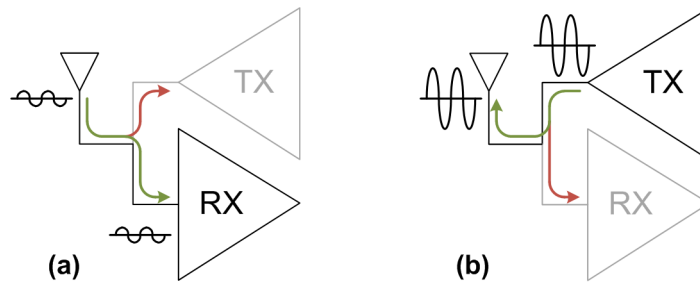
Figure 3.3: Conceptual TDD front-end in (a) receive mode, (b) transmit mode.

Fig. 3.3(b) shows the TDD transceiver in transmit mode. The receiver is now turned off, and all transmitter output signal should appear at the antenna. Similar to the receive scenario, the receiver should not provide a leakage path for the output signal nor degrade the transmitter's output impedance characteristics. However, an even more critical concern

in a transmit scenario is the PA's large output power and high voltage swings. For example, peak output power for cellular standards can be higher than 30dBm, which translates to 10V peak amplitude when referenced to a 50Ω antenna. This voltage is much higher than modern on-chip supplies and transistor tolerances. A T/R switch is thus needed to isolate the receiver not only to preserve TX performance, but also to protect the receiver itself from breakdown due to the PA's high output power.

Fig. 3.4(a) shows a simple single-pole double-throw (SPDT) switch that could be used to perform T/R selection. The switch needs to provide strong isolation between TX and RX sides for the aforementioned reasons. In addition, because the switch is inserted in series between the antenna and the transceiver, the switch must also have low insertion loss (IL). The loss of the switch impacts transmitter performance directly in the form of lower output power at the antenna. The attenuated power output also, in turn, degrades PA efficiency since the PA still consumes the same amount of DC power.
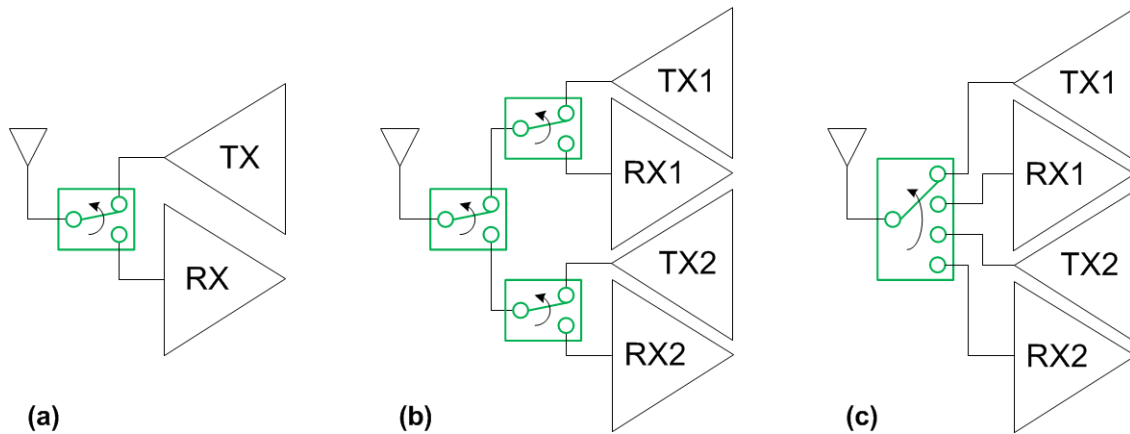
Figure 3.4: Example TDD front-ends with (a) SPDT switch, (b) cascaded SPDT switches, (c) SP4T switch.

On the receiver side, switch IL, by attenuating the input signal, increases NF and degrades sensitivity. The system RX noise factor including an input TRSW is given by:

$$F_{RX} = 1 + (F_{RAW} - 1) * IL_{SW} \tag{3.1}$$

where F_{RAW} is RX noise factor without the switch and IL_{SW} is the insertion loss of the TRSW. The NF penalty due to TRSW is roughly 0.5-1dB per dB of switch IL for realistic RX system noise characteristics, indicating that switch IL needs to be below 1-2dB to be at all useful.

In modern RF systems, a single antenna is often not used by only one transceiver but by multiple transceivers servicing different RF standards and different bands. Depending on application, TRSWs could be chained as in Fig. 3.4(b) to select between different transceivers,

or more commonly, single-pole multiple-throw (SPMT) switches could be used. Fig. 3.4(c) illustrates an example SP4T switch selecting between two transceivers, but in realistic multi-standard multi-band systems, T/R- and band-switching are performed with switch modules of up to SP10T or even higher. Switch IL increases accordingly with switch complexity.

Historically as well as in the majority of modern systems, TRSWs are discrete off-chip components. Commercial state-of-the-art TRSW modules still need and use expensive materials such as GaAs or thick-film SOI to implement pHEMT or PIN diode switching devices [25]. In addition to being costly from both BoM and PCB area perspectives, off-chip TRSWs are not reconfigurable and incur extra losses and parasitics from packages, bondwires, and PCB routing. Integrating TRSWs onto standard silicon CMOS processes, however, is challenging due to the aforementioned specifications of good isolation, low IL, and ability to withstand high PA output powers of 30dBm or more. In a survey of published state-of-the-art SoC systems, many works still do not have integrated T/R- or band-switching functionality [26][27][28][29].

3.2   Integrated T/R Switches

As discussed in Sec. 3.1.2, most TDD systems still rely on off-chip discrete components to perform T/R switching. These discrete components add BoM cost, PCB area, and losses from PCB and package parasitics. Furthermore, these components are narrowband and not configurable. Consequently, there is much interest in integrated wideband T/R switches to support modern and future multi-band radios.

In this section, Sec. 3.2.1 and Sec. 3.2.2 delves into the challenges associated with integrated TRSW design such as substrate loss, power-handling capabilities, and device breakdown. Sec. 3.2.3 analyzes a few novel integrated TRSWs from recent works, and Sec. 3.2.4 presents a survey of historical and state-of-the-art integrated TRSWs from literature.

3.2.1   Basic TRSWs and Substrate Loss

A basic T/R switch is shown in Fig. 3.5. This is the fundamental transistor topology used in both discrete GaAs switches and integrated CMOS switches. When a port - TX or RX - is active, it is connected to the antenna through a series transistor, M1 or M2. The inactive port is pulled to ground with a shunt transistor, M3 or M4. In addition, while not shown in Fig. 3.5, TRSWs are AC-coupled so there is never DC voltage drop across or DC current through any device. AC-coupling capacitors can be placed in series at the TX and RX ports, or between DC ground and the shunt switches.

All devices in a TRSW are biased through large gate resistances (R_G) and their gate nodes are high impedance. A gate node in this case swings with the signal and maintains a more consistent V_{GS}. Given that TRSWs must accommodate large signal swings of many volts from the PA, floating gate nodes are required to ensure that the device never turns off when
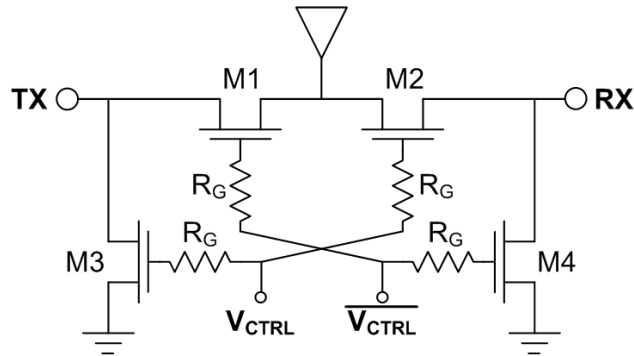
Figure 3.5: Fundamental topology for a T/R switch.

it should be on, or turns on when it should be off. A more consistent V_{GS} is also beneficial from a small signal perspective by improving switch linearity.

Fig. 3.6 illustrates a port-to-antenna interface in its active and inactive states. This could represent either TX or RX to antenna; the mechanism is the same. Fig. 3.6(a) shows the active state, where the series switch is ON, with resistance R_{ON}, to connect the port to antenna. The shunt switch is OFF, and its OFF resistance is assumed to be much greater than port impedance and not shown.
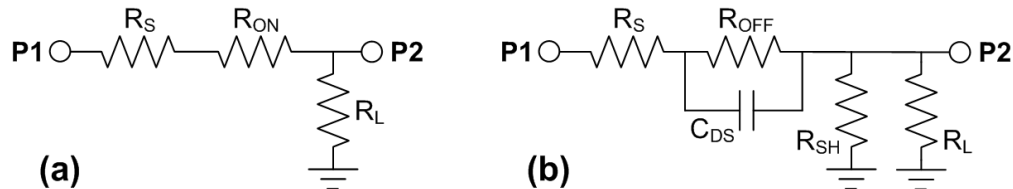
Figure 3.6: Port-to-antenna interface in (a) active state, (b) inactive state.

Insertion loss (IL) is defined as the ratio of power available from source to power delivered to a load. Given matched condition $R_S = R_L$, the contribution to insertion loss from R_{ON} is:

$$IL = \frac{P_{AVS}}{P_L} = \left(\frac{R_{ON} + 2R_S}{2R_S} \right)^2 \tag{3.2}$$

Fig. 3.6(b) shows the network when the port is inactive. The shunt switch is ON, and it pulls the port to AC ground with ON resistance R_{SH}. The series switch is OFF to disconnect the port from the antenna. However, some portion of the signal will leak through C_{DS}, the total capacitive coupling from the series device's drain to source.

In the inactive state, the ratio of power available from source to power delivered to the load is defined as isolation, where a high isolation with minimum power delivered is desirable. Given that R_{OFF} is very large, in most practical RF designs, the combination of C_{DS} and R_{SH}

determines isolation performance. Larger devices with higher W/L ratios have smaller R_{ON}, but greater parasitic capacitances. Therefore, a larger series transistor will contribute less IL but also less isolation. This is the fundamental trade-off of traditional GaAs TRSWs [30].

We now examine the active state in more detail. Eq. 3.2 considered only the low-frequency case; at RF, parasitic capacitances need to be taken into account. Fig. 3.7(a) shows a schematic of the active port network with the series transistor's capacitances added. The gate node is high-impedance as previously mentioned, and the body of the device is connected to ground through some amount of substrate resistance R_{SUB}. Assuming R_{ON} is small, the circuit can be simplified to the schematic in Fig. 3.7(b) [31]. Resistance R_{ON} is extracted, and all capacitances are combined into some effective capacitance C_{EFF} to the substrate.
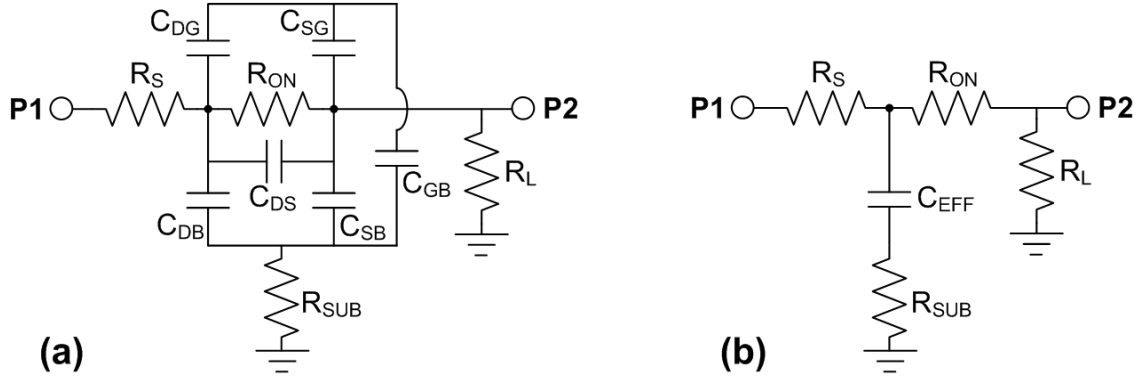
Figure 3.7: Schematic of port network in active state at RF: (a) detailed, (b) simplified.

In GaAs and some SOI processes, substrate resistance is extremely high. Capacitance C_{EFF} has little effect as it just couples to a high-impedance. In bulk silicon however, substrate resistance is much lower and can cause significant IL degradations as the signal leaks through the capacitances to the substrate. Including the effects of R_{SUB} and C_{EFF}, the insertion loss of the network is:

$$IL = \left(\frac{R_{ON} + 2R_S}{2R_S} \right)^2 \left| \frac{1 + sC_{EFF} \left(R_{SUB} + R_S \frac{R_{ON}+R_S}{R_{ON}+2R_S} \right)}{1 + sC_{EFF} R_{SUB}} \right|^2 \tag{3.3}$$

The insertion loss from Eq. 3.2 is now multiplied by a frequency-dependent factor that is always greater than 1. As expected from looking at Fig. 3.7(b), IL increases with C_{EFF} until a limit where C_{EFF} acts like an AC short and R_{SUB} determines maximum possible IL.[2] In discrete TRSWs that use processes with high substrate resistance, larger switch transistors always resulted in smaller IL, albeit at the cost of worse isolation. On bulk silicon however, there is a counter-effect from the device capacitance that degrades IL, and there is a point

---

[2]For a given C_{EFF} however, IL has a maximum at the point where $R_{SUB} = 1/(\omega C_{EFF})$ [31].

where increasing device size will start to worsen IL. This demonstrates one of the challenges of integrated TRSWs.

3.2.2 Power Handling

Another common limitation of integrated CMOS TRSWs is their power-handling capability. Specifically, this refers to the maximum amount of PA output power the switch can tolerate before either IL increases dramatically or the device breaks down altogether. To explore the factors that affect and constrain power-handling ability, Fig. 3.8 illustrates a basic TRSW in TX and RX modes.

Fig. 3.8(a) shows a scenario in which TX port is active and RX is isolated. Switches M1 and M4 are ON while M2 and M3 are OFF. The large PA output signal is propagated to the antenna through M1, while M4 pulls RX input to AC ground. M2 and M3, the OFF devices, must both withstand the full TX signal swing across their drain-to-source junctions. For WLAN and cellular applications, peak transmit power has voltage amplitudes that are many times higher than the oxide breakdown voltage of modern CMOS transistors. This presents a limitation on integrated TRSWs' power handling capabilities, although we will later discuss techniques for improvement.
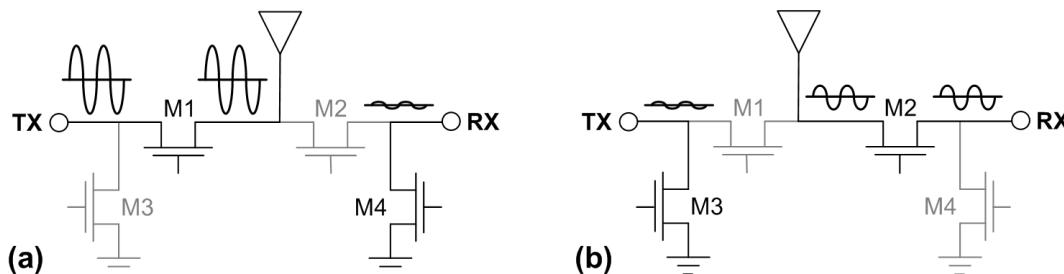
Figure 3.8: TRSW in (a) TX and (b) RX scenarios.

In RX mode in Fig. 3.8(b), the received signal is of small to moderate strength, and no device in the system experiences voltage swings beyond its breakdown limit. Thus, the stressed devices are only M2 and M3, and for this reason, M3 is often omitted in integrated TRSW designs. M3 presents a significant design challenge while its only benefit is increased TX isolation, which, unlike RX isolation, is not a critical metric. RX isolation ensures that high TX output signals will not inadvertently turn the LNA on or cause device breakdown. The purpose of TX isolation, on the other hand, is to prevent the inactive PA from adding noise and degrading LNA performance in RX mode. Sufficient TX isolation can generally be achieved with careful design of M1 alone, or by ensuring that the PA has high output impedance when it is inactive.

However, the RX series switch M2 must still be able to accommodate PA output voltage swings of 10V or more. CMOS transistors can handle maximum AC voltage amplitude of

$2 * V_{DD}$, where V_{DD} is the devices' maximum DC voltage rating. For modern processes, AC amplitudes are limited to 2.4V for thin-oxide devices ($V_{DD} = 1.2$V), and up to 6.6V for thick-oxide IO devices ($V_{DD} = 3.3$V). Thick-oxide devices present a significant improvement in power-handling ability at the cost of larger area and higher capacitances, but they are nevertheless inadequate for practical wireless applications.

To improve power-handling capabilities, transistors must be stacked in series as shown in Fig. 3.9(a). All devices are OFF, and ideally, the total voltage amplitude splits equally across each transistor, so that the drain-to-source voltage of any one device is within safe limits. Fig. 3.9(a) shows an example stack with 3 devices, but actual required number of devices would be determined by the type of transistor used (thin- or thick-oxide) and desired power tolerance. Similar to using longer gate lengths, a stacked topology suffers from higher total ON resistance, larger area, and higher parasitic capacitances. For a stack of x devices, effective gate length becomes $L_{EFF} = x * L$, total $R_{ON} \propto W/L_{EFF}$, and $C_{EFF} \propto W * L_{EFF}$.[3]
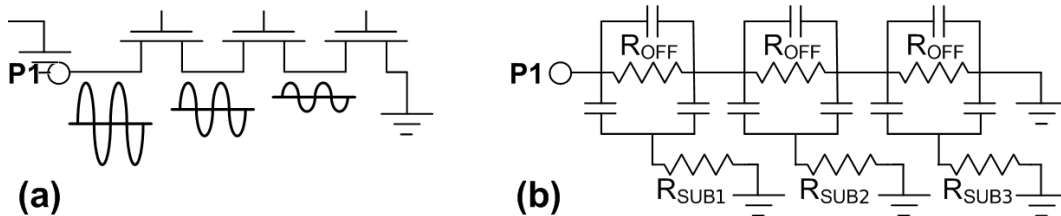
Figure 3.9: Stacked switch in OFF mode: (a) schematic, (b) high-frequency circuit model.

Fig. 3.9(b) shows a circuit model of a stacked switch at RF with parasitics included. Transistors in the stack experience different substrate resistances depending on layout and distance to substrate tap. At low frequencies, or if substrate resistance were very high, R_{SUB} can be neglected. The drain-to-source impedance of each device would be equal, and voltage would split evenly amongst the transistors in the stack as in a voltage divider. At RF and with low substrate resistance however, the devices' effective drain-to-source impedances become unequal due to unequal R_{SUB}, and signal voltage may not be evenly distributed. Often extra margin is required to ensure robustness. Alternatively, explicit source/drain-to-gate or source/drain-to-body capacitors can be manually and selectively placed to equalize signal voltage distribution [32].

In addition to gate oxide breakdown, substrate and well diodes present another limitation on the power-handling capabilities of integrated TRSWs. Fig. 3.10(a) illustrates a physical diagram of an NMOS transistor in bulk silicon. If there is a large voltage swing on the source or drain of the device, and if R_{SUB} is small, the substrate-to-source/drain diodes will

---

[3]Effective ON resistance and parasitic capacitances will be even worse in practice due to larger physical layout.

become forward-biased. The diodes turning on incurs a large leakage current that degrades IL dramatically. In contrast to gate oxide breakdown, the substrate-diode limitation on power-handling exists regardless of whether the switch is on or off. The diodes not only limit M2 and M3 in Fig. 3.8 while they are off, but also limit M1 while it is on in TX mode.
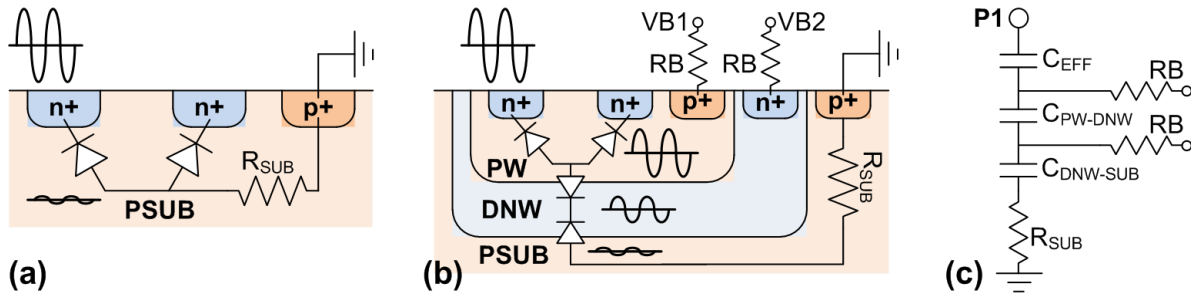
Figure 3.10: Diagram of a physical NMOS switch: (a) without isolated P-well, (b) with isolated P-well, (c) circuit model of triple-well structure.

With the ubiquity of triple-well capability in modern CMOS processes, the substrate diode issue can be rectified by placing NMOS switch devices in isolated P-wells. Fig. 3.10(b) illustrates an NMOS in a triple well processes. The isolated P-well (PW) is separated from the larger substrate (PSUB) with a deep N-well (DNW) layer. The P-well tap is biased through a large resistor R_B so that the well presents a high impedance and swings along with the signal at the device's source or drain. The deep N-well is generally biased through a large resistor as well for improved power-handling capabilities.

Fig. 3.10(c) shows an equivalent circuit model for the triple-well structure. The capacitance C_{EFF} is the the device's parasitic capacitance, which used to couple directly to the low substrate resistance, but now goes through several layers of well capacitance before reaching R_{SUB}. Assuming R_B is large, the capacitances act like a capacitive voltage divider and gradually attenuate the signal as illustrated in Fig. 3.10(b). This effect, along with biasing each well at an appropriate DC voltage, ensures that no substrate or well diode ever becomes forward-biased. Furthermore, because the effective capacitance to R_{SUB} is reduced in a triple-well process, IL is also improved per Eq. 3.3.

The penalty for using isolated P-wells, on the other hand, is greater layout area and its associated drawbacks. There are layout design rules governing minimum enclosure distance for each well layer, and minimum gap distance between each isolated well. For a stacked switch, there are multiple devices each needing its own isolated P-well, and the resulting area penalty can be high. Larger layout area results in additional degradations from parasitic resistances and capacitances due to routing.

3.2.3   Alternate TRSW Topologies

In recent state-of-the-art, alternate topologies for integrated TRSWs have been introduced to address some of the drawbacks of the basic TRSW design from Sec. 3.2.1. The basic topology was adapted directly from discrete GaAs TRSWs and modified to compensate for the shortcomings of bulk silicon. Newer integrated TRSW topologies, on the other hand, evolved to exploit the advantages of integration and work symbiotically with the rest of the TDD system.

One primary weakness of the basic TRSW design is the necessity for series-stacked transistors to meet robustness and power-handling specifications. The stacked switch significantly increases IL, parasitics, and area, and is a dominant limitation on the design of high-frequency, low-loss integrated TRSWs. Assuming the TX shunt switch (M3 in Fig. 3.5) is omitted, the only switch that encounters high voltage swings across its source-to-drain junction is the RX series switch (M2 in Fig. 3.5). One technique for high-performance integrated TRSWs is to eliminate this RX series switch altogether by absorbing its function into the LNA input match network.
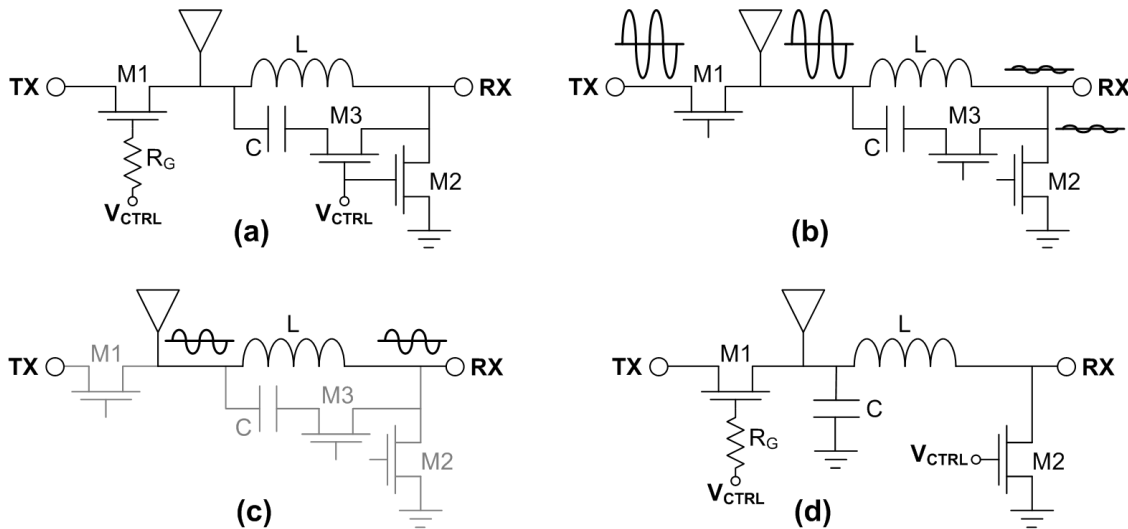
Figure 3.11: Inductor-resonance TRSW: (a) schematic, (b) TX mode, (c) RX mode, (d) alternate topology.

Fig. 3.11(a) shows a TRSW topology in which the RX series switch has been eliminated [33]. Placing an inductor in series at the LNA input is a common impedance matching technique, mostly for inductively degenerated common-source LNAs, but can be adapted for other LNA topologies as well. The design takes advantage of the series inductor, an already-existing component of the LNA front-end, and utilize it to perform T/R switching.

Fig. 3.11(b) shows the T/R switching network in TX mode. The TX series switch (M1) as

well as both of the RX shunt switches (M2 and M3) are ON. The inductor and capacitor form a parallel LC tank that resonates and presents a high-impedance at the signal frequency. M2 and M3 are both functionally RX shunt switches to ground; they provide isolation at the RX port and do not experience the large transmit signal swings. Instead, the large TX voltage swing occurs across the inductor and capacitor, which, as purely passive components, can easily be designed to withstand high powers without having to compromise on loss.

Fig. 3.11(c) shows the T/R switching network in RX mode. All switches are OFF, TX is isolated by M1, and the LNA is connected directly to the antenna through the series inductor matching network as if no TRSW exists. The challenging and high-loss series RX switch has thus been eliminated. The TX series switch M2 still exists and still requires an isolated P-well. However, compared to the RX series switch, the TX switch does not require stacked transistors and is therefore much less lossy. Fig. 3.11(d) shows a variation of the inductor-resonance TRSW topology where the capacitor is absorbed into the RX matching network [34]. This eliminates the capacitor switch M3 which de-Qs the LC tank.

The main limitation of the TRSW topology in Figs. 3.11(a) and (d) are that they are inherently narrowband. They depend on the resonance of the LC tank, which, to present a high impedance in TX mode and provide good RX isolation, fundamentally must be high-Q.

An alternate TRSW structure that can be wideband is shown in Fig. 3.12(a) [35]. Two transformers are stacked in series, one each for TX and RX ports, and the RX input and TX output are combined through the transformer structure at the antenna port. This transformer-based TRSW again exploits existing transceiver front-end components - transformers and integrated balun - to provide T/R switching function.
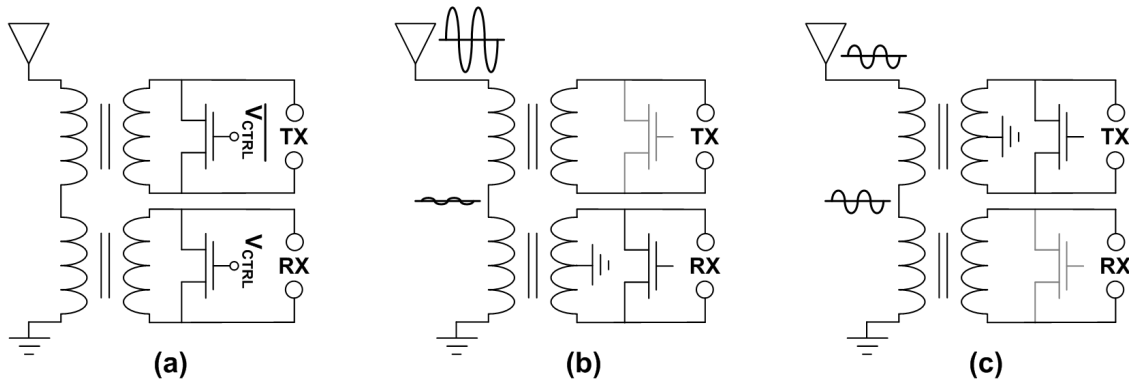
Figure 3.12: Transformer-based TRSW: (a) schematic, (b) TX mode, (c) RX mode.

Each transformer has a shunt switch that controls whether the TX or RX port is active. Fig. 3.12(b) shows the TRSW in TX mode. The RX switch is connected to DC ground and turned ON, which shorts both the bottom transformer winding and RX port. Like the inductor-resonance TRSW from Fig. 3.11, no RX series switch is required for RX isolation,

and the high TX voltage swings are absorbed across passive components - in this case, the top transformer winding. The TX shunt switch, however, like M3 in Fig. 3.5, must still be designed to withstand large AC voltages. Fig. 3.12(c) shows the transformer-based TRSW in RX mode. The TX switch now turns ON to short the upper transformer winding and the TX port, and the RX is connected to the antenna effectively through just the RX transformer.

In contrast to the inductor-resonance TRSW, this transformer-based TRSW does not rely on high-Q resonance and is thus more wideband. The transformer-based TRSW does need a TX shunt switch however, which potentially contributes significant parasitics and loss and is the major drawback of this topology. If the TX shunt switch is not sufficiently low resistance, it could de-Q the RX transformer and increase RX IL similar to an RX series switch. On the other hand, an advantage of the TX shunt switch over the RX series switch is that its parasitic capacitances can be absorbed into the transformer frequency response. This leads to better high-frequency performance because the switch capacitances do not direct affect sensitive parameters such as RX isolation or IL through substrate resistance.

3.2.4 State-of-the-Art in Integrated TRSWs

The earliest integrated TRSW works mimicked the symmetric series-shunt topology of Fig. 3.5 directly from discrete GaAs switches, and consequently, encountered the insertion loss and power-handling limitations described in Sec. 3.2.1 and Sec. 3.2.2. For example, [31] demonstrated the first CMOS switch with reasonable performance in a 0.5μm 3V process and for 900MHz applications. The standalone switch achieved 0.7dB IL, 42dB isolation, and 17dBm P1dB. The switch IL, while worse than discrete counterparts, was adequate, and it was achieved by optimizing switch size, its on resistance, its parasitic capacitance, and substrate resistance. Reasonable isolation performance was achieved as well. However, linearity was very limited despite using 3V devices and a 6V control voltage.

Another early integrated TRSW work, [30], used a 0.18μm 1.8V process and targeted 2.4GHz BT and WLAN applications. This work achieved 1.5dB IL, 24dB isolation, and 11dBm P1dB. As shown, all performance metrics degraded significantly at higher frequencies and with a more scaled process. This demonstrated the limitations of the basic series-shunt TRSW topology in bulk silicon.

More recent works, utilizing the advances in silicon processing technology, have had more success. For example, [36] implemented the series-shunt topology in a 0.13μm triple-well process. The work was targeted at both 900MHz and 2.4GHz applications, with the switch being wideband up to about 3GHz. The TRSW used thick-oxide 0.26μm 3.3V devices, placed them in isolated floating P-wells, and tripled-stacked them in series for the RX series and TX shunt switches. Feedforward capacitors were added to equalize the voltage drop across each device in a stack.

With the combined effects of the aforementioned techniques, the work in [36] achieved P1dB

of 31.3dBm at 900MHz and 28dBm at 2.4GHz. This is an order-of-magnitude improvement in power-handling capability over the earliest works from [31] and [30], and demonstrated the feasibility of integrated TRSWs for practical cellular and WLAN applications. Performance gains in other metrics, however, were limited. The TRSW achieved TX and RX ILs of 0.5dB and 1dB respectively at 900MHz, and 0.8dB and 1.2dB respectively at 2.4GHz. This is comparable to the early works and still somewhat worse than discrete TRSWs. RX IL has obviously been degraded by the use of stacked-transistor switches. This work achieved TX and RX isolations of 29dB and 24dB respectively, which, again, is comparable to the early works and worse than discrete TRSWs.

The work in [37] used similar techniques - series-shunt topology, stacked-transistor switches, isolated wells - to implement an SP3T switch with an integrated LNA. This work was implemented in a 0.18μm 3.3V process and targeted at 2.4GHz BT and WLAN applications, although the switch itself is wideband similar to [36]. The SP3T switch selects between TX, RX (with attached LNA), and BT ports. All series and shunt switches are composed of either double- or triple-stacked transistors. Similar to [36], this work achieved excellent linearity of 33dBm P1dB, and moderate ILs of 1.3dB for TX and 1.45dB for BT. LNA NF is 1.5dB standalone and 3dB including the switch. While the switch adds a significant NF increase, the total NF is nevertheless reasonable for practical applications. The achieved isolation is 28dB, similar to other works.

The performance exhibited thus far could also be attained at much higher frequencies. The TRSW in [38] targeted UWB applications from DC to 20GHz. To accommodate the significantly higher frequency, [38] used a series-only topology due to the shunt switches adding too much parasitic capacitance. Both TX and RX shunt switches were eliminated from the design. To achieve good isolation, the design relied on increasing finger spacing in layout to minimize drain-to-source capacitance, at the cost of larger area and higher drain/source-to-body capacitance. With these modifications, and in a 0.13μm 1.2V process with 2V control signals, this work achieved 30dBm P1dB, 2dB IL, and 21dB isolation at its highest frequencies up to 20GHz.[4] Performance was only slightly degraded compared to other works at low GHz frequencies. In fact, the high-frequency techniques improved performance at lower frequencies as well; for example, this work achieved 0.9dB IL and 42dB isolation at 2.4GHz.[5]

The inductor-resonance TRSW topology described in Sec. 3.2.3 has shown superior performance over the traditional series-shunt topology (and its variations), albeit with the tradeoff of being narrowband. The work in [33] pioneered the inductor-resonance topology and implemented the TRSW shown in Fig. 3.11(a) with integrated PA and receiver. The design was targeted at 2.4GHz WLAN applications. In a 90nm 1.2V process and without the usage of isolated wells, the TRSW achieved TX and RX IL of 0.4dB and 0.2dB respectively, TX

---

[4]This design used a fully-differential topology, which automatically provides an extra 3dB in power-handing ability. An external balun would have been required to interface with a single-ended antenna.

[5]Power handling was somewhat worse at lower frequencies, with P1dB=24dBm at 1GHz.

and RX isolation of 16dB and 30dB respectively, and 30dBm P1dB.[6]

The elimination of both the RX series and TX shunt switches allowed [33] to attain good linearity without a triple-well process. The TX series switch was optimized for high power-handling and low IL by optimizing substrate resistance through careful control of well area, spacing, and contact placement. Low RX IL was achieved due to the lack of RX series switches, and overall, [33] had excellent IL performance on par with or even better than its discrete counterparts. However, the switch required precise EM simulations of the substrate to ensure performance and robustness.

The work in [39] replicated the work from [33] in a 32nm process. The TRSW in [39], using the same topology and targeting the same frequency band as [33], achieved TX and RX IL of 1.3dB and 1.1dB respectively, 32dB RX isolation, and 34dBm TX P1dB.[7] Although IL worsened significantly, it was nevertheless reasonable when compared against other integrated TRSWs and adequate for the target application. With great linearity and moderate but sufficient IL and isolation, [39] verified the feasibility of integrated TRSWs at highly scaled technology nodes.

The alternate inductor-resonance TRSW topology shown in Fig. 3.11(d) was implemented by [34] in a 55nm process and also targeted at 2.4GHz WLAN. The integrated transceiver including the TRSW had 29dBm PA saturated power and 3.3dB LNA NF, which were adequate for the target application. TX IL was 0.6dB. This work further confirmed the high power-handling and low IL properties of the inductor-resonance TRSW.[8]

Although [33], [39], and [34] all achieved good integrated TRSW performance, these works were all narrowband and targeted specifically at 2.4GHz applications. Consequently, they contribute limited utility towards the area of integrated front-ends for multi-standard, multi-band radios. To remedy this, the authors of [35] proposed the transformer-based TRSW from Sec. 3.2.3 and Fig. 3.12 as a wideband integrated TRSW for multi-band radios.

The work in [35] used thick-oxide 3.3V devices from a 90nm process. The TRSW had a wide passband with -1dB bandwidth from 5GHz to 7GHz. The switch had good linearity and isolation performances, with 35.7dBm P1dB and 42dB isolation.[9] However, the switch incurred significantly higher loss. TX and RX IL was 2.65dB and 2.52dB respectively, more than 1dB above other works. Additionally, broadband impedance match was not achieved, and S11 was only below -6.5dB over the 5GHz to 7GHz band.[10]

The work in [40] adopted the transformer-based TRSW into a fully integrated transceiver

---

[6]This work was also fully differential and required an off-chip balun.

[7]Isolation performance was estimated from simulations.

[8]In [34], both the series inductor and shunt capacitor used for LNA matching network and T/R switching were implemented off-chip.

[9]P1dB was calculated by [35] from measured IIP3.

[10]The TRSW did achieve good impedance match over a narrower bandwidth at midband.

front-end and made several improvements. Firstly, while [35] used identical transformers and switches for the TX and RX ports, [40] optimized each front-end passive network to the specific needs of the PA and LNA, resulting in asymmetric transformers and switching structures. For example, the LNA used a 1:3 transformer while the PA used 1:1.[11] Secondly, the system used the PA itself to short the TX transformer during RX mode instead of an additional explicit switch. This eliminated the challenging TX shunt switch, which was a significant contributor to parasitics and IL. Thirdly, a capacitor was inserted in series in between the two stacked transformers on the antenna-side winding. The capacitor gave an additional degree of freedom to the matching network for improved PA/LNA impedance co-match. With the modified matching network, this design, in contrast to [35], became narrowband.

The system in [40] was implemented in a 90nm process and targeted at 5GHz applications. The full TDD front-end, including TRSW, achieved 3.2dB LNA NF and 25.9dBm PA saturated power at 5.2GHz. No standalone TRSW measurements were made, but IL improvements over [35] can be extrapolated from the low NF. This work validated the feasibility of the transformer-based TRSW topology in practical applications, but only for a narrowband system. High-performance, wideband integrated TRSWs for multi-band radios still have not been demonstrated.

A wideband TDD front-end from 1.3GHz to 3.3GHz for LTE applications was presented in [32]. To achieve broadband T/R switching, [32] returned to the traditional series-shunt TRSW topology. The RX port had a stacked series switch with 5 transistors and a standard RX shunt switch. The TX port connected directly to the antenna, with both the TX series and shunt switches omitted. The exclusion of the TX switches was possible due to the system's use of a digital PA. When it is off, a digital PA exhibits high output impedance, and the transceiver in [32] depended on its own PA output impedance to provide TX isolation.

One major drawback of [32] was its dependence on frequency tuning to achieve the targeted wide bandwidth. The system had a capacitor bank at the antenna port, and the capacitor bank switches, when they are off, encounter the full TX signal swing across their drain-to-source junctions. Although the TX series and shunt switches were omitted in the system, the capacitor bank switches effectively were TX shunt switches from a switch design perspective, with all the associated costs. The capacitor bank switches were implemented as stacked switches with 4 transistors.

The system in [32] was implemented in 45nm SOI. Because the PA output was directly joined with the antenna, the PA had good broadband performance, achieving 27.7dBm saturated power and 25% to 30% total efficiency. On the other hand, the RX switch IL was estimated to be about 2dB - high due to the stacked RX series switch - and the LNA NF varied widely between 2.8dB and 6dB across the targeted 1.3GHz to 3.3GHz band. Input impedance

---

[11]In [35], both transformers were 1:1.

match for RX was not reported.

We discuss one final state-of-the-art TDD front-end. The work in [41] is once again narrowband targeted at 2.4GHz BT and WLAN applications. However, it is of some interest as it combined several integrated TRSW techniques to create an SP3T switch for TX, RX, and BT ports.[12] The TX and RX ports used the inductor-resonance TRSW topology from Fig. 3.11(d). Meanwhile, the BT port was connected in series with the PA output balun transformer, similar to a transformer-based TRSW, but with the key difference of omitting the BT transformer and connecting the port directly.

When [41] is in TX or RX mode, a shunt switch grounds the BT port, similar to the RX shunt switch from Fig. 3.12(b), and the system operates as a standard inductor-resonance TRSW. In BT mode, the PA shorts its output transformer with its own devices as in [40], the BT shunt switch turns off, and BT is connected through the shorted transformer winding directly to the antenna. Because BT has a much lower maximum output power than WLAN or cellular, the BT shunt switch is less burdensome to design and incurs less parasitics and loss. This system in [41], including integrated SP3T, achieved 27.8dBm PA saturated power and 3dB RX NF in a highly scaled 28nm process.

Table 3.1 presents, in chronological order, a summary of the switch characteristics from the works described in this section.

Table 3.1: Summary of integrated TRSW works.

	Technology	System	Freq. (GHz)	SW Topology	IL (TX/RX) (dB)	Isolation (TX/RX) (dB)	P1dB (dBm)
[31]	0.5μm	TRSW	0.9	series-shunt	0.7	42	17
[30]	0.18μm	SW+PA+LNA	2.4	series-shunt	1.5	24	11
[36]	0.13μm	TRSW	0.9-2.4	series-shunt	0.5-0.8/1-1.2	24-29	28-31.3
[38]	0.13μm	TRSW	0-20	series only	0.7-2	21-42	24-30
[33]	90nm	SW+PA+RX	2.4	inductor res.	0.4/0.2	30/16	30
[37]	0.18μm	SP3T+LNA	2.5	series-shunt	1.3/2.7	28	33
[39]	32nm	SW+PA+RX	2.4	inductor res.	1.3/1.1	-/32	34
[35]	90nm	TRSW	5-7	transformer	2.65/2.52	42	35.7
[34]	55nm	SW+PA+RX	2.4	inductor res.	0.6/NF=3.3dB	-	29
[32]	45nm SOI	SW+PA+LNA	1.3-3.3	RX series-shunt	2	-	27.7
[40]	90nm	SW+PA+LNA	5.2	transformer	-/NF=3.2dB	-	25.9
[41]	28nm	SP3T+PA+RX	2.4	inductor res.+xfmr	-/NF=3dB	-	27.8

---

[12] "TX" port is WLAN TX. "RX" port is shared RX for both WLAN and BT. "BT" port is BT TX.

3.3   Proposed Wideband T/R Switching Technique

TDD co-existence with high-performance integrated TRSWs have thus far been limited to narrowband systems serving 2.4GHz BT and WLAN applications These works rely on sharp resonance to provide adequate transmit/receive isolation, and cannot be readily adapted for wideband or reconfigurable radio applications.

Wideband solutions have incurred higher loss, demonstrating ILs of ~1-1.5dB for standalone switches [36][37], ~2dB when integrated with a transceiver [32], and ~2.5dB at higher frequencies above 5GHz [35]. Isolation performance have also been limited to the ~20-30dB range. In contrast, at gigahertz frequencies, discrete TRSWs can have ILs of ~0.5dB and isolation greater than 40dB [25].

A primary bottleneck gating the design of low-loss, wideband TRSWs is the RX series switch. The need for stacking transistors in series degrades IL and adds parasitic capacitances significantly. IL at higher frequencies is especially limited. Furthermore, to ensure robustness, designs often require isolated wells, accurate knowledge of substrate characteristics, and fine-tuning substrate impedance through layout [37][33]. These components all add to design complexity and die area.

In this section, we propose an integrated T/R switching technique for wideband TDD co-existence that eliminates both the TX and RX series switches. Instead of using switches at all to isolate and select between PA and LNA blocks, we re-use the PA as an LNA during receive mode. Fig. 3.13 illustrates the proposed scheme conceptually.
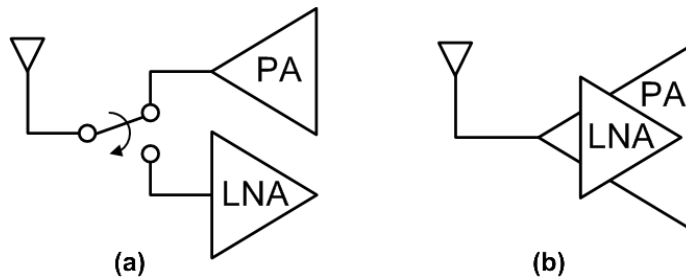
Figure 3.13: Conceptual diagram of TDD front-end: (a) conventional, (b) proposed.

Isolation is no longer an issue as the PA and LNA are the same block, with a single antenna port. Power and control switches are used to enable the PA to LNA transformation. These switches, in contrast to conventional TRSW devices, are DC and insensitive to parasitic capacitance and other high-frequency sources of loss.

Sec. 3.3.1 gives an overview of the inverse class-D PA that we will re-use for our proposed T/R switching scheme. Sec. 3.3.2 describes the detailed mechanism of the PA to LNA transformation.

3.3.1 Switching PAs

Digital switching PAs have been gaining popularity both as an area of research and in commercial systems due to its potential for higher efficiency. PA efficiency, denoted by η, is defined in Eq. 3.4.[13] Efficiency measures the relationship between a PA's output power (P_L) and the DC power drawn from the PA supply (P_{SUP}). Ideally, efficiency is desired to be as high as possible up to 100% so that all power drawn from the supply is delivered to the load, and none wasted on overhead.

$$\eta = \frac{P_L}{P_{SUP}} \tag{3.4}$$

The classic class-D switching PA is shown in Fig. 3.14(a). The PA is essentially an inverter, and it outputs a square wave at its drain node. The series LC at the output filters harmonics and allows only sinusoidal current at the fundamental frequency to reach the load. The ideal drain current and voltage waveforms of the PA are plotted in Fig. 3.15(a). Because the drain current and voltage are orthogonal, overhead power is zero and efficiency is 100% in the theoretical ideal case. In contrast, conventional analog class-A and class-B PAs have maximum theoretical efficiencies of 50% and 78.5% respectively.

The class-D PA is a voltage-switching PA; its current-switching dual is the inverse class-D PA shown in Fig 3.14(b). The inverse class-D PA is essentially a current source driven by an input square wave, resulting in a square wave drain current at the PA's output. In duality with class-D, the drain voltage of the inverse class-D PA is filtered into a half-wave rectified sine wave by a parallel LC tank at the output. Fig. 3.15(b) shows the PA's ideal drain current and voltage waveforms, and they are also orthogonal and result in theoretical 100% maximum efficiency.
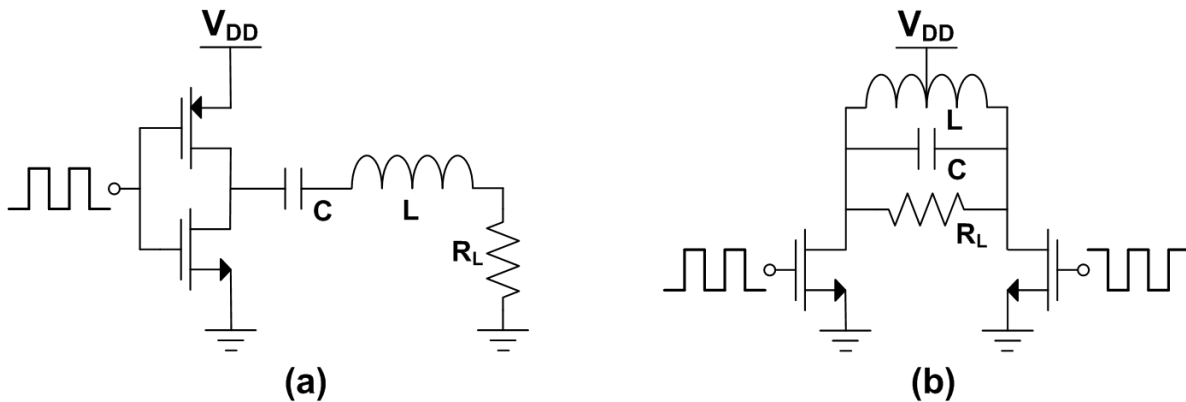
Figure 3.14: Circuit schematic of switching PA: (a) class-D, (b) inverse class-D.

---

[13]All discussions of PA efficiency in this thesis, unless otherwise specified, refer to drain efficiency.
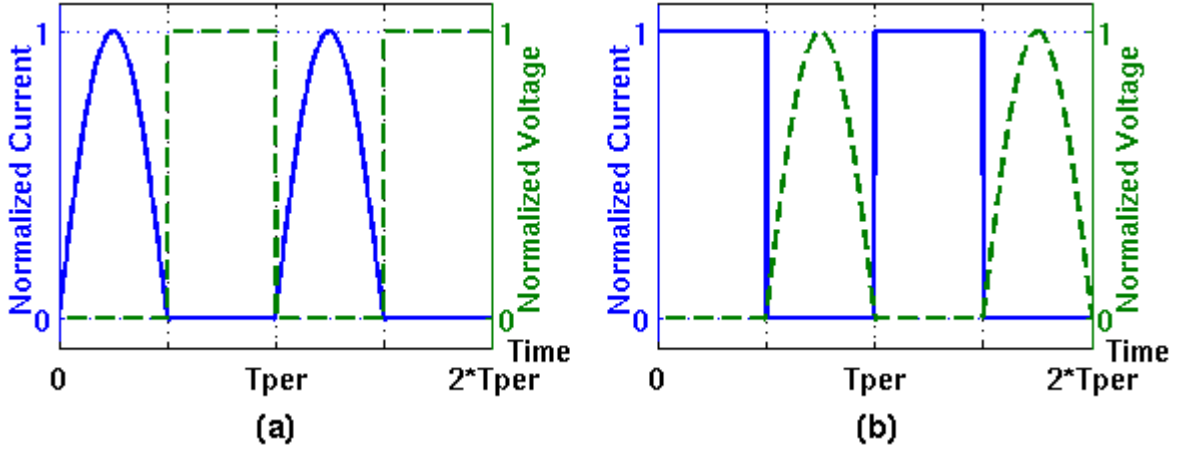
Figure 3.15: Normalized drain current and voltage waveforms of switching PA: (a) class-D, (b) inverse class-D.

A main advantage of the inverse class-D is that the PA's large drain capacitance, as well as any other parasitic capacitances at the drain node, is absorbed into the output passive network. In comparison, in a voltage-switching class-D PA, the PA's drain capacitance works in opposition to and degrades the fidelity of the output LC filter. In a practical implementation, the bottom plate capacitance of the series capacitor is also a significant source of loss. For these reasons, the inverse class-D PA has demonstrated better power and efficiency performance, and we use this PA topology for our proposed T/R switching technique.

Next, we briefly analyze and discuss design considerations for an inverse class-D PA. When the current source switch is on, some current I_{PK} flows through the device. The switch also has some finite ON resistance R_{ON} which prevents the drain voltage from reaching zero. We refer to this non-zero low voltage as the PA's knee voltage V_K [42]. When the current source is off, the drain voltage is sinusoidal with a peak voltage that we denote V_{PK}. These drain current and voltage relationships are summarized in Eq. 3.5.

$$
\begin{aligned}
I_{SW-ON} &= I_{PK} & V_{SW-ON} &= V_K = I_{PK} * R_{ON} \\
I_{SW-OFF} &= 0 & V_{SW-OFF} &= V_{PK}
\end{aligned}
\tag{3.5}
$$

The drain current in time domain is a square wave with a low of zero and a high of I_{PK}. The drain voltage is V_K plus a half-wave rectified sine with amplitude $V_{PK} - V_K$. The components at only the fundamental frequency are delivered to the load, and Eq. 3.6 gives the amplitudes of the current and voltage waveforms at the fundamental [42]. In addition, the drain voltage is DC-biased through the inductor to PA supply V_{DD}; therefore, the DC component of the

drain voltage waveform is exactly equal to V_{DD} and given in Eq. 3.7 [42].

$$I_{FUND} = \frac{2}{\pi} I_{PK}$$
$$V_{FUND} = V_{PK} - V_K$$

(3.6)

$$V_{DC} = V_{DD} = V_K + \frac{1}{\pi}(V_{PK} - V_K)$$

(3.7)

Eq. 3.8 defines the power delivered to the load P_L as a function of voltage and current amplitudes at the fundamental, as well as a function of the current and load resistance R_L [42].

$$P_L = \frac{1}{2} * V_{FUND} * I_{FUND} = I_{PK}(V_{DD} - V_K) = I_{PK}(V_{DD} - R_{ON} * I_{PK})$$
$$P_L = \frac{1}{2} * I_{FUND}^2 * R_L = \frac{2}{\pi^2} * I_{PK}^2 * R_L$$

(3.8)

Given Eqs. 3.6, 3.7, and 3.8, PA parameters I_{PK} and P_L can be defined in terms of design variables V_{DD}, R_{ON}, and R_L [42]:

$$I_{PK} = \frac{V_{DD}}{\frac{2}{\pi^2}R_L + R_{ON}}$$
$$P_L = \frac{2}{\pi^2}R_L \frac{V_{DD}^2}{\left(\frac{2}{\pi^2}R_L + R_{ON}\right)^2}$$

(3.9)

As seen from Eq. 3.9, higher supply voltage results in larger current and greater PA output power. For a given supply voltage, on the other hand, higher output power can be obtained with a smaller load resistance. Transistor oxide breakdown constrains the maximum peak drain voltage that can be tolerated, and this in turn constrains the maximum supply voltage that could be used.[14] Thus, to attain high output power with moderate V_{DD}, low load resistances smaller than 50Ω are typically used.

Switch resistance R_{ON} degrades PA output power as expected. Furthermore, the switch consumes some amount of power ($I_{PK}^2 * R_{ON}$) that does not reach the load. Taking into account finite switch resistance, the efficiency of an inverse class-D PA becomes [42]:

$$\eta = \frac{1}{1 + (\frac{\pi^2}{2})(\frac{R_{ON}}{R_L})}$$

(3.10)

A circuit diagram of a common, practical implementation of the inverse class-D PA is shown in Fig. 3.16(a). There are two main modifications from the basic topology from Fig. 3.14(b).

---

[14]Re-arranging V_{DD} equation from Eq. 3.6, we obtain $V_{PK} = V_K + \pi(V_{DD} - V_K)$. Note that V_{PK} is approximately 3x V_{DD} assuming V_K is small.

Firstly, cascode devices are generally required for device robustness in the presence of high peak drain voltages, and they also give additional benefits of input-output isolation and higher output impedance. The cascodes could be thin- or thick-oxide devices depending on design specifications. Secondly, an output 1:n transformer, where n>1, is added to transform the 50Ω antenna impedance into the desired lower load resistance for the PA.
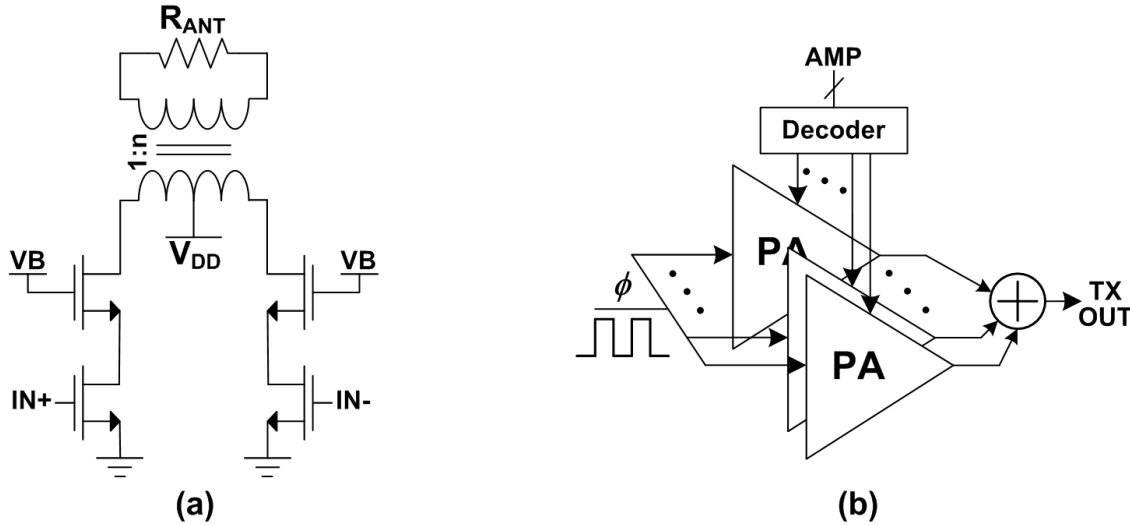
Figure 3.16: Implementation of mixed-signal transmitters: (a) circuit schematic of a practical inverse class-D PA, (b) block diagram of polar transmitter.

Mixed-signal polar transmitters employing inverse class-D PAs have demonstrated good power and efficiency performance at gigahertz frequencies, and a high-level diagram of such a polar transmitter is shown in Fig. 3.16(b) [43]. Current-switching PA cells are segmented and digitally controlled, identical to a current DAC. The amplitude code sets total I_{PK} and thus output power magnitude. The transmitter's output phase is set by the phase of the switching input to the PA.

3.3.2 PA to LNA Transformation

We now describe our proposed transformation of an inverse class-D PA into an LNA.

Fig. 3.17(a) illustrates the typical inverse class-D current-switching PA. Transistors M1 and M2 are the switched PA input devices, and M3 and M4 are cascodes to support high PA output power. This fundamental topology of an input pair plus a cascode pair is identical to a cascoded common-gate LNA, and we exploit this similarity to transform the PA into a wideband LNA.

Fig. 3.17(b) illustrates the same structure in LNA mode. Supply and ground have been flipped, as have the source and drain of all transistors. In PA mode in Fig. 3.17(a), M3 and
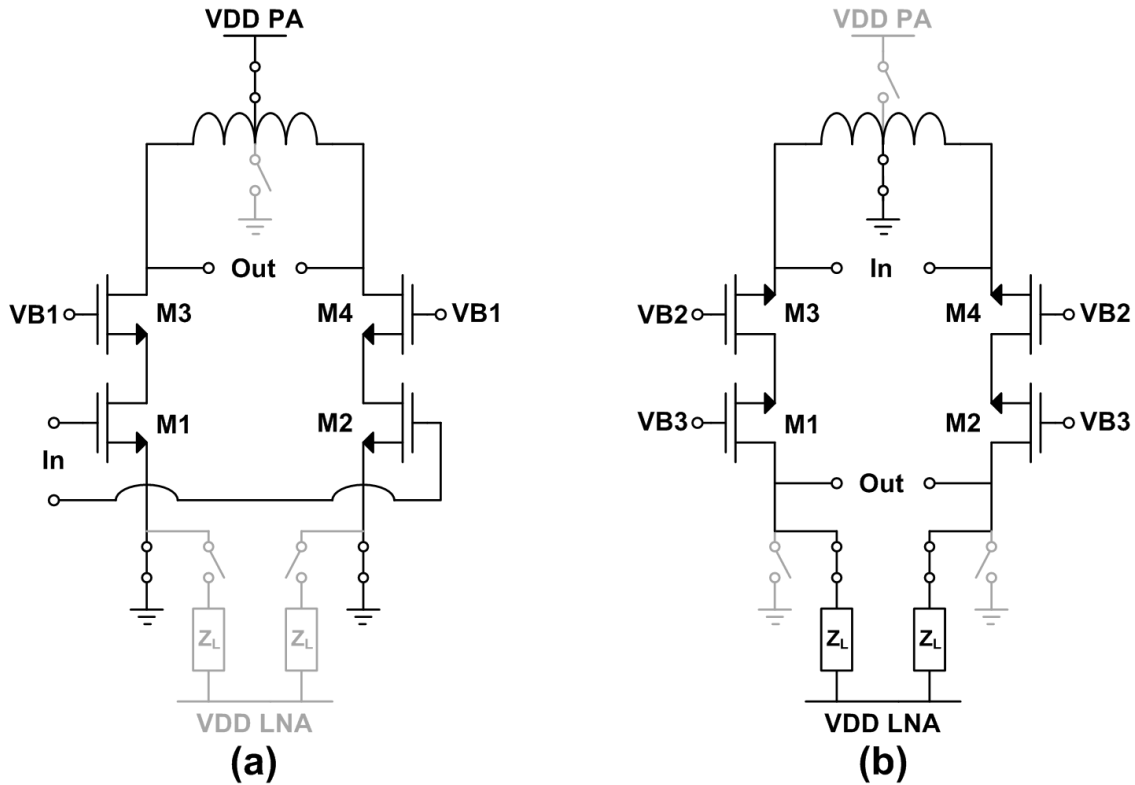
Figure 3.17: PA to LNA transformation: (a) PA mode, (b) LNA mode.

M4 were cascodes, and their drains formed the PA output port. In LNA mode, M3 and M4 are now LNA input devices, and their sources (formerly drains in PA mode) form the LNA input port. Similarly, M1 and M2 are now LNA cascodes, and their drains (formerly sources in PA mode) are the LNA output node and are connected an LNA load and supply.

The PA has thus been transformed into an LNA using only DC power switches. The PA output and LNA input now share a single port, and that port is connected to PA output and LNA input devices directly without any series switches. Isolation is no longer relevant as there is no idle block to isolate during either TX or RX modes. In addition, since the PA is necessarily designed to withstand its own output power, there are no parts of the circuit that require special techniques and processes, such as stacked switches and isolated wells, to tolerate high PA signal swings.

The power switches required for PA/LNA transformation will still have some effect on performance. IR drop from the PA power switch especially can degrade PA output power and efficiency. However, unlike conventional TRSWs whose parasitic capacitances are a severe limitation on insertion loss and isolation at higher frequencies, power switches are DC and can be made large with fewer negative side effects. The design of the power switches benefit from being outside of sensitive RF signal paths.

For these reasons, the proposed PA re-use technique demonstrates great potential for enabling high-frequency integrated T/R switching that is wideband and low loss. The proposed switching technique sidesteps the issues of isolation, power handling, and high-frequency loss that have plagued existing integrated TRSW designs. The next chapter will describe the design and implementation of a wideband TDD front-end utilizing this PA re-use technique.

Chapter 4

Design of a Wideband TDD Front-End with Integrated T/R Switching via PA Re-Use

In the previous chapter, we demonstrated conceptually how a PA can be transformed into an LNA for TDD co-existence purposes. However, while the PA and LNA share a similar overall topology, the two blocks in practice have drastically different design specifications. In this chapter, we describe the design and implementation of a wideband TDD front-end implementing the PA re-use integrated T/R switching technique. We focus mainly on the challenges associated with the design of a shared front-end transformer, the shared PA/LNA core, and the power and mode switches use to enable PA/LNA transformation. We target frequency bands in the 3GHz to 6GHz range, since this is where most TDD LTE bands occur. However, this work is primarily a proof-of-concept demonstrating the feasibility of the PA re-use T/R switching technique in a wideband gigahertz system, and it is not specifically targeted for a certain wireless standard. Measured results from the implemented system are presented at the end of this chapter.

4.1 Front-End Transformer Design

Front-end passive networks are a critical aspect of both PA and LNA design. Conventionally, PA output transformers use 1:2 or higher turns ratio, where the primary is the PA and the secondary is the antenna. This transformation enables the PA to see a load impedance lower than the 50Ω antenna, which, in turn, enables a lower voltage swing at the PA output for a given output power. Low voltage swing is desirable for compatibility with the low supply and breakdown voltages of modern CMOS devices.

In contrast, LNAs conventionally should be input impedance matched to 50Ω or higher. An input transformer that boosts source impedance also provides passive voltage gain at the

input of the LNA, which is desirable for good noise figure. A conventional PA transformer
that reduces antenna impedance and attenuates voltage from antenna to LNA input, on the
other hand, would be devastating for noise performance.

Shared PA/LNA front-end passive networks from existing TDD works are all 3-port systems:
antenna port, PA output, and LNA input. These designs manage the opposing PA/LNA
requirements for antenna impedance by having separate transformers for the PA and LNA,
and combining them in such a way that each transformer can be shorted while the other
is active [35][40]. Our proposed TDD system, however, is a 2-port system, with only the
antenna port plus a shared PA output/LNA input port. It is therefore impossible to short
one port while letting the desired impedance appear on another port, since the PA and LNA
are the same port.

One possible solution is to use a compromise impedance; for example, a 2:3 transformer
providing $\sim 30\Omega$. Naturally, by the definition of compromise, this scheme is problematic
for both the PA and LNA. The PA would need higher supply voltage to achieve desired
peak output power, and maintaining device robustness in the presence of high supply and
high voltage swings would be challenging. The LNA would need to consume more current
to match to the lower input impedance while incurring a large noise penalty from passive
voltage attenuation.

Another possible solution is to reconfigure the transformer to have a variety of transformation
ratios. However, any attempt at reconfigurability - tapping at different spots, disconnecting
turns, even switching to a separate transformer entirely - requires in-line switches in the
signal path. These switches would not only heavily de-Q the transformer, but at the front-
end, they would also experience the full swing of the PA output. They would de facto
become conventional T/R switches and incur all the associated penalties, which defeats the
fundamental purpose of the proposed combination PA/LNA.

For our TDD front-end, we use transformer-based power combining as the method of reconfig-
uring antenna impedance for PA and LNA modes. Compared to the other possible solutions,
this method incurs almost no overhead and degradation for either the PA or the LNA. In
this section, Sec. 4.1.1 describes the theory of transformer-based power combining and how
it can be applied to provide a PA/LNA co-match. Sec. 4.1.2 demonstrates problems with a
conventional power-combining architecture, and Sec. 4.1.3 describes modifications made to
solve those problems. Sec. 4.1.4 and Sec. 4.1.5 describe the silicon implementation of the
designed transformer.

4.1.1   Stacked Transformer for Impedance Co-design

Transformer-based power combining is a PA technique that enables dynamic modulation of
PA load impedances with minimal degradation [44][45]. Fig. 4.1(a) illustrates the concept:
a PA is separated into several sub-PAs, each with its own output transformer, which are

stacked in series.
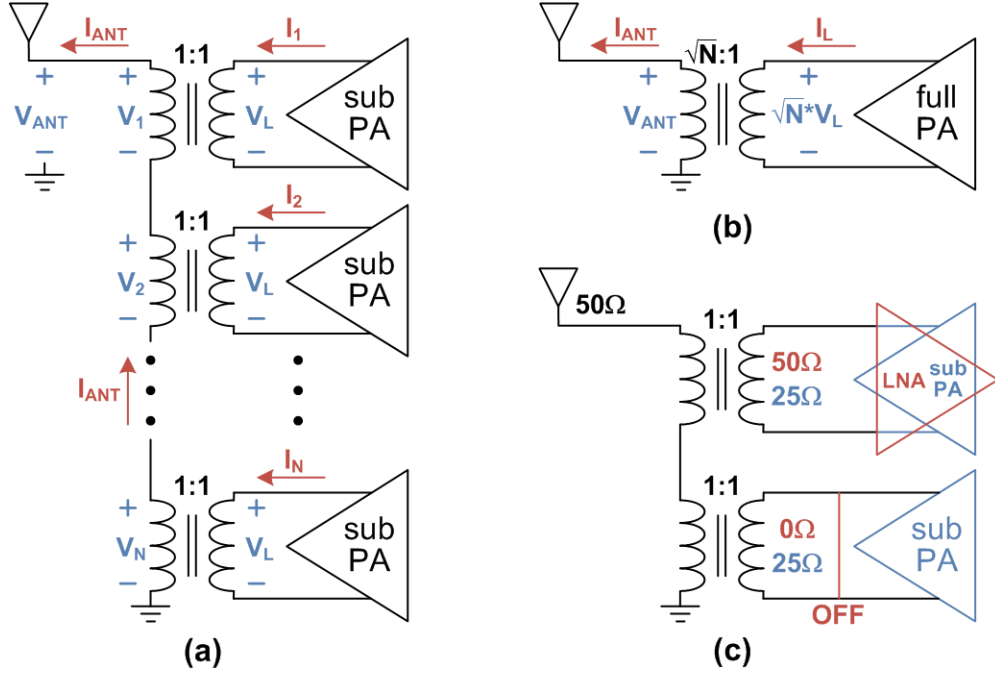
Figure 4.1: Transformer-based power combining: (a) combined PA using N sub-PAs and stacked 1:1 transformers, (b) equivalent PA with single transformer, (c) reconfigurable impedance for PA/LNA.

If all transformers are identical and 1:1, and if all sub-PAs are also identical and in-phase, then the power delivered to the antenna is simply the sum of the output power from each sub-PA.

$$V_{ANT} = V_1 + V_2 + ... + V_N = N * V_L \tag{4.1}$$

$$I_1 = I_2 = ... = I_N = I_{ANT} = \frac{V_{ANT}}{R_{ANT}} = \frac{N * V_L}{R_{ANT}} \tag{4.2}$$

The load impedance seen by each sub-PA is therefore:

$$R_{1,2,...N} = \frac{V_L}{I_{1,2,...N}} = \frac{R_{ANT}}{N} \tag{4.3}$$

And the total power delivered to the antenna is:

$$P_{ANT} = \frac{1}{2} * \frac{V_{ANT}^2}{R_{ANT}} = \frac{1}{2} * \frac{N^2 * V_L^2}{R_{ANT}} \tag{4.4}$$

This effectively creates a 1:N transformation using only 1:1 transformers.[1]

---

[1]An "A:B" transformer or transformation in text refers to the number of physical turns, while a "1:X" label in figures refers to the transformation ratio. Note that $N = \sqrt{B/A}$ is the theoretical ideal.

In contrast to a single-transformer PA with the same load impedance and output power (shown in fig. 4.1(b)), one main advantage of this architecture is its high efficiency when operating in power backoff, by way of dynamic load modulation. PAs are by necessity designed for peak output power, and efficiency drops sharply as output voltage swing becomes a tiny fraction of its peak while supply voltage remains constant [42].[2] Modern modulation schemes such as OFDM have high peak-to-average power ratios (PAPRs), meaning a PA would rarely operate at its peak power. Consequentially, conventional PAs achieve low average efficiency even with good peak efficiency. Power-combining PAs, on the other hand, can boost average efficiency by turning off one or more sub-PAs in power backoff. DC power consumption would be reduced, load impedance increased, and output voltage swing boosted for higher efficiency [45].

We exploit specifically this impedance modulation property of transformer-based power combining for co-matching our PA/LNA structure. Fig. 4.1(c) demonstrates a design with two sub-PAs, one of them repurposed into a convertible PA/LNA. If the bottom sub-PA is turned off and its transformer shorted, the top structure would experience only a 1:1 transformation, and the LNA would see a desired 50Ω source impedance. When both sub-PAs are on, however, the impedance seen by each sub-PA is $50\Omega/2 = 25\Omega$, giving the PA an effective 1:2 transformation and its desired low load impedance. Reconfigurable matching impedance for PA/LNA modes has thus been achieved.

There are also a number of additional advantages to using transformer-based power combining. Firstly, practical on-chip transformers achieve higher Q with lower transformation ratios. Therefore it can be less lossy to realize a 1:N transformation with N 1:1 transformers rather than a single 1:N transformer, especially when N becomes greater than two [46].[3] Secondly, since each sub-PA is responsible for only a fraction of the total output power, its peak output voltage decreases by a factor of $\sqrt{N}$. This allows the use of lower PA supply voltages and even thin-oxide devices, reducing parasitics and boosting high frequency performance. These advantages contribute directly to our design by enabling a higher performance PA as well as a lower loss front-end useful for both PA and LNA modes.

To short out the transformers of turned-off sub-PAs, we appropriate the PA devices themselves as transformer switches [45]. This eliminates the need for an extra switch that would have added parasitics and degraded PA performance, especially given that it would have had to withstand high PA output power. Furthermore, the transformer switch must have very low ON resistance, or else it would degrade the effective Q of the other transformers in the stack and add significant loss. Since PAs are large, the existing PA devices inherently form a huge switch with low resistance, making them an ideal candidate for acting as the transformer switch.

---

[2]This assumes digital switching PAs, whose DC current reduces together with output power. Efficiency in backoff is even worse for traditional class A or class AB PAs [42].

[3]Theoretically, the transformers for power combining don't have to be 1:1, but for this reason, they always are in practical implementations.
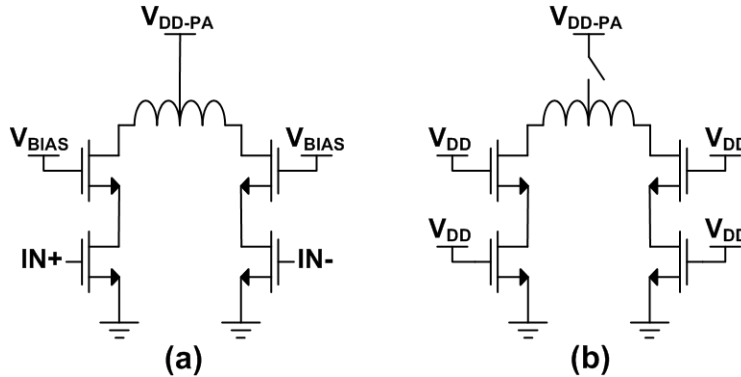
Figure 4.2: Re-using PA as transformer switch: (a) PA in standard configuration, (b) PA re-configured to short transformer.

Fig. 4.2 illustrates how a sub-PA becomes a transformer switch. Fig. 4.2(a) shows a sub-PA under normal operating conditions, with a pair of input devices and a pair of biased cascodes. Fig. 4.2(b) shows the sub-PA when it's shorting the transformer. All 4 devices are hard turned on by having their gates pulled to VDD, and thus both ends of the transformer winding are pulled to ground. The PA supply must be switched off to avoid short-circuit current from supply to ground.

4.1.2 Design Considerations for LNA Noise and Bandwidth

Aside from the impedance co-design issue tackled in Sec. 4.1.1, there are a number of additional challenges in the practical design of a shared PA/LNA front-end. Specifically, transformer non-idealities, as dictated by PA requirements, can significantly restrict LNA noise performance and bandwidth. In fact, for a wideband CG LNA, NF and bandwidth are intertwined and mutually dependent, and in a shared PA/LNA design, both metrics suffer under the same constraints from the PA.

To understand the source of this degradation, we first analyze the effect of transformer parasitic resistance on LNA noise figure. Fig. 4.3(a) shows a simple CG amplifier with a front-end transformer composed of two coupled inductors. The source impedance, representing antenna impedance, is R_S, k is the transformer coupling coefficient, and Z_L is the amplifier load impedance. The capacitor C tunes the front-end tank to the desired resonance frequency.

Fig. 4.3(b) shows the equivalent circuit model with 1:1 source transformation. To keep the analysis simple, we assume infinite output impedance for the transconductor and perfect coupling for the transformer. The parasitic resistance from transformer inductors are modeled as R_P, one for each inductor. If $R_P \ll R_S$, we can ignore R_P on the antenna side since R_P on LNA side will be the major noise contributor, for reasons that will be illuminated below. Fig. 4.3(c) shows the final noise model.
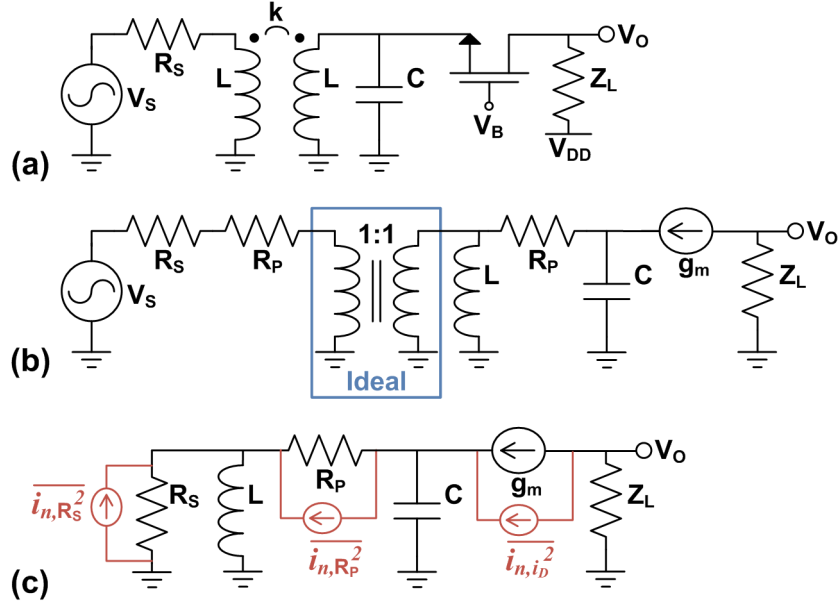
Figure 4.3: LNA noise model: (a) CG LNA with source transformer, (b) equivalent circuit model, (c) approximate noise model.

From Fig. 4.3(c), the noise power at LNA output contributed respectively by R_S, i_D, and R_P are:[4],[5]

$$\overline{v_{on,R_S}^2} = \frac{4kT}{R_S}(g_m Z_L)^2 \left(R_S || sL || \left(R_P + \frac{1}{sC} || \frac{1}{g_m} \right) \right)^2 \left(\frac{\frac{1}{sC} || \frac{1}{g_m}}{\frac{1}{sC} || \frac{1}{g_m} + R_P} \right)^2$$

$$= 4kT R_S (g_m Z_L)^2 * \tag{4.5}$$

$$\left(\frac{sL/R_S}{s^2 LC(1 + R_P/R_S) + (sL/R_S)(1 + g_m(R_S + R_P)) + sCR_P + (1 + g_m R_P)} \right)^2$$

$$\overline{v_{on,i_D}^2} = 4kT\gamma g_m Z_L^2 \left[1 - g_m \left(\frac{1}{sC} || \frac{1}{g_m} || (R_P + sL || R_S) \right) \right]^2$$

$$= 4kT\gamma g_m Z_L^2 * \tag{4.6}$$

$$\left(\frac{s^2 LC(1 + R_P/R_S) + sL/R_S + sCR_P + 1}{s^2 LC(1 + R_P/R_S) + (sL/R_S)(1 + g_m(R_S + R_P)) + sCR_P + (1 + g_m R_P)} \right)^2$$

---

[4]We are looking only at front-end contributors to NF; Z_L noise is ignored.

[5]Note that impedance looking into input of the LNA is approximated as $1/g_m$.

$$\overline{v_{on,R_P}^2} = \frac{4kT}{R_P}(g_m Z_L)^2 *$$

$$\left[\left(R_S||sL||\left(R_P + \frac{1}{sC}\Big|\Big|\frac{1}{g_m}\right)\right)\left(\frac{\frac{1}{sC}||\frac{1}{g_m}}{\frac{1}{sC}||\frac{1}{g_m} + R_P}\right) - \left(\frac{1}{sC}\Big|\Big|\frac{1}{g_m}||(R_P + sL||R_S)\right)\right]^2 \quad (4.7)$$

$$= 4kTR_P(g_m Z_L)^2 *$$

$$\left(\frac{1 + sL/R_S}{s^2 LC(1 + R_P/R_S) + (sL/R_S)(1 + g_m(R_S + R_P)) + sCR_P + (1 + g_m R_P)}\right)^2$$

Combining the above, the noise factor of the LNA is:

$$F = 1 + \frac{\overline{v_{on,i_D}^2}}{\overline{v_{on,R_S}^2}} + \frac{\overline{v_{on,R_P}^2}}{\overline{v_{on,R_S}^2}}$$

$$= 1 + \frac{\gamma}{g_m R_S}\left(\left|\frac{s^2 LC(1 + R_P/R_S) + sL/R_S + sCR_P + 1}{sL/R_S}\right|\right)^2 + \frac{R_P}{R_S}\left(\left|\frac{sL + R_S}{sL}\right|\right)^2 \quad (4.8)$$

We apply the impedance match condition $R_S = 1/g_m$ and take the magnitude of F at its resonance frequency ω_o:

$$F = 1 + \gamma\left(\frac{L/R_S + CR_P}{L/R_S}\right)^2 + \frac{R_P}{R_S}\left(\frac{(\omega_o L)^2 + R_S^2}{(\omega_o L)^2}\right) \quad (4.9)$$

To make this result more intuitive, we substitute into the expression $Q_L = \omega_o L/R_P$, for inductor Q, and $Q_{tank} = R_S/(\omega_o L)$, for the front-end tank:[6]

$$F = 1 + \gamma\left(1 + \frac{Q_{tank}}{Q_L}\right)^2 + \frac{R_P}{R_S}(1 + Q_{tank}^2) \quad (4.10)$$

The LNA suffers a heavy noise penalty because the Q_{tank}/Q_L ratio in practice can be quite large, and because the square terms cause NF to rise sharply even when $R_P \ll R_S$. In contrast, the NF contribution from the antenna-side R_P we ignored earlier is only R_P/R_S, without a square multiplier, and thus negligible if R_P is very small. The LNA-side R_P, on the other hand, can contribute significant noise depending on Q_{tank} even with small R_P.

Fig. 4.4 plots LNA NF as a function of Q_{tank}, assuming $Q_L = 10$ and $\gamma = 1$. The plot shows noise contributions due to the i_D and R_P terms separately, as well as their combined effect. While i_D noise is strictly better for lower Q_{tank}, R_P noise has a minimum at $Q_{tank} = 1$. This creates a minimum in the overall NF response; however, even that minimum point presents a 0.7dB increase over the baseline 3dB NF.

---

[6]Actual front-end tank Q is half of the given expression due to impedance match, but Q_{tank} as defined gives a more intuitive NF expression.
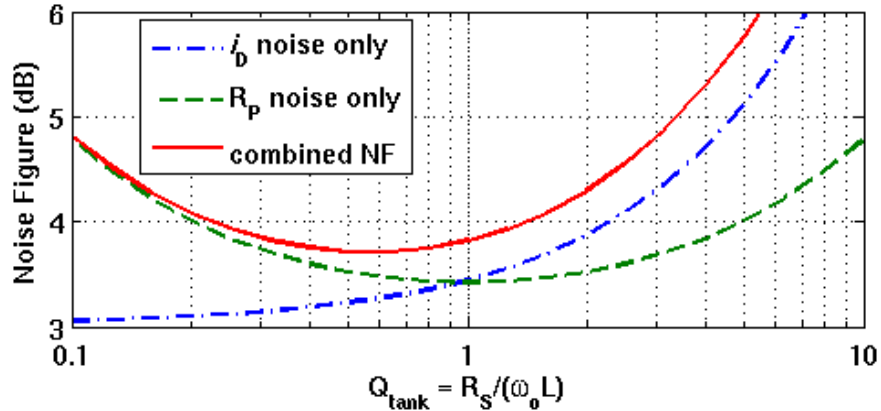
Figure 4.4: LNA NF as a function of Q_{tank}.

We can make a rough estimate of $C = 1.5\text{pF}$ for each sub-PA,[7] which, for a midband resonance frequency at 4.5GHz, limits L to about 830pH. This results in $Q_{tank} = 2.1$, where NF according to Fig. 4.4 is 4.4dB. This is fairly high for a theoretical ideal calculation that does not yet include noise from layout parasitics or even Z_L. The PA's large output capacitance, through limitations on L and Q_{tank}, has thus constrained LNA NF.

Furthermore, in a wideband LNA, we are concerned with performance across the entire band and not just at resonance. In Fig. 4.5, we plot the NF and normalized gain of the LNA across frequency, using the same estimated C and L values as above. Both gain and NF are due to only the front-end tank; the LNA transconductor and load are assumed to be ideal and not frequency-selective. The effect of inductor parasitic resistance R_P is still included.
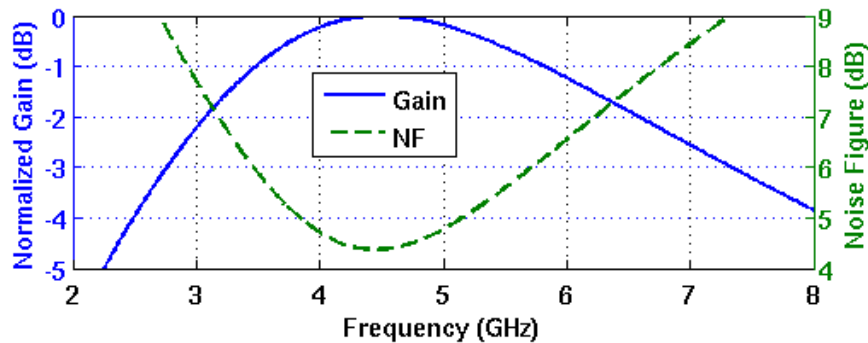
Figure 4.5: Frequency response and NF of LNA due to front-end tank.

The -3dB bandwidth of the gain response is very wide, and it tracks with our estimated $Q_{tank} = 2.1$ and bandwidth $= (2 * 4.5\text{GHz})/(Q_{tank}) = 4.3\text{GHz}$.[8] However, for every 1dB

---

[7]Estimate includes PA drain capacitance, routing parasitics, and antenna pad capacitance. Note that antenna capacitance is doubled due to impedance transformation.

[8]Note that Q_{tank} is defined to be $R_S/(\omega_o L)$. The factor of 2 is due to matching.

drop in gain, NF rises more than 1dB. At the edges of -3dB bandwidth, NF has risen more than 4dB from its lowest value at resonance. This, again, is unacceptably high. NF should be below a certain threshold across the entire band. Consequently, the -3dB bandwidth must actually be much greater than the desired frequency range. However, like NF at resonance, bandwidth is also constrained by PA output capacitance by way of L and Q_{tank}.

Frequency tuning - a simple and common method for extending bandwidth - could not be used due to high noise from the requisite front-end capacitor bank.[9] It is instructive, however, to briefly analyze the noise contributions from such a capacitor bank, despite it not being a part of our final design. The cause of its high noise penalty is similar to that of the transformer parasitic resistance R_P, and this noise mechanism will show up again in other parts of our system.
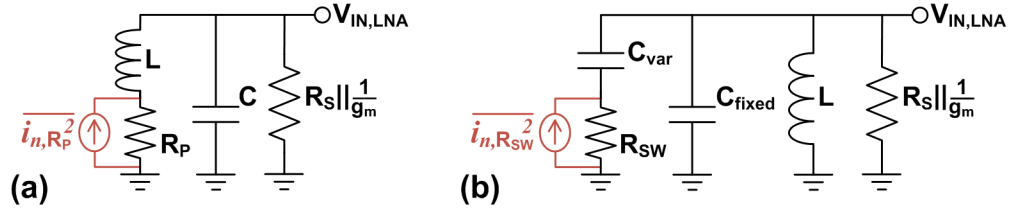
Figure 4.6: Front-end noise models for (a) inductor parasitic resistance and (b) capacitor switch resistance.

First, we revisit transformer noise with a different approach, less precise but perhaps more intuitive. Fig. 4.6(a) illustrates an alternate approximation for the transformer noise model. If we make a series-to-parallel transformation for R_P, the equivalent parallel resistance would be:

$$R_{P,parallel} = R_P * (Q_L^2 + 1) \approx R_P * Q_L^2 \tag{4.11}$$

We could also think of this as the inductor and capacitor forming an L-match between R_P and the LNA input node. Assuming $R_{P,parallel}$ is much larger than $R_S||(1/g_m)$, its contribution to noise factor would be:

$$\frac{\overline{v_{on,R_P}^2}}{\overline{v_{on,R_S}^2}} \approx \frac{R_S}{R_P Q_L^2} = \frac{R_P}{R_S} Q_{tank}^2 \tag{4.12}$$

This is approximately the result from Eq. 4.10.

Fig. 4.6(b) models the same front-end, but with a variable capacitor instead of transformer parasitics. Tank capacitance is split between C_{fixed} and a tunable portion C_{var}, with R_{SW}

---

[9]Device breakdown in the presence of high PA output power is also a potential challenge for a front-end capacitor bank, but because it was already infeasible due to noise, other implementation issues were not investigated.

representing the ON resistance of capacitor bank switches. The same series-to-parallel trans-
formation (or L-match) that applied to R_P earlier now applies to R_{SW}. Its noise contribution
accrues the same Q_{tank}^2 factor, albeit with the mitigating condition that only a fraction of
the total capacitance is being switched:

$$\frac{\overline{v_{on,R_{SW}}^2}}{\overline{v_{on,R_S}^2}} \approx \frac{R_S}{R_{SW}Q_{C_{var}}^2} = \frac{R_{SW}}{R_S}Q_{tank}^2 \left(\frac{C_{var}}{C_{var} + C_{fixed}}\right)^2 \tag{4.13}$$

As before, we estimate $C_{fixed} = 1.5\text{pF}$ but now center resonance at 5GHz, making $L =$
680pH. To shift resonance to 3.5GHz, $C_{var} = 1.5\text{pF}$ is needed. With these values, we would
need $R_{SW} < 2.5\Omega$ to keep added NF below even 0.3dB, resulting in unrealistically large
switches.[10] Considering that the intended purpose of frequency tuning was to mitigate NF
variations across the band, including a capacitor bank with such a severe noise penalty would
be counter-productive.

4.1.3   Stacked Transformer with Reconfigurable Inductance

The problems with LNA NF and bandwidth are fundamentally due to the transformer in-
ductance, as calculated in Sec. 4.1.2, being too small. While increasing inductor Q and
lowering its parasitic resistance would decrease noise, these parameters are limited by the
physical realities of on-chip inductors. Front-end tank Q is therefore left as the only mal-
leable variable from the LNA noise factor expression in Eq. 4.10. A large Q_{tank} not only
narrows bandwidth to create sharper NF increases at band edges, it also amplifies the noise
from inductor parasitic resistance to create higher absolute NF across the entire band. To
reduce Q_{tank}, transformer inductance must be increased.

Sec. 4.1.2 claimed that transformer inductance is limited by the PA's large output capaci-
tance, but this is not entirely true. The limitation applies *only to the PA*; in LNA mode,
front-end capacitance is smaller. Transformer inductance in LNA mode, then, is not con-
strained by PA output capacitance as much as by the lack of reconfigurability inherent to
on-chip inductors.

Let us take another look at our proposed stacked transformer architecture from Fig. 4.1(c),
reproduced here in Fig. 4.7(a) for convenience. In LNA mode, the bottom transformer should
be shorted, but doing so is wasteful. The LNA would be leaving unused an entire transformer
that already exists in the system. In practice, the bottom transformer is not only wasted,
it also degrades the Q of the top transformer through parasitic resistances from both the
windings and the shorting switches.

Ideally, the LNA should re-use the bottom transformer in such a way that it contributes
to performance rather than degrade it. Since the two transformers are already stacked in

---

[10]The situation is actually quite a bit worse because the capacitance being switched is doubled for a
differential capacitor, so R_{SW} would need to be halved compared to single-ended case.
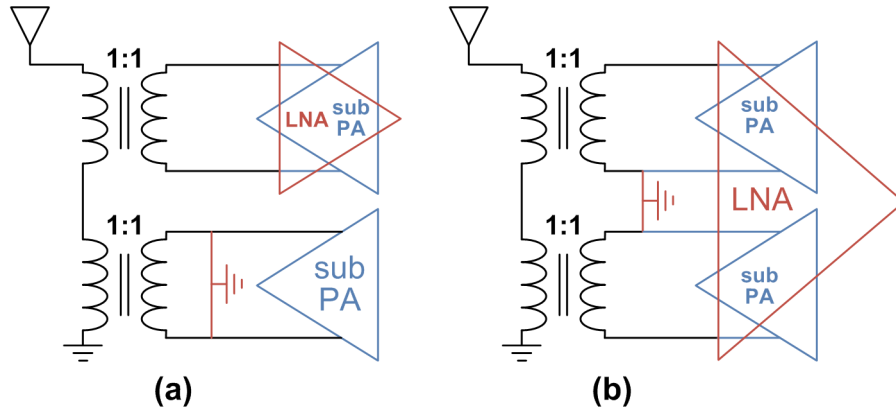
Figure 4.7: Stacked transformer architecture: (a) conventional, (b) transformer-reuse.

series, a natural strategy is to simply treat them as a single, larger 1:1 transformer, as shown in Fig. 4.7(b). This configuration re-uses the bottom transformer while maintaining impedance match for the LNA. More importantly, with the two transformers in series, the resultant combined transformer has doubled inductance. In effect, inductance has been "reconfigured" for LNA mode with no architectural changes or additional overhead to the front-end transformer structure.

The revised stacked transformer configuration of Fig. 4.7(b) has doubled inductance and halved Q_{tank} as compared to the original structure in Fig. 4.7(a). Using the same estimated circuit parameters from Sec. 4.1.2, the new Q_{tank} becomes 1.1, which results in $NF = 3.8$dB according to Fig. 4.4. This noise performance is not ideal, but nevertheless a significant improvement. Performance is now very close to the minimum achievable NF in Fig. 4.4 and acceptable for our target application. Fig. 4.8 plots NF and gain frequency response for the revised transformer structure. In-band gain variation is now less than 1dB, and NF increases by about 1dB at band edge from its resonance NF. In contrast to Fig. 4.5, the LNA's bandwidth and noise performance is now reasonable.
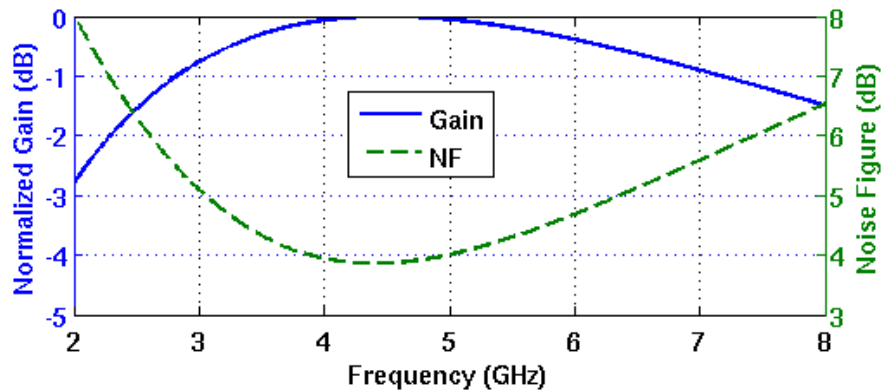

Figure 4.8: Frequency response and NF of LNA with transformer-reuse.

The mechanism that allows the two transformers to be combined into a single transformer relies on the LNA's center tap connection to ground, shown in Fig. 4.7(b). To create that ground connection, we simply re-use the PA devices as ground switches, identical to the method described in Sec. 4.1.1 and Fig. 4.2. However, while those two center nodes can both easily be shorted to ground, they cannot be easily shorted to each other, at least not without additional switches and thus overhead and degradations. In other words, the center tap can be exploited to combine exactly two transformers, and this configuration unfortunately cannot be extended to use more stacked transformers to achieve even higher inductance. Nonetheless, the performance achieved with only two transformers is adequate for our application. Furthermore, this improvement was gained with no additional overhead to the previous stacked transformer architecture and with no modifications with respect to PA mode.

4.1.4   Implementation of 1:1 Transformer

To implement our front-end transformer design on-chip, we first investigate optimal layout for a 1:1 transformer. The maximum theoretical gain of an impedance matched 1:1 transformer is [47]:

$$G_{MAX} = 1 + \frac{2}{Q^2 k^2} - 2\sqrt{\frac{1}{Q^4 k^4} + \frac{1}{Q^2 k^2}} \tag{4.14}$$

To maximize gain and achieve low insertion loss, it is desirable to maximize both inductor Q and transformer coupling factor k. In an on-chip implementation, these parameters are limited by process design rules governing metal width, gap, and density. In addition, Q and k are also limited by parasitic capacitance. For example, using wider metal traces will increase Q but also increase capacitance to substrate, and the close magnetic coupling necessary for high k is accompanied by high capacitive coupling between the windings as well. Both capacitance to substrate and interwinding capacitance can limit self-resonance frequency (SRF) and cause higher insertion loss.

Fig. 4.9 shows several 1:1 transformer implementations that we experimented with. Fig. 4.9(a) illustrates the most basic version, which is simply two single-turn inductors implemented on a thick metal layer. The outer blue loop is the primary winding, on the antenna side, and the inner red loop is the secondary, on the PA/LNA side. A center-tap port for the secondary winding is included. Single-turn inductors were chosen here, despite their larger area, for their higher Q compared to multi-turn structures. The octagon shape also contributes higher Q since it more closely approximates a circle - the shape with maximum Q per unit inductance - compared to squares or rectangles [47].

Since process design rules limit maximum metal width, one method to increase Q beyond that limit is to simply put two loops in parallel. If the loops of the two inductors are interlaced as shown in Fig. 4.9(b), there are additional benefits beyond just an increase in
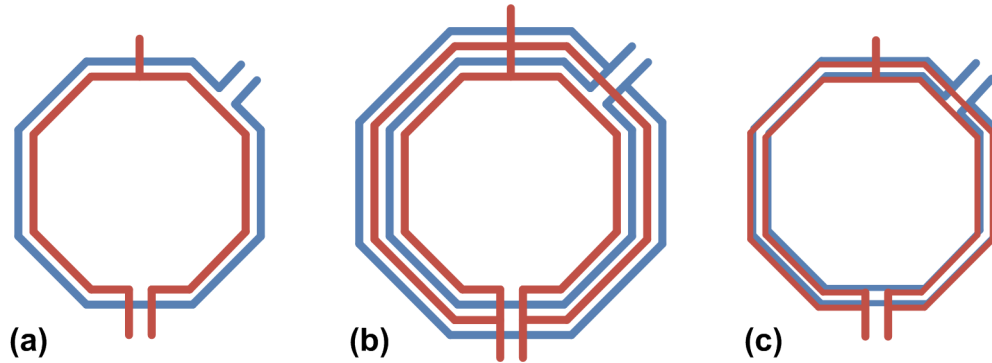
Figure 4.9: Transformer layout implementations: (a) single loop, (b) parallel loops, (c) broadside coupling.

effective width. Firstly, interlacing creates more coupling between the windings, improving k. Secondly, at high frequencies, current in a transformer does not flow uniformly around the surface of the conductor. Current will instead crowd along the edge nearest to the coupled winding, whose current is traveling in the opposite direction [48]. In the interlaced structure of Fig. 4.9(b), the middle two loops have coupled windings on both sides, which helps even out current density and improve high-frequency Q.

Another option to further increase Q is to strap multiple layers of metal together, so that they act as a thicker metal with lower resistance. In our process, two thick metal layers are available, plus an ultra thick aluminum layer with similar unit resistance. A secondary effect of strapping multiple layers is increased interwinding coupling, since there is more surface area between the windings. This is beneficial for k, but also increases capacitive coupling.

Fig. 4.9(c) shows an alternate implementation using broadside coupling - the primary is on one metal layer while the secondary is directly underneath on a lower metal layer, or vice versa. This layout ensures high k due to its small gap size and high coupling surface area. The problem of current crowding at high frequencies is also alleviated. However, coupling capacitance is high, and it is not possible to strap multiple layers together for higher Q with this structure. Table 4.1 summarizes the performance of the various 1:1 transformer options. All transformers are implemented as octagons with 150µm inner radius, 10µm width, and 2µm gap, where dimensions are constrained by design rules for metal width, gap, and local density. The transformer models are analyzed using Integrand EMX software and simulated under ideal matched conditions. Transformer parameters are taken at midband frequency of 4.5GHz.[11]

The performance gains of moving from single loop to double loop, and from single layer metal to two layers, are both significant. Moving to three layers of strapped metal, however,

---

[11]The L and Q shown in the table, as well as mentioned in text hereafter, are averages of their respective values from the primary and secondary windings.

Table 4.1: Comparison of transformer layout implementations.

Implementation	L(nH)	Q	k	S21(dB)
Single loop, single layer metal	0.87	7.9	0.66	-1.7
Parallel loops, single layer metal	0.72	10.4	0.80	-1.1
Parallel loops, 2 layers strapped metal	0.70	14.3	0.83	-0.8
Parallel loops, 3 layers strapped metal	0.68	15.8	0.85	-0.7
Parallel loops, broadside coupling	1.0	9.6	0.96	-0.9

only gives a slight performance boost for the price of higher interwinding capacitance. The broadside coupling structure, despite achieving reasonable $IL < 1$dB, has extremely low SRF due to its tight coupling. Thus, for our design, we choose the parallel loop structure of Fig. 4.9(b) with two thick metal layers strapped together. The simulated parameters of our final 1:1 transformer are: $L = 700$pH, $Q = 14.3$, $k = 0.83$, and $IL = 0.8$dB.

4.1.5   Implementation of Stacked Transformer

Fig. 4.10 illustrates the complete stacked transformer structure, which uses two mirrored copies of the 1:1 transformer design. The gap distance between the outer edges of the two 1:1 transformers is 20µm. In LNA mode, the stacked transformer, simulated under matched conditions at 4.5GHz, has parameters: $L = 1.8$nH, $Q = 9.6$, $k = 0.81$, and $IL = 1.0$dB.

Figure 4.10: Layout implementation of stacked transformer.
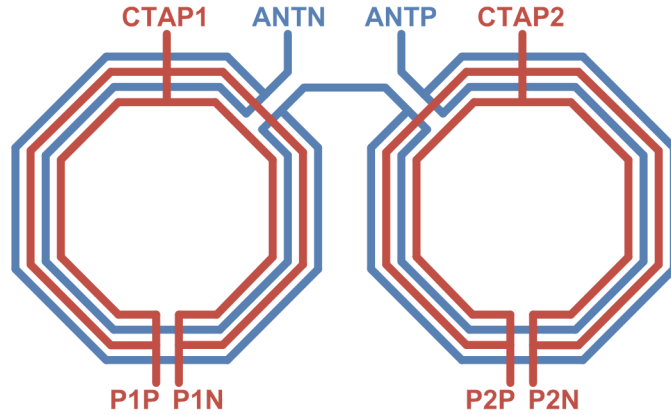
The stacked transformer is de-Q-ed by two main sources of loss, each contributing 0.1dB of additional IL. The first is the center bridge on the primary connecting the two 1:1 transformers. The second is routing from primary ports to the chip's antenna pads, which is not shown in Fig. 4.10 but is included in our simulations. The inductance of the stacked transformer

is somewhat higher than twice the inductance of the single 1:1 transformer. This effect is
attributed to a combination of additional routing and lowered SRF.

In the above simulations, the transformer is unbiased and floating. When the transformer
is biased into balun mode - primary winding has one port grounded and is single-ended,
secondary winding is center-tapped to ground - its frequency response is changed as both
substrate capacitance and interwinding capacitance are subject to Miller effect. The capaci-
tances can appear larger or smaller depending on if they appear between signals of equal or
different strengths, and of equal or opposite polarity. Consequently, IL of stacked transformer
in balun mode increases to 1.3dB.

Furthermore, due to the Miller capacitance effect, there also arises a significant performance
difference between applying the balun in an inverting versus non-inverting configuration.
Fig. 4.11 illustrates the two configurations. The two structures are functionally identical; in
both, two 1:1 transformers are connected in series. However, when the effects of interwinding
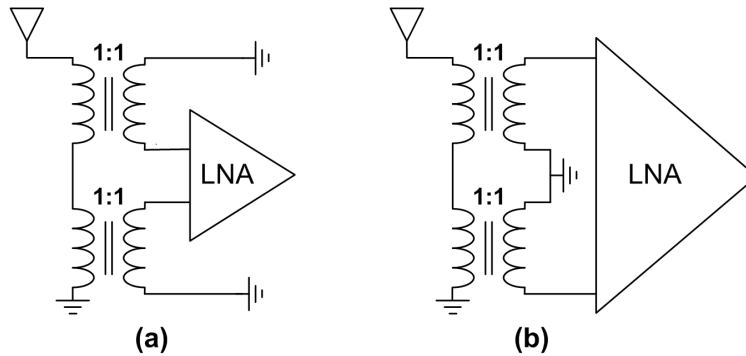capacitances are included, their frequency responses differ.

Figure 4.11: Stacked transformer in LNA mode with (a) inverted coupling, (b) non-inverted
coupling.

To formally analyze this difference, we apply the SRF model in Fig. 4.12(a). For simplicity,
only interwinding capacitance is included, and k is set to 1. Each transformer port is assumed
to couple symmetrically to the inverting and non-inverting ports of the opposite winding.

Fig. 4.12(b) shows the SRF model for a stacked transformer with inverted coupling. Balun-
mode biasing forces $v_p = v_1$ and $v_n = v_1 - v_2$. Solving for impedance seen from the single-
ended antenna port results in the following frequency response:[12]

$$Z_{INV} = \frac{2sL}{1 + 10s^2 LC_C} \tag{4.15}$$

This is the impedance of an inductor with inductance $2L$ and SRF $\omega_{SRF} = \sqrt{1/(10LC_C)}$.

---

[12]The blue box denotes an ideal transformer. Interwinding capacitances are not drawn but are included
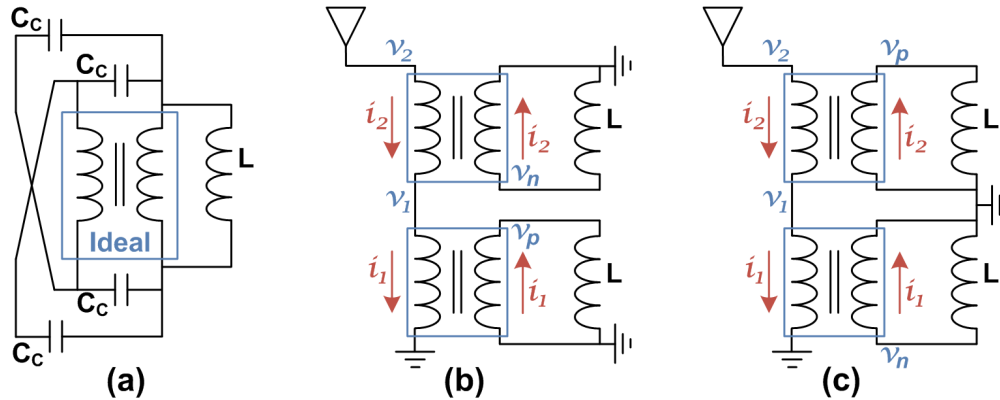in the analyses as shown in Fig. 4.12(a).

Figure 4.12: SRF models for (a) single 1:1 transformer, (b) inverting stacked transformer,
(c) non-inverting stacked transformer.

Fig. 4.12(c) shows the SRF model for non-inverted coupling. In this configuration, $v_p = v_2 - v_1$ and $v_n = -v_1$. Solving for impedance from antenna port results in:

$$Z_{NON-INV} = \frac{2sL(1 + 6s^2 LC_C)}{(1 + 10s^2 LC_C)(1 + 2s^2 LC_C)} \tag{4.16}$$

The inductance and SRF match those of the inverting configuration. However, there is now also a zero at $\omega_Z = \sqrt{1/(6LC_C)}$, which is 1.3x SRF and forms a doublet with it. In balun mode, SRF is sufficiently low frequency that the doublet does affect in-band frequency response of the LNA input tank.

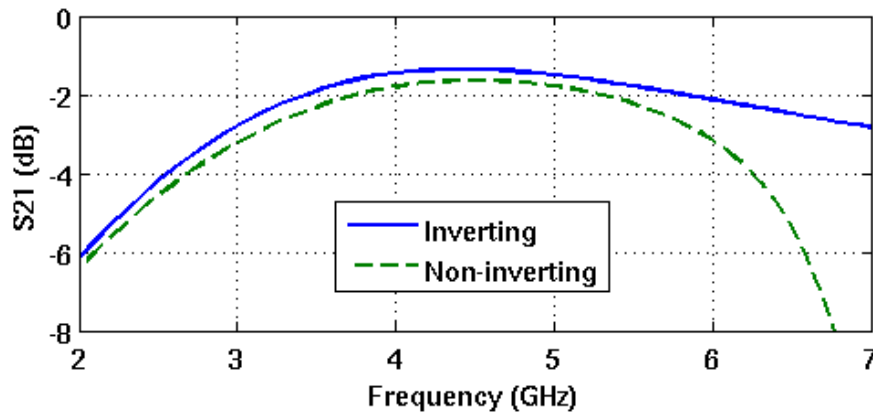Figure 4.13: S21 of stacked transformer in LNA mode.

Fig. 4.13 plots simulated S21 of the stacked transformer in LNA balun mode. Response with inverted coupling is shown in blue, and a peak IL of 1.3dB is achieved, as reported earlier. Response with non-inverted coupling is shown in green, and the degradation from the doublet at the upper edges of the band is apparent. Furthermore, IL across the entire

band, including peak IL at midband, is degraded as well. Thus, in LNA mode, we operate
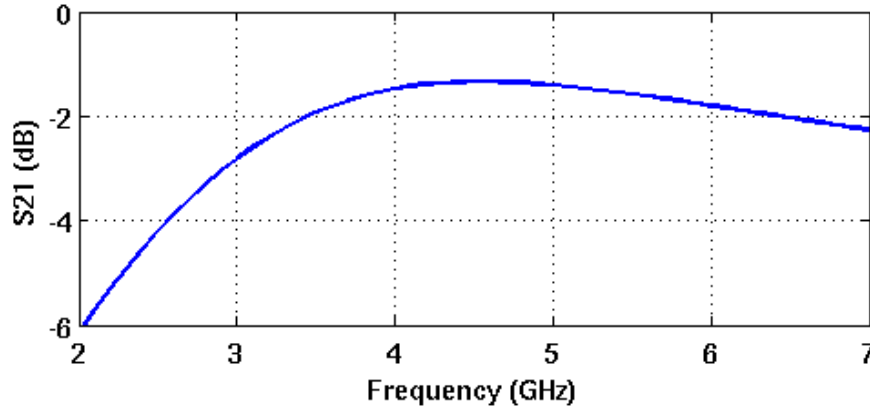the stacked transformer in the inverting configuration of Fig. 4.11(a).

Figure 4.14: S21 of stacked transformer in PA mode.

In PA mode, the two single 1:1 transformers are separately driven, and thus inverted versus
non-inverted coupling does not matter. IL of the stacked transformer in PA mode matches
that of LNA mode: $IL = 1.0$dB when the transformer is unbiased, and $IL = 1.3$dB when the
primary winding is made single-ended and the secondary windings center-tapped. Fig. 4.14
plots simulated S21 of the stacked transformer in PA mode.

4.2 Design of PA/LNA Core

Sec. 3.3.2 described conceptually how an inverse class-D digital PA could be re-purposed
into a common gate LNA; this section delves into the implementation details of designing a
shared PA/LNA. Sec. 4.2.1 and Sec. 4.2.2 describe the standalone PA and LNA architectures,
respectively, as well as how they are combined. Sec. 4.2.3 describes the design of the PA
supply which, which was the most complex switch in enabling transformation between PA
and LNA operations. Sec. 4.2.4 describes the remaining PA/LNA mode control switches in
the system.

4.2.1 PA Core

In an inverse class-D digital PA, each PA cell consists of a digitally-switched differential pair,
similar to a differential current DAC. The input pair, when the cell is ON, are driven by a
differential LO at the desired PA output frequency. The drain outputs of all cells connect to
the front-end transformer, which provides the desired load impedance and tank frequency
response.

Fig. 4.15 shows the PA cell implementation. Drivers locally regenerate the inputs to each
cell, and the supply bypass capacitors for the drivers are laid out locally within each cell
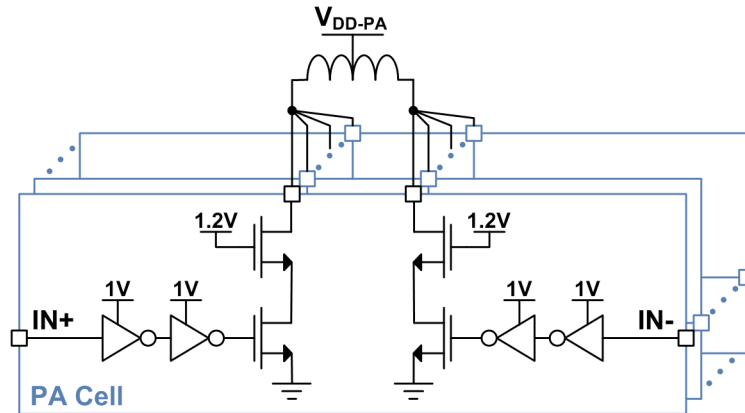
Figure 4.15: Schematic of a PA cell.

as well. Cascode devices provide robustness to withstand high PA output power, as well
as additional benefits of input-output isolation and high output impedance. Both the input
and cascode transistors are implemented with 1.2V thin-oxide devices.

We target peak PA output power of 23dBm. Estimating 2dB loss through the transformer
and routing, and with power split between two sub-PAs, each sub-PA must have peak output
power of 22dBm. Assuming a load resistance of 25Ω, the voltage at the drain of each PA
cell theoretically swings from 0V to 2.8V. However, given non-idealities such as finite ON
resistance of the PA devices, harmonic distortion, and asymmetric power combining,[13] the
drain voltage in practice could swing as high as 3.2V. A thin-oxide transistor can tolerate
gate-to-source or gate-to-drain voltage of $1V_{DD}$ DC and up to $2V_{DD}$s AC. If the cascode gate
is biased at 1.2V, a thin-oxide device is still adequate for robustness given our target output
power, plus some margin.

The ideal drain efficiency of an inverse class-D PA is $\eta = 100\%$, as its drain current is
orthogonal to drain voltage, shown in Fig. 4.16(a). However, achieving the theoretical ideal
requires perfect control of the output tank impedance. The ideal current waveform - a square
wave - contains only odd harmonics, and the voltage waveform - a half-wave rectified sinusoid
- contains only even harmonics. Therefore, the ideal front-end must not only present the
desired load resistance at the fundamental frequency, but also present an open circuit at all
even harmonics and a short at all odd harmonics [42].

In a practical implementation, only the fundamental can be well controlled, which reduces
theoretical achievable efficiency to 78.5% [42]. Efficiency is further eroded by two main
components. The first is transformer loss, which reduces output signal power while DC power
consumption remains constant. The second is finite ON resistance of the PA devices, which

---

[13]The top and bottom sub-PAs see different load impedances across frequency due to the inherent asym-
metry of the transformer being used as a balun. Consequently, their drain voltage swings are slightly different
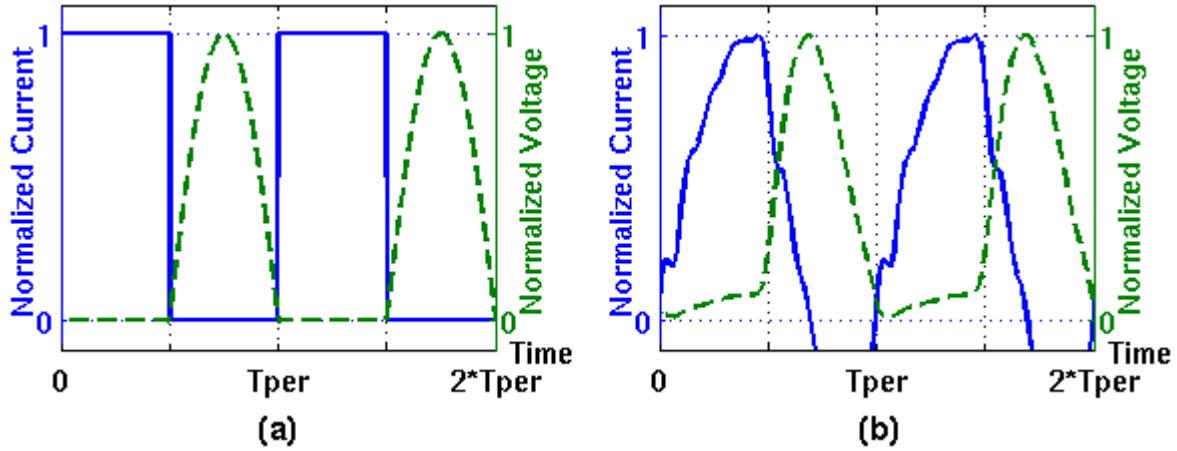and not symmetrically split.

Figure 4.16: PA drain current and voltage waveforms: (a) theoretical ideal, (b) actual implementation.

consumes a portion of power that would otherwise have been transferred to the antenna.

Our implemented transformer has $IL = 1.3$dB. The ON resistance of the combined input-and-cascode PA devices, under realistic current and voltage bias conditions, is about 1.5Ω per sub-PA. These two non-idealities each reduce efficiency by approximately $15\% \sim 16\%$. Parasitics from top level routing, pads, and bondwire inductance cause an additional 4% degradation. The full pure PA implementation (without LNA modifications) achieves 24.3dBm peak output power at 42.7% drain efficiency. Fig. 4.16(b) shows the drain current and voltage waveforms of the implemented PA.
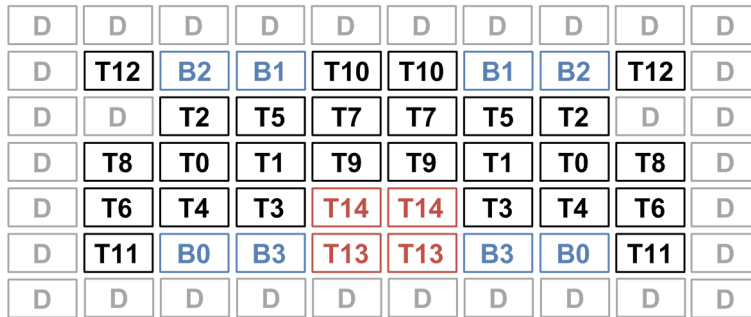
Figure 4.17: Layout diagram of full PA core.

Fig. 4.17 shows the full PA core layout containing both sub-PAs; each block represents a PA cell. Each sub-PA consists of 15 thermometer cells (T1-T15) and 4 binary cells (B1-B4). A ring of dummy cells surround the entire PA core structure to ensure optimal layout matching amongst cells, as every active cell would be abutted by another cell in all directions. The lower, more critical thermometer cells are placed in the center, while higher cells are placed

more outwardly in a pattern that would average out and minimize deterministic variation.
The two sub-PAs are mirrors of each other.

The LNA will use the highest two thermometer cells, T13 and T14, of each sub-PA. These
four cells are therefore placed in the center adjacent to each other, allowing the LNA to
straddle the two sub-PAs with good layout matching and minimal extraneous routing. In
PA mode, T13 and T14 is unused for the vast majority of the time, and are only activated
when the PA needs to operate at or near peak output power. Therefore, modifying these
specific cells to accommodate the LNA will have the least impact on average PA performance.

4.2.2   LNA Design

Fig. 4.18 illustrates the LNA architecture. The LNA, because it re-uses PA core devices, is
by necessity a cascoded common gate topology. The LNA input devices are the PA cascode
devices of Fig. 4.15. In PA mode, the PA cascode drains are output nodes connected to
the front-end transformer. In LNA mode, the same nodes are the sources of the LNA
input devices - LNA input nodes - and are DC-coupled to ground through the front-end
transformer.

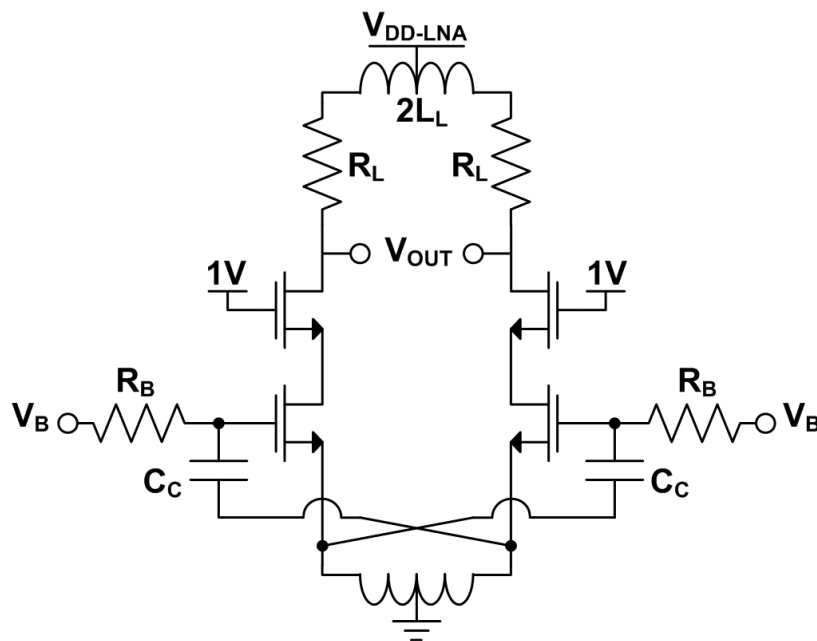Figure 4.18: Schematic of LNA architecture.

The LNA cascode devices are PA core input devices, which, as shown in Fig. 4.15, are
digitally switched in PA mode to either 0V or 1V. Since 1V is a reasonable gate bias voltage
for an LNA cascode, the bias can be generated by manipulating the digital input to the
PA, and no additional circuitry is needed inside the PA core. It was especially important to

avoid overhead in biasing the LNA cascode because the gates of the PA input devices are sensitive, high-speed nodes where any additional parasitics is undesirable.

Capacitive cross-coupling at the LNA input is added to boost g_m and help noise and power performance. The effective transconductance G_M of a capacitively cross-coupled CG LNA is given in Eq. 4.17 below. The intrinsic gate-to-source capacitance of the LNA input device is C_{GS}, and any parasitic capacitance to ground at the input device gate node is represented by C_{G0}.[14]

$$G_M = g_m * \left(\frac{2C_C + C_{G0}}{C_C + C_{GS} + C_{G0}} \right) \tag{4.17}$$

The base noise factor of a capacitively cross-coupled CG LNA, when only input device noise is included and under input match condition $G_M R_S = 1$, is:

$$F = 1 + \frac{\gamma}{G_M R_S} \left(\frac{g_m}{G_M} \right) = 1 + \gamma \left(\frac{g_m}{G_M} \right) \tag{4.18}$$

Conventionally, $C_C \gg C_{GS}$ and $C_C \gg C_{G0}$ is chosen to get the maximum amount of g_m-boost possible. However, this LNA design is constrained by the amount of extra capacitance (in the form of C_C) the PA can absorb into its output tank. The C_{GS} of the LNA is about 100fF, $C_{G0} \approx 200$fF due to overhead from PA/LNA mode switching, and $C_C = 1$pF is the maximum tolerable capacitance by the PA. Fig. 4.19 plots NF and normalized bias current as a function of C_C for a constant G_M. Despite $C_C = 1$pF being not overwhelmingly large, it nevertheless attains the vast majority of the noise and power benefit from g_m-boosting, with 41% reduction in I_D and 1.0dB reduction in base NF.

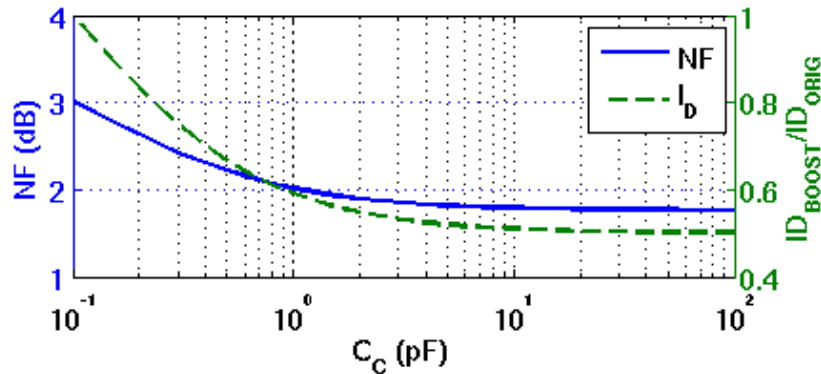Figure 4.19: NF and normalized bias current of capacitively cross-coupled CG LNA as a function of C_C, for $C_{GS} = 100$fF and $C_{G0} = 200$fF.

The output RC corner of a standard resistively-loaded LNA is too low for our target frequency range. Thus, we employ a shunt-peaking load for the LNA, as shown in Fig. 4.18. The shunt-

---

[14]The conditions $C_C > 0$ and $\omega \gg 1/(C_C R_B)$ are assumed.

peaking inductor is placed above the resistor and implemented as a single center-tapped inductor for better layout matching and area efficiency.
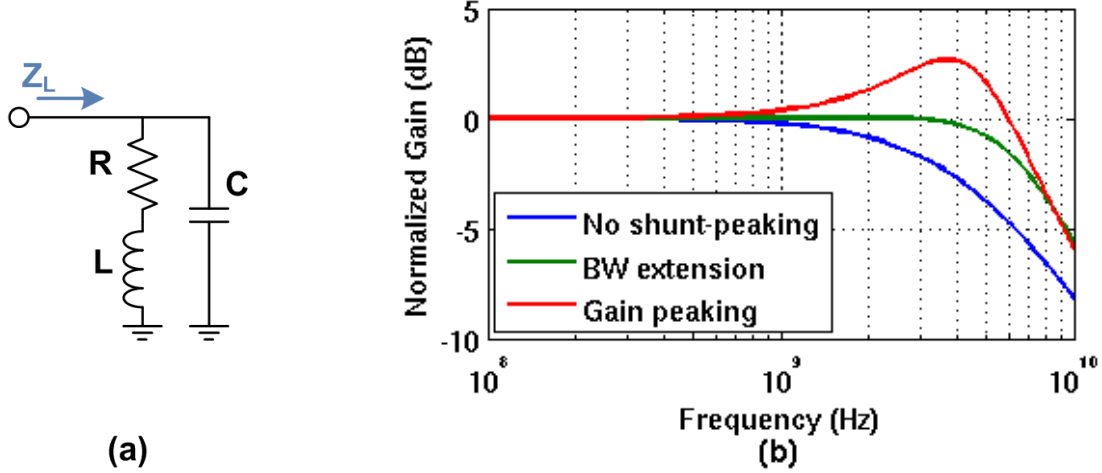

Figure 4.20: Shunt-peaking load: (a) schematic, (b) frequency response.

The impedance of a shunt-peaking load, shown in Fig. 4.20(a), is:

$$Z_L = (sL + R) \, || \left(\frac{1}{sC} \right) = \frac{sL + R}{s^2 CL + sCR + 1} \tag{4.19}$$

A shunt-peaking load adds a zero to the transfer function, which can offset RC rolloff and extend bandwidth. The blue and green frequency responses in Fig. 4.20(b) illustrate this effect. Depending on how the zero is placed, the gain around the rolloff frequency could also peak above the amplifier's low frequency gain, shown in red in Fig. 4.20(b). Since low-frequency gain is unnecessary for our application, we purposely peak the frequency response to provide additional gain at our target frequencies. For similar values of L, C, center frequency, and gain, using a shunt-peaked load to gain-peak enables a more wideband response than the parallel RLC load typically used in narrowband LNAs.

To analyze the shunt-peaking response in more detail, we first define a variable:

$$\alpha = \frac{R^2 C}{L} \tag{4.20}$$

Assuming gain-peaking from the shunt-peaking load exists, the peak frequency is:

$$\omega_o^2 = \frac{1}{LC} * \left(\sqrt{1 + 2\alpha} - \alpha \right) \tag{4.21}$$

And the impedance at the peak frequency is:

$$|Z_{L,peak}| = R * \sqrt{\frac{1}{2\alpha\sqrt{1 + 2\alpha} - \alpha(2 + \alpha)}} \tag{4.22}$$

Fig. 4.21(a) and (b) plot the gain and bandwidth of a shunt-peaked amplifier as a function
of α. Peak frequency is set to 4.5GHz, and peak gain is arbitrarily normalized. As expected,
the higher the peak gain, the narrower the bandwidth, with smaller α indicating higher
gain and narrower band. Furthermore, larger C results in lower gain for the same α, while
bandwidth stays constant. Thus, despite using shunt-peaking to eliminate the problem of
low RC corner, it is still advantageous to minimize C as much as possible. The output of the
LNA is loaded, in addition to its own drain capacitance, by an estimated 160fF consisting
of 80fF from the subsequent stage, 30fF from routing parasitics, and 50fF from overhead
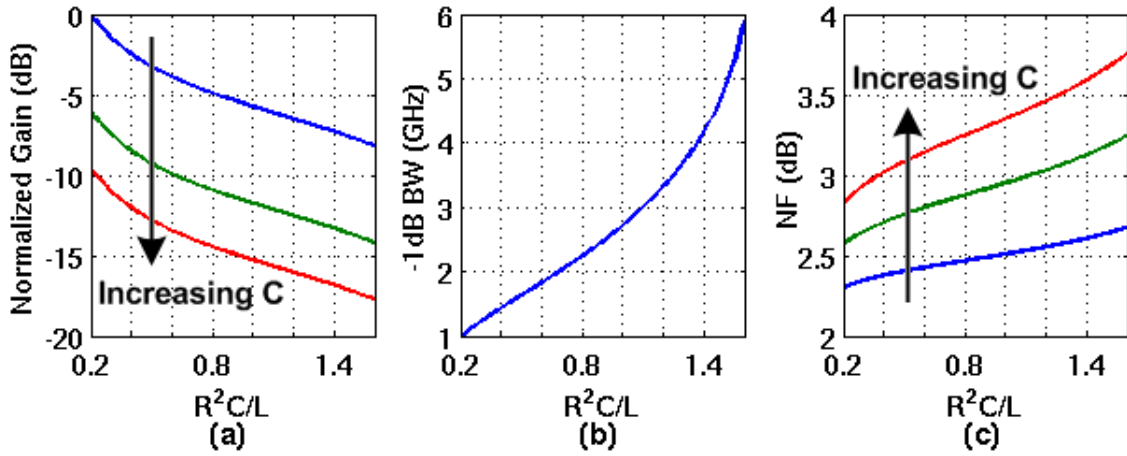circuitry to enable PA/LNA mode switching.

Figure 4.21: Gain-peaking parameters as a function of α: (a) normalized peak gain, (b) -1dB
bandwidth, (c) NF.

Next, we analyze the noise contribution of the shunt-peaking load. Only R_L is a noise
generator, and its noise contribution at the output node of the LNA is modulated by the
shunt-peaking frequency response.

$$\overline{v_{on,R_L}^2} = \frac{4kT}{R_L} \left(\frac{R_L}{s^2 C_L L_L + s C_L R_L + 1} \right)^2 = \frac{4kT}{R_L} Z_L^2 \left(\frac{R_L}{R_L + s L_L} \right)^2 \tag{4.23}$$

The noise factor of the base LNA including load noise, at peak frequency $\omega = \omega_o$, is:

$$F = 1 + \gamma \left(\frac{g_m}{G_M} \right) + \frac{4}{G_M R_L} \left(\frac{\alpha}{\sqrt{1 + 2\alpha}} \right) \tag{4.24}$$

Noise figure trends are shown in Fig. 4.21(c). As expected, NF closely tracks with LNA gain,
where a lower gain, such as from higher α or higher C_L, results in higher NF. As a balance
between gain, bandwidth, and NF, we use $\alpha \approx 0.8$ and $R_L \approx 120\Omega$.

Fig. 4.22 shows the layout of the shunt-peaking inductor. A high inductance was required,
while a high Q was unnecessary since any parasitic resistance from the inductor can be

absorbed into R_L. We therefore use a multi-turn square inductor for maximum area efficiency. The differential inductor, consisting of $2 * L_L$, has 5 turns and uses 60μm inner radius, 4μm trace width, and 4μm trace gap. The thin trace width and large gap were used to lower parasitic capacitance and achieve adequate SRF. The implemented differential inductor has $L = 7.7 \sim 12.7$nH from 3GHz to 6GHz, $Q = 5.7 \sim 7.0$, and SRF= 8.5GHz.
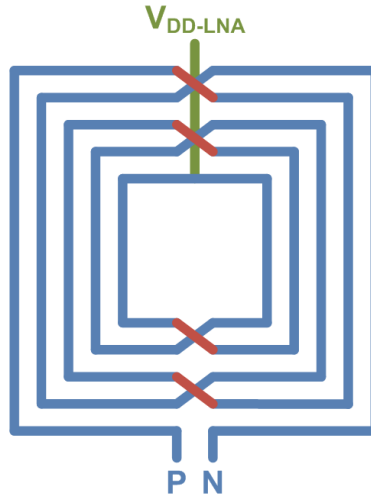

Figure 4.22: Layout of shunt-peaking inductor.

The simulated full LNA, embedded within the PA core and including the front-end transformer, achieves 18.1dB gain and NF=4.1dB at its peak frequency, while consuming 7.8mW. The LNA is highly sensitive to components and parasitics at its input node; consequently, a significant portion of the noise arises from the front-end. The base LNA, without front-end components but with fully-implemented load, has only NF=1.9dB. Its low NF is due to intentional source impedance mismatch in order to raise G_M. The front-end transformer itself contributes 1.2dB to NF, due to both direct noise generation and IL. The transformer center tap for the LNA is switched to ground with PA cells, and the finite ON resistance of these cells, 0.6Ω in total, contributes 0.5dB to NF, again due to both direct noise generation and IL. Finally, parasitics at the front-end such as top-level routing, pads, and bondwire inductance contribute a further 0.5dB to NF.

4.2.3   PA Supply Switch and Center Tap Design

Fig. 4.23 illustrates the PA core when it is configured in LNA mode. The PA cells sharing a drain node with LNA cells are turned off, while those on the opposite side of the LNA cells are turned on to create a connection to ground. The center tap of each sub-transformer must be open circuit and disconnected from the PA supply.

The most straightforward method of disconnecting the PA supply is to simply insert a supply switch. First, we analyze the design constraints on this supply switch and demonstrate why
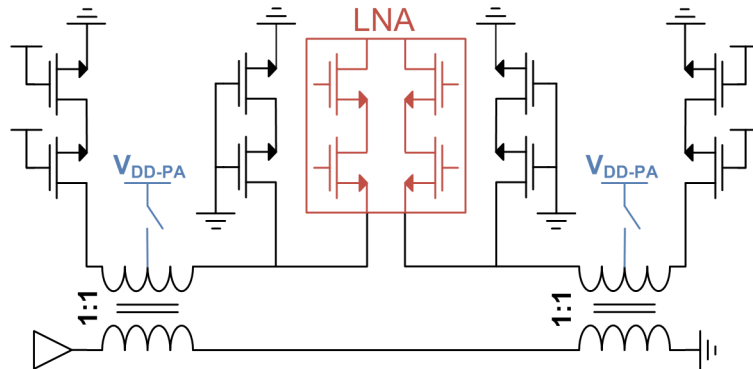
Figure 4.23: Schematic of PA core in LNA mode.

this naive approach is inadequate. To maintain good PA performance, the switch should
have very low ON resistance. Otherwise, IR drop across the switch effectively lowers the
supply voltage and reduces PA output power, while additional DC power consumed by the
switch degrades efficiency. The effects of switch R_{ON} on peak output power and peak drain
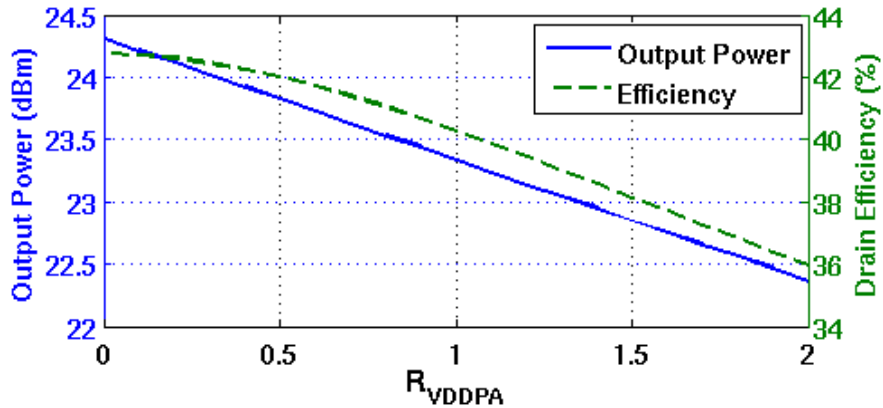efficiency are shown in Fig. 4.24.

Figure 4.24: PA peak output power and drain efficiency as a function of supply switch
resistance.

A large switch is obviously needed to achieve low R_{ON}; but in addition, because V_{DD-PA}
up to 1.5V may be used to attain target peak output power, the supply switch must also
be thick-oxide PMOS and thus *especially* large physically. Consequently, while the switch is
OFF in LNA mode, it contributes a proportionately large parasitic drain capacitance at the
sub-transformer center tap nodes. Fig. 4.25(a) illustrates this effect for one sub-transformer.
If C_P is large, its resonance frequency with the transformer inductance could easily fall in
or even below the band of interest.

Fig. 4.25(b) plots S21 of the front-end transformer in LNA mode for several switch sizes,
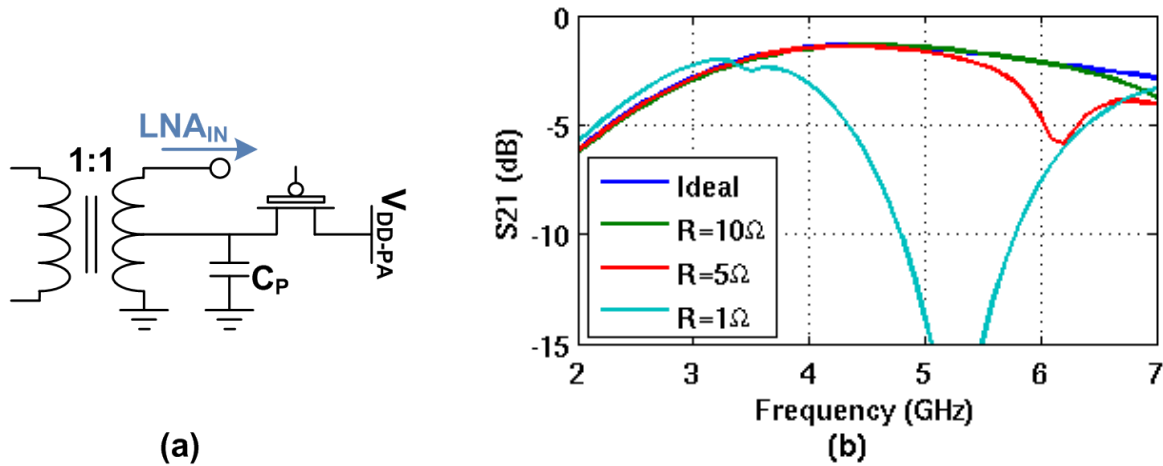denoted by their R_{ON}. To generate a small enough C_P that does not affect in-band per-

Figure 4.25: Parasitic OFF capacitance of PA supply switch: (a) schematic, (b) S21 of front-end transformer in LNA mode.

formance, an $R_{ON} > 10\Omega$ is needed, which is not tolerable by the PA. In contrast, at the $R_{ON} < 1\Omega$ preferred by the PA, the transformer in LNA mode is fundamentally not functional due to resonance of C_P with transformer inductance. Using a naive switch to disconnect PA supply has been shown to be insufficient, and another approach is needed.

There are two possible ways to accommodate a large C_P in LNA mode. The first is to simply resonate it out with another inductor in parallel, as shown in Fig. 4.25(a). There are several major drawbacks to this method however. Firstly, the parallel tank composed of C_P and L_P is high-Q, as it must be to create the open circuit desired by the LNA. Thus, the tank is narrowband and cannot protect the entire frequency band of interest from C_P degradation.

Secondly, there is a high noise penalty. Fig. 4.25(b) illustrates the noise model for R_P - the parasitic resistance of L_P - and it resembles the noise models of Fig. 4.6 in Sec. 4.1.2. Identical to the mechanism described in Sec. 4.1.2 and Eq. 4.12, R_P noise here is also multiplied by Q^2_{tank} factor through effective L-match networks. In fact, Q^2_{tank} factor is applied twice here: once through the high-Q C_P and L_P network, and once more through $L/2$ and C_S.[15] This leads to extremely high noise contribution at LNA input node.

Thirdly, the resonance frequency of C_P and L_P cannot be well-controlled because an accurate value for C_P cannot be easily obtained. The large supply switch is composed of thousands of device fingers, and even an approximate layout extraction would be extremely burdensome. For all of the above reasons, a parallel tank solution to eliminate C_P is not viable.

The second method for eliminating C_P is shown in Fig. 4.25(c). An inductor L_P is inserted in series with the supply switch, acting as a choke and isolating C_P from the center tap

---

[15]Note that $L_P \ll L/2$ when C_P is large. The loading of front-end tank components on the C_P and L_P network can therefore be ignored for an approximation.
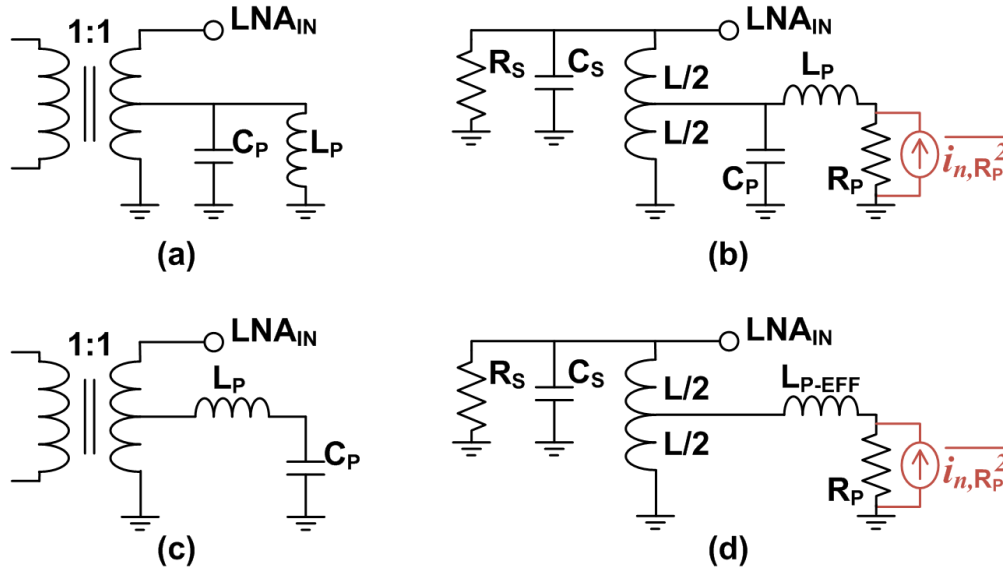
Figure 4.26: Options for addressing center tap capacitance C_P: (a) parallel resonance, (b) noise model for parallel resonance, (c) inductor choke, (d) noise model for inductor choke.

node. A central benefit of this method is that it does not rely on resonance, and therefore it achieves wideband functionality and insensitivity to exact C_P value. In fact, a larger C_P from a larger switch, layout parasitics, or other sources is now beneficial, as it would tune out a smaller fraction of L_P, leaving its remaining inductance to be a stronger choke.

Fig. 4.25(d) illustrates the noise model for the choke method. The inductor's parasitic resistance R_P is the noise contributor, and L_{P-EFF} is the effective inductance after C_P has been tuned out at the frequency of interest. There is still a Q_{tank}^2 noise multiplication factor from the $L/2$ and C_S network, but if L_{P-EFF} is an effective choke with $L_{P-EFF} \gg L/2$, then very little noise appears at the center tap node to be multiplied. Thus, noise contribution at LNA input remains low.

Next, we briefly discuss the effect of center tap inductance on PA performance. Fig. 4.27 shows a basic PA core with the component models that make up its output tank. The differential and common mode resonance frequencies of the tank are given below in Eq. 4.25. Differential resonance affects the PA's fundamental output and is what we have thus far been referring to when discussing front-end resonance. Common mode resonance, on the other hand, should ideally occur at the PA output's 2nd harmonic in order to create an open circuit at that frequency [42].

$$1/\omega_{DIFF}^2 = \left(\frac{C}{2} + C_{DIFF} \right) * L$$

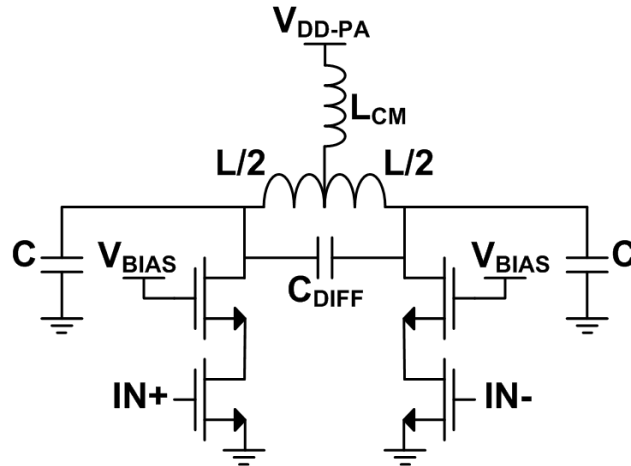$$1/\omega_{CM}^2 = 2C * \left(\frac{L}{4} + L_{CM} \right)$$

(4.25)

Figure 4.27: Schematic of PA core showing differential and common mode output resonance.

The vast majority of front-end capacitance is common mode C, comprising of, for example, the drain capacitance of PA cells. Thus, ω_{CM} in practice is approximately equivalent to ω_{DIFF} and significantly below the 2nd harmonic frequency even with $L_{CM} = 0$. Increasing L_{CM} lowers ω_{CM} even further, but the additional difference at the 2nd harmonic is minor. The greater problem for the PA arises from the choke inductor's parasitic resistance R_P, which degrades PA performance in the same manner as the supply switch's R_{ON}. Since attainable Q for on-chip inductors is limited, the maximum inductance the PA can tolerate is also effectively limited, although only for reasons of IR drop at DC and not RF response.
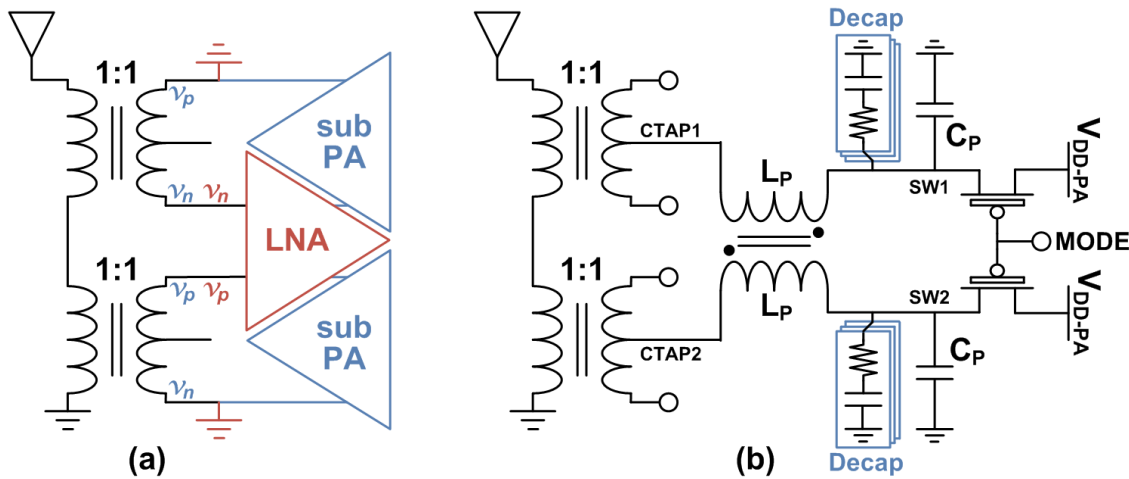

Figure 4.28: Design of center tap choke: (a) polarity in PA/LNA modes, (b) schematic of PA supply and center tap network.

In order to boost the LNA's choke inductance without using large inductors, we couple the two choke inductors from the two sub-transformers into a center tap transformer. In LNA

mode, the two center tap nodes have opposite signal polarity, as shown in Fig. 4.28(a). If we negatively couple the two center tap inductors, they generate positive mutual inductance and boost each other's effective inductance. On the other hand, in PA mode, the two center tap nodes see identical, in-phase signals. The negative coupling now creates a negative mutual inductance that cancels the center tap inductors' respective self inductances. This PA inductance cancellation effect does not change R_P, but it gives some marginal benefit from 2nd harmonic effects and is a net positive.

If the self-inductance of each choke is L_P, then the effective center tap inductance after coupling in PA and LNA modes are:

$$L_{EFF-LNA} = L_P * (1 + k)$$
$$L_{EFF-PA} = L_P * (1 - k)$$
(4.26)

The layout of the center tap transformer, composed of the two coupled choke inductors, is shown in Fig. 4.29. The transformer needs high Q for low PA supply IR drop and low LNA noise, and high k for strong mutual inductance. Thus, we use two strapped thick metal layers and three loops in parallel per inductor. For optimal Q, the inductors are single-turn and octogonal. The transformer uses 70μm inner radius, 10μm trace width, and 2μm trace gap.

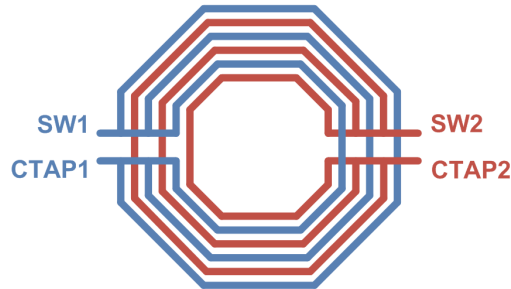Figure 4.29: Layout of center tap choke transformer.

At 4.5GHz, the implemented transformer has $L = 293$pH, $Q = 10.5$, and $k = 0.71$. The simulated effective inductance is 86pH for PA mode and 555pH for LNA mode. The LNA inductance, although not significantly greater than front-end transformer inductance $L/2$, was nevertheless adequate for LNA performance.

Fig. 4.29(b) illustrates the full center tap network and supply switching architecture. The implemented PMOS switch has post-extraction $R_{ON} = 420$mΩ. With its drain capacitance no longer a concern, the switch is limited primarily by physical area and diminishing returns due to routing parasitics. The supply bypass capacitors ("decap") for V_{DD-PA} are placed on the transformer side of the supply switch to help maximize choke inductance. The decap is implemented as a bank of RC branches to de-Q the bypass network for improved PA stability, identical to the decap from [48].
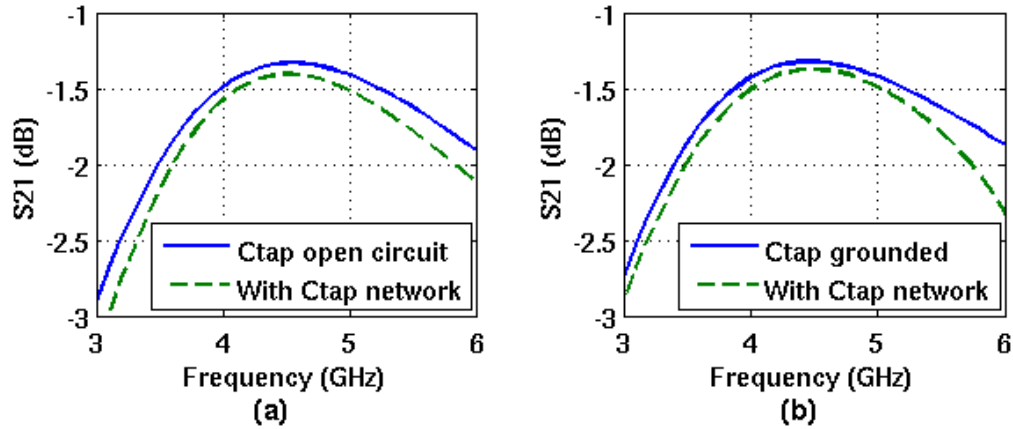
Figure 4.30: S21 of front-end transformer with center tap network: (a) LNA mode, (b) PA mode.

Fig. 4.30 plots S21 of the front-end transformer with and without center tap network. At peak frequency, degradation from the complete center tap network is below 0.1dB for both PA and LNA modes. Combined with the the PA and LNA cores, the center tap network contributes -0.3dB to LNA gain and 0.1dB to its NF. In PA mode, the center tap network contributes 1.0dB degradation in peak output power and 3.4% degradation in drain efficiency.

4.2.4 PA/LNA Mode Switching

Fig. 4.31 illustrates the full PA and LNA cores along with the front-end transformer. The PA core contains a total of 20 cells per sub-PA: 15 thermometer, 4 binary, and 1 dummy. Each cell is differential, composed of two halves that connect to opposing ends of each sub-transformer. The blue and red labels indicate how each PA device is driven in PA and LNA modes.

The two center branches in the diagram are LNA cells, and they connect to the LNA load. The LNA uses half of two thermometer cells per sub-PA. Adjacent to the LNA branches are the remaining 18 half-cells per sub-PA not used by the LNA but share the same PA drain/LNA source node. These are all switched off in LNA mode. The two outside branches comprise of the full 20 half-cells per sub-PA that connect to the opposing side from the LNA cells. These are all hard switched ON in LNA mode.

The top row of transistors are PA input devices. In PA mode, these are driven by the LO if the cell is active, and turned off if the cell is not. In LNA mode, logic in the PA pre-driver forces the PA inputs to ground or 1V (VDD of the PA pre-driver) depending on cell function. This mechanism will be described in more detail in Sec. 4.3.1. The bottom row of transistors are PA cascode devices. In PA mode, they are all biased with 1.2V. In LNA mode, aside from the LNA cells, they are driven to ground or 1.2V with additional custom logic using
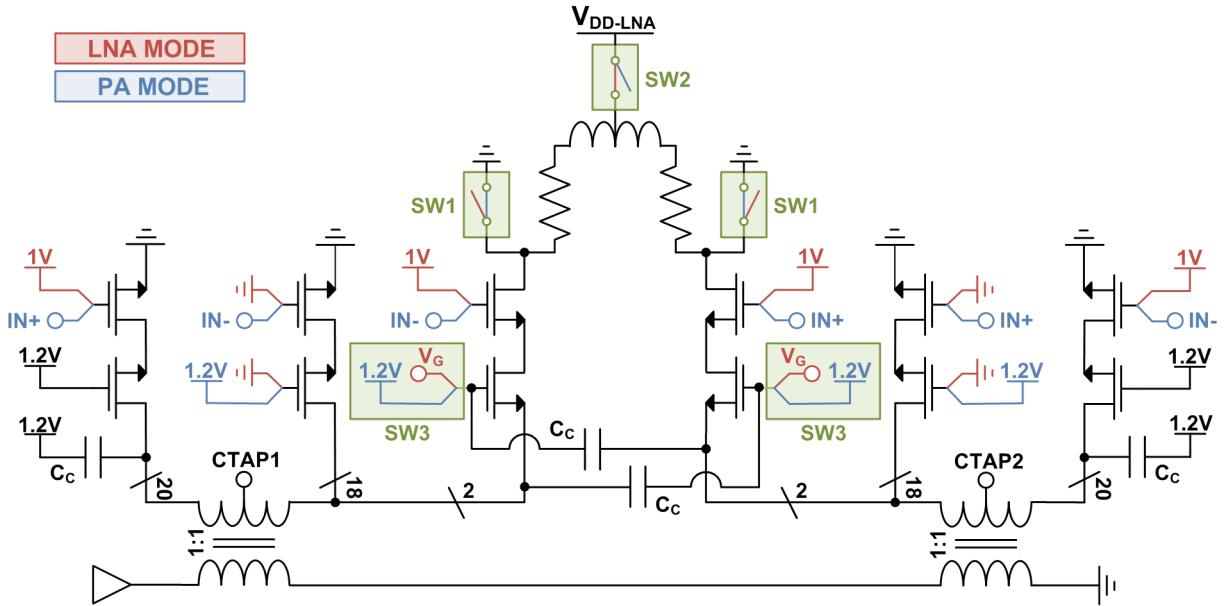
Figure 4.31: Schematic of PA/LNA core.

VDD=1.2V.

The two C_C capacitors in the center are the g_m-boosting capacitors for the LNA. The outer two C_Cs exist to balance the g_m-boosting capacitors so that the PA sees a symmetric load. Because the g_m-boosting C_Cs connect to 1.2V in PA mode, the outer C_Cs are connected to 1.2V as well for symmetry.

Aside from PA/LNA mode switching that occurs through digital logic, there are four discrete switches in the system. The first is the PA supply switch on the transformer center taps, which was covered extensively in Sec. 4.2.3. The remaining three are the 3 switches shown in Fig. 4.31. SW1 is the PA ground switch, SW2 is the LNA supply switch, and SW3 is the LNA bias switch.

The PA ground switch, SW1, is a thin-oxide NMOS that, in LNA mode, disconnects the PA input devices of the LNA cells (LNA cascode devices) from ground. LNA mode prefers a small SW1 so that it contributes little capacitance to the sensitive LNA output node, while PA mode prefers a large SW1 to provide low resistance to ground. Striking a balance between these two needs, the implemented SW1 has $R_{ON} = 6.3\Omega$ and $C_{OFF} = 51$fF. The switch makes negligible difference to PA performance, while its parasitic capacitance is manageable and included in the LNA design in Sec. 4.2.2.

Unlike all other mode switches, the PA ground switch is uniquely placed inside the PA cell in layout. This scheme avoids long traces routing to locations outside the PA core and enables minimized resistance to ground. To ensure good layout matching between PA cells, every

single PA cell contains the PA ground switch, although the switch is only active in the LNA cells. In all other cells, the switch is shorted with metal routing and acts as a dummy.

The LNA supply switch, SW2, disconnects the LNA supply in PA mode, so that there is no short circuit current from V_{DD-LNA} to ground. SW2 is placed at the center tap of the shunt-peaking inductor, where it can be made arbitrarily large with no adverse effects. SW2 is completely outside of the signal path, and it neither capacitively loads the signal nor degrades linearity as a signal path switch would. The switch is thick-oxide PMOS to accommodate V_{DD-LNA} up to 1.5V, and it has $R_{ON} = 2.9\Omega$. SW2 has no discernible effect on LNA performance.
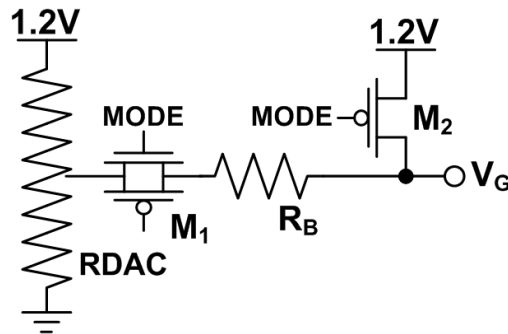

Figure 4.32: Schematic of LNA bias network.

The LNA bias switch, SW3, encompasses not a single switch, but a bias network for the LNA shown in Fig. 4.32. In LNA mode, MODE=1 and the gate bias of LNA input devices, V_G, is set by a resistive voltage DAC. The transmission gate $M1$ is small, as low R_{ON} is unnecessary: M1 resistance adds directly onto R_B, which is already large to enable g_m-boosting. The DAC is 6-bits and centered around nominal V_G with 5mV step.

In PA mode, MODE=0 and V_G is pulled up to 1.2V for PA cascode bias. The pull-up switch, M2, is thin-oxide PMOS, and it must be moderately large to create a strong bias that suppresses gate swing due to C_C. However, if M2 is too large, its parasitic capacitance would load the LNA input gate and reduce its g_m-boost. M2 has $R_{ON} = 3.6\Omega$ and $C_{OFF} = 160\text{fF}$. Like the PA ground switch, the parasitics of SW3 are already included in the LNA design, while they have negligible effect on PA performance.

The three switches described above, in addition to the center tap network from Sec. 4.2.3 and the logic that drives the PA cells, make up the "T/R switch" of our TDD front-end. In contrast to conventional in-line T/R switches, these PA/LNA mode switches are all low frequency and outside of the RF signal path. Although some of these switches have non-trivial design challenges, they nevertheless result in an efficient design that contributes minimal degradation to both PA and LNA performances.

4.3 System Implementation

The PA re-use TDD front-end is embedded in a full digital polar transmitter as proof-of-concept. Fig. 4.33 shows a block diagram of the implemented system.
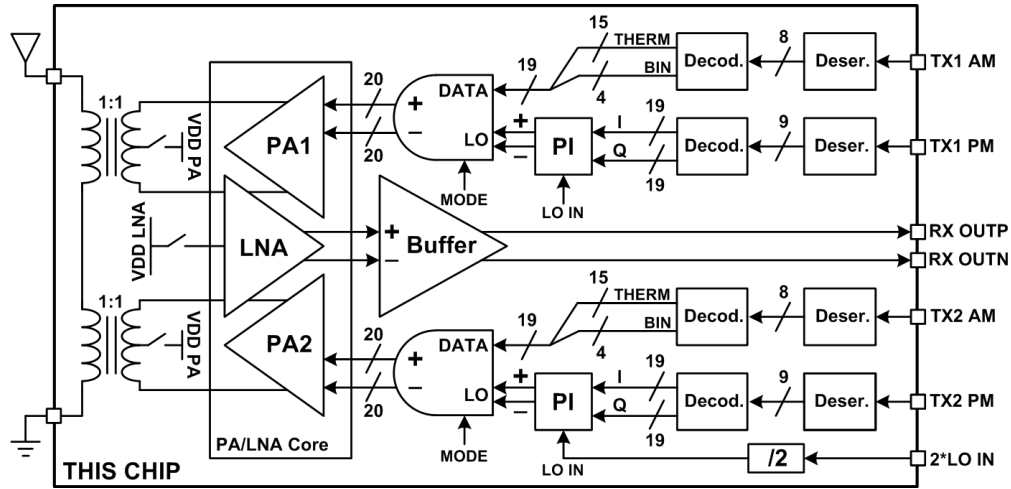

Figure 4.33: Block diagram of implemented system top-level.

We use the digital transmitter from [45], which has an 8-bit amplitude modulator and 9-bit phase modulator. For each sub-PA, identical AM and PM input data are deserialized and decoded. A phase interpolator (PI) uses the phase code to generate a differential LO waveform at the desired PA output frequency and phase, and a PA driver combines this LO with amplitude code to produce the switching inputs that drive the cells of the PA core. Sec. 4.3.1 delves into the detailed architecture of each transmitter block.

On the RX side, the LNA outputs are directly buffered off-chip for measurement. Sec. 4.3.2 describes the design of the RX buffer.

4.3.1 TX System

Fig. 4.34 shows a block diagram of the TX data deserializer. Both amplitude and phase data arrive to the chip differentially in serial 10-bit frames. The input signals are terminated, amplified, and then translated to digital domain with strong-arm latch-based comparators. A fast clock ("CLKF") at half the frequency of the input bitrate latches each data bit on both rising and falling edges, while a 5x slower clock ("CLKS") at the frequency of the frame rate latches the 10-bit parallel data after deserialization. Only 8 bits of the amplitude data are valid, and they are decoded into 15 thermometer and 4 binary bits, each controlling a corresponding PA cell. The phase decoder similarly resolves 9 bits of input phase data into I/Q 19-bit phase codes - 15 thermometer bits, 3 binary, and 1 sign bit.
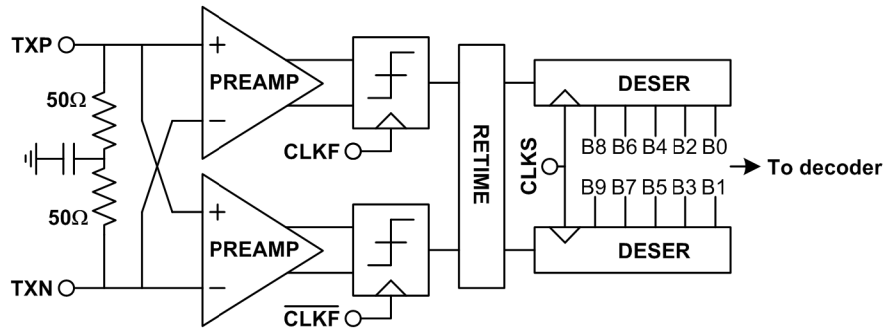
Figure 4.34: Block diagram of TX data deserializer.

Fig. 4.35 illustrates the clock distribution routing for the input 2*LO signal, which is at twice
the frequency of the desired PA output frequency. The signal is routed on differential trans-
mission lines, and the traces are laid out in the twisted fashion shown in Fig. 4.35 to improve
symmetry and reduce deterministic mismatch. A resistive splitter branches the input 2*LO
into two paths, with each feeding the transmitter chain for one sub-PA. The splitter ensures
that all 3 of its ports are simultaneously impedance-matched when properly terminated.
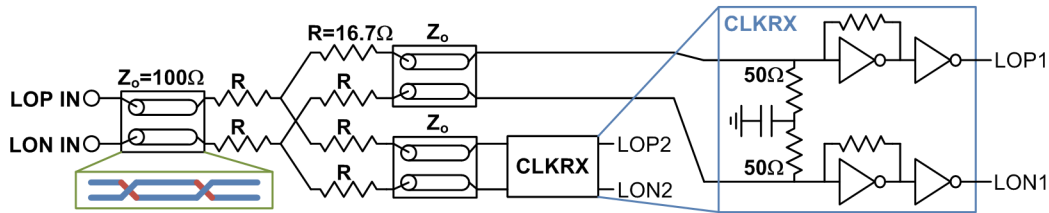Finally, inverter-feedback clock receivers regenerate the 2*LO signal for subsequent blocks.

Figure 4.35: Schematic of high-frequency LO splitter and clock receiver.

Fig. 4.36 presents a block diagram of the phase interpolation path. A frequency divider
creates quadrature LO waveforms, at the desired PA output frequency, from the input 2*LO.
I/Q current DACs, controlled by 19-bit I/Q phase codes, mix with the quadrature LO to
produce weighted quadrature currents. These currents are summed to generate a combined
LO output with desired phase. Current integrators limit the amplitudes of the quadrature
LO driving the Gilbert cell mixers. At the output, a CML-to-CMOS converter and buffers
regenerate the PI output to full rail-to-rail swing.

Finally, Fig. 4.37 illustrates the PA driver. There are 20 PA drivers cells per sub-PA, one
for each PA cell including the dummy. Fig. 4.37(a) shows the base PA driver cell, which is
a simple AND operation between the PI output LO and amplitude code. Fig. 4.37(b) shows
modifications added to the base cell to enable PA/LNA mode switching. In PA mode, when
MODE=1, the logic acts the same as the base cell and outputs the LO and amplitude code's
AND result. In LNA mode, when MODE=0, the bottom logic chain forces a steady VDD
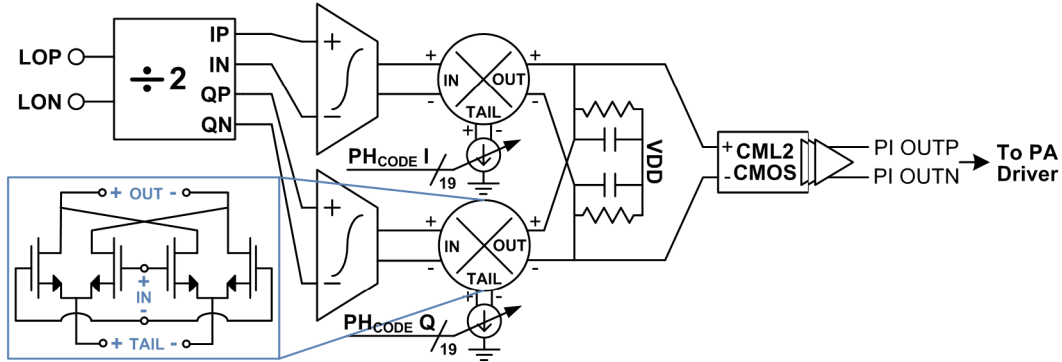
Figure 4.36: Block diagram of phase interpolation path.

output while the top chain forces steady ground. Each PA driver cell uses the appropriate logic depending on the function of its corresponding PA cell in LNA mode. The two logic chains are delay-equalized with each other.
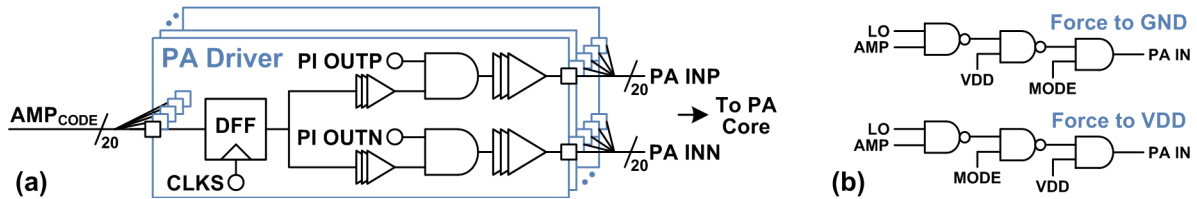

Figure 4.37: Block diagram of PA driver: (a) base, (b) with added logic for T/R mode switching.

4.3.2 RX Buffer

A schematic of the RX buffer is shown in Fig. 4.38. The buffer consists of two common-source (CS) gain stages followed by a pseudo-differential output driver. The gain stages can be bypassed, and devices added for bypassing function are shown in blue in Fig. 4.38. Gain mode, in which the CS stages are included, is used to measure NF, while bypass mode, in which the CS stages are bypassed, is used to measure linearity.

Source-follower topology was chosen for the output driver to achieve the best linearity. Since the g_m of the source follower is limited by 50Ω output impedance match requirement, the gain stages are necessary to limit the high noise contribution of the output driver to RX system NF. On the other hand, RX linearity should be limited by the LNA and not the buffer in order to facilitate meaningful measurements. Thus, the gain stages are made bypassable.

Shown as V_X in Fig. 4.38, the input node to the output driver selects between the output of the gain stages and the buffer input. The bypass path uses transmissions gates to achieve maximum linearity, while PMOS switches are sufficient for the gain path. In bypass mode,
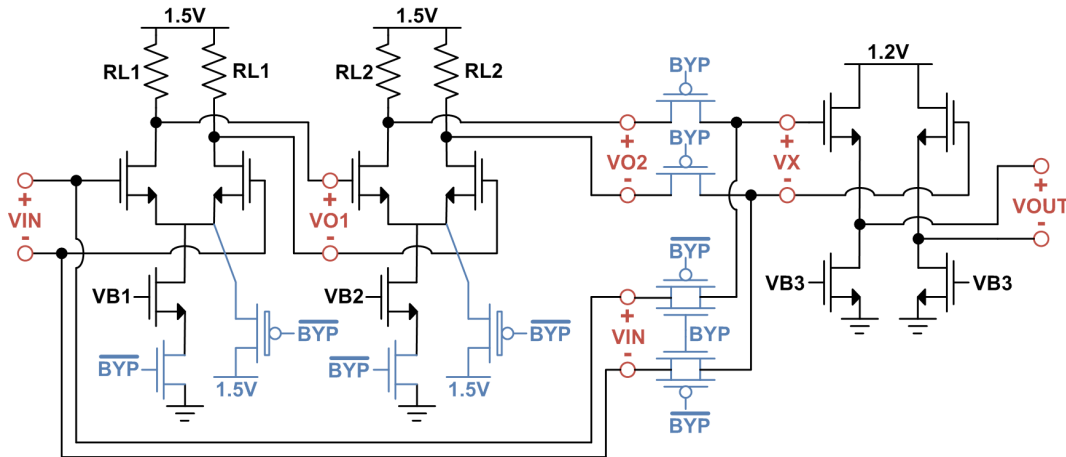
Figure 4.38: Schematic of RX output buffer.

the gain stages are turned off to minimize unnecessary capacitive loading at the LNA output.
While the gain stages are off, their tail current sources are disconnected, and their tail nodes
are driven to VDD to prevent high voltage stress across any gate-source, gate-drain, or
source-drain junctions.

Fig. 4.39(a) plots the simulated gain of the entire RX system in both gain and bypass modes.
The gain of the standalone LNA block is also shown. The CS stages have about 9dB of gain
each, while the output driver has approximately -8dB gain, including the 6dB due to output
impedance match. Fig. 4.39(b) similarly plots the simulated NF of the entire RX along with
standalone LNA NF. In gain mode, the RX buffer contributes an additional 0.2dB NF to
the standalone LNA, raising LNA NF from 4.2dB at its lowest point to 4.4dB.

Figure 4.39: Simulated system RX performance in gain and bypass modes: (a) gain, (b) NF.

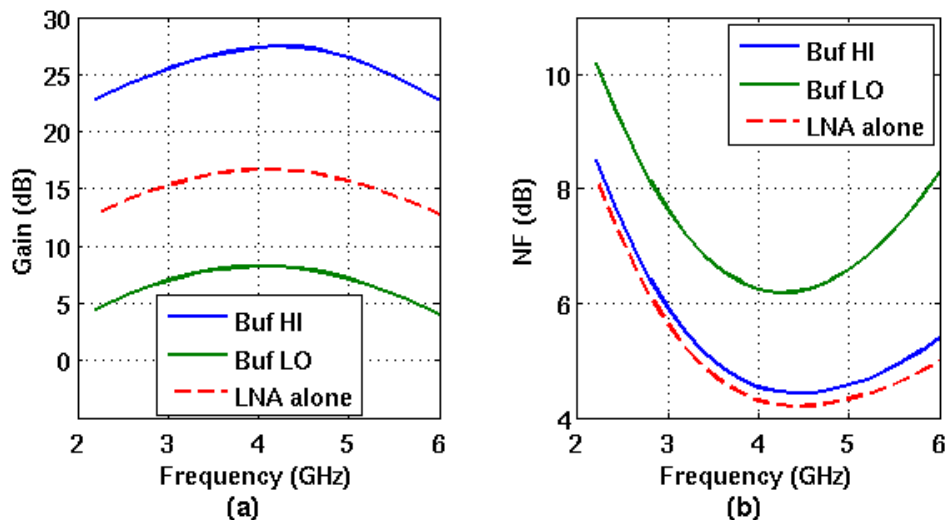The RX buffer in bypass mode has simulated input P1dB of 11.6dBm. This is greater than the standalone LNA's output P1dB and therefore ensures that RX linearity measurements will be defined by the LNA's linearity performance.

4.4 Measurements

A prototype of the TDD system was fabricated and measured. In this section, Sec. 4.4.1 describes the fabricated chip as well as the design of the test PCB. Sec. 4.4.2 describes the test setup for both TX and RX measurements. Finally, Sec. 4.4.3 presents the measured performance results.

4.4.1 Chip Implementation and PCB Design

A system prototype was fabricated in TSMC 65nm flip-chip, and a chip micrograph is shown in Fig. 4.40. The chip measures 2.31mm x 2.61mm but is pad-limited. Assuming that the PA, LNA, and front-end transformer exist in comparable TDD front-ends, the area overhead used to implement T/R switching functionality reduces down to the PA supply switching structure, which has active area of 0.25mm^2 including the supply transformer. All other mode switches are negligible in area in comparison.

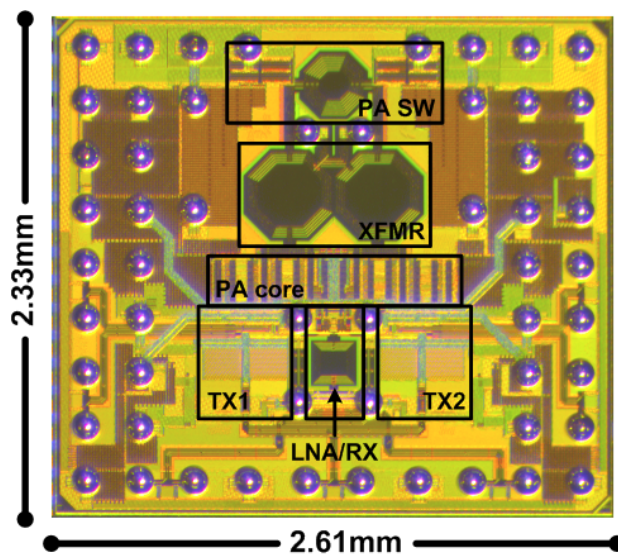Figure 4.40: Chip micrograph of implemented TDD system.

The flip-chip die is attached directly to PCB with no intermediate package. This method keeps bondwire inductance to under 200pH. In contrast, other methods such as chip-on-board wirebonding or a flip-chip packaging have bondwire inductances of 2nH to 5nH. We use a 10-layer FR-4 PCB consisting of top and bottom signal layers, 2 internal signal layers,

2 split-plane power layers, and 4 ground planes. Fig. 4.41(a) shows a diagram of the PCB
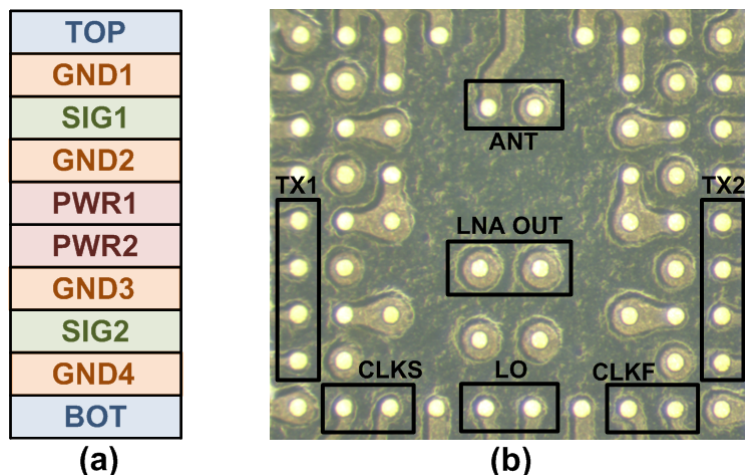stack-up.

Figure 4.41: PCB implementation: (a) layer stack-up, (b) micrograph of die area from
fabricated PCB.

Fig. 4.41(b) shows a micrograph of the the die area from the fabricated PCB. The light green
areas are top layer copper pads and traces, and the yellow circles are soldermask openings.
For die-attach, we use soldermask-defined pads that are 4mil in diameter - same as solder
ball size - and have 10mil (254µm) pitch. High-frequency ports such as the antenna, clocks,
and TX data inputs are placed along the perimeter of the chip, so that they can be routed
directly on PCB with top layer traces.[16,17] Interior pads not on the perimeter are routed
using via-under-pads. These structures are identifiable in Fig. 4.41(b) from their larger
copper diameter to accommodate via drill feature sizes.

All high-frequency signals are routed with controlled-impedance traces and shielding. Fur-
thermore, ground planes are voided under SMA pads for all sensitive signals. Because the
board is relatively thin with many layers, SMA pad capacitance would otherwise be ex-
tremely high ($\approx$800fF) due to the thin gap between a signal and its nearest ground plane.
The topmost ground layer is also voided under the transformer and inductor areas of the die
to reduce leakage and parasitic capacitance.

---

[16]The LNA output port, despite being high-frequency, is not placed on the perimeter due to limited space
and the prioritization of other signals.

[17]The antenna port ("ANT") is configured in balun mode. Its positive terminal is routed single-endedly
while its negative terminal is routed using via-under-pad to the ground planes.

4.4.2 Measurement Setup

Fig. 4.42 illustrates the measurement setup in PA and LNA modes. Chip control signals and TX data inputs are generated from a Xilinx VC707 FPGA board, which communicates with the test PCB through an FMC interface. The FPGA configures PCB components such as supply regulators as well. A signal generator provides the required 100MHz reference clock for the FPGA.
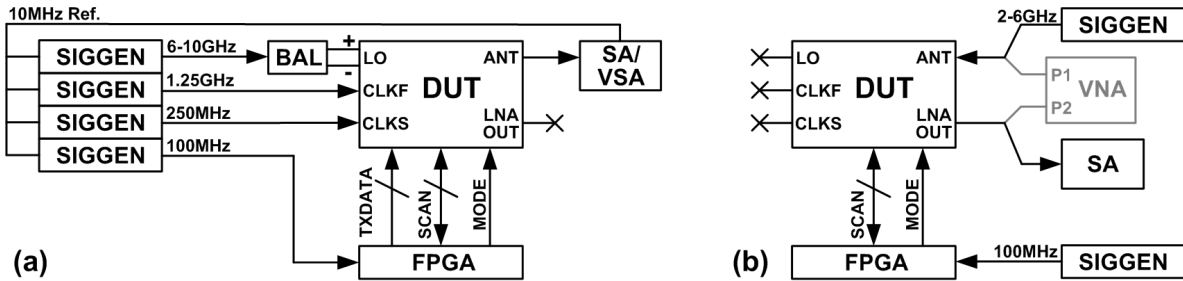

Figure 4.42: Measurement setup in (a) PA mode, (b) LNA mode.

In PA mode, signal generators produce the TX deserialization clocks as well as the $2*LO$ PA clock input. The FPGA outputs TX data at 2.5Gbps datarate, resulting in CLKF=1.25GHz and CLKS=250MHz. The deserialization clocks have on-board baluns while an SMA connector balun was necessary for LO due to its much higher frequency. The PA output is measured at the antenna port with a spectrum analyzer. Built-in vector signal analysis software in the spectrum analyzer was used for modulation tests. All equipment, including the FPGA, share a common 10MHz reference for timing alignment.

In LNA mode, S-parameters are measured with a 2-port vector network analyzer, and noise figure and linearity are measured with a spectrum analyzer. The LNA output is measured both differentially with an external SMA connector balun, and single-endedly in certain cases due to the complexities of de-embedding the external balun and its cables. Single-ended measurements are taken by terminating one port of the differential LNA output with a broadband 50Ω load and measuring the other port.

4.4.3 Measurement Results

Fig. 4.43 plots the gain and noise performance of the system in LNA mode across the frequency band of interest. Cables and PCB traces have been de-embedded. Including the buffer in high gain mode, the full RX path has peak gain of 26.8dB and -3dB bandwidth of 2.7GHz. Noise figure, also including RX buffer, is 5.1dB at its lowest point, and the NF +1dB bandwidth is 2GHz. Compared to simulation results, both the LNA's gain and noise figure performance are degraded by about 0.5dB. However, both quantities are very wideband. The LNA consumes 9mA from 1.5V.
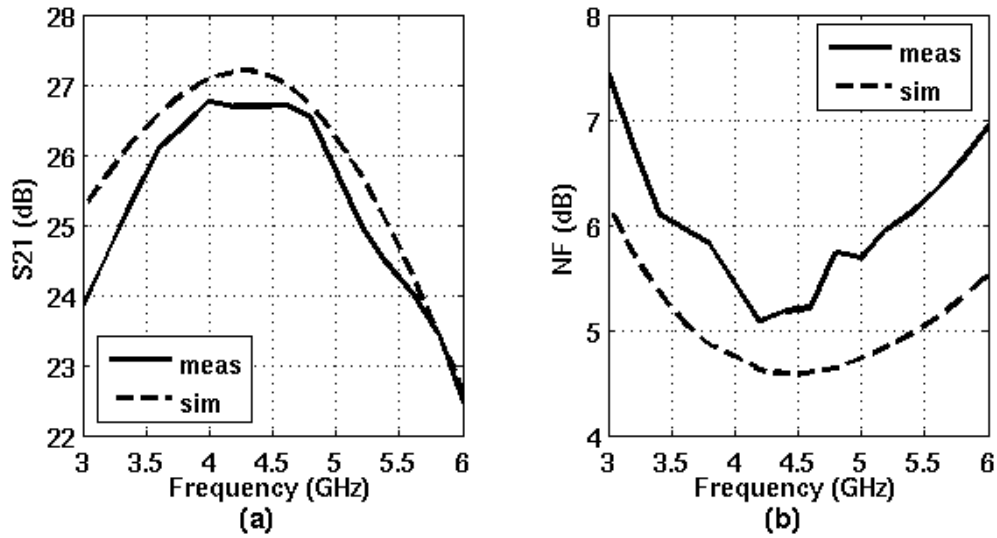
Figure 4.43: Measured RX (a) S21 and (b) NF performance across frequency.

Fig. 4.44 plots measured S11 and S22 of system RX across the frequency band of interest. S22 is below -10dB across the entire band. S11 is degraded at the lower edge of the band and reaches -10dB at the midband frequency of 4.5GHz. Measured S11 performance, although sup-optimal, is comparable to simulation results that displayed similar degraded S11 behavior. Input impedance match was compromised during system implementation from the amalgamation of various parasitics such as top level routing, pads, bondwire inductance, and PCB vias and traces. This is an area for improvement in future designs.
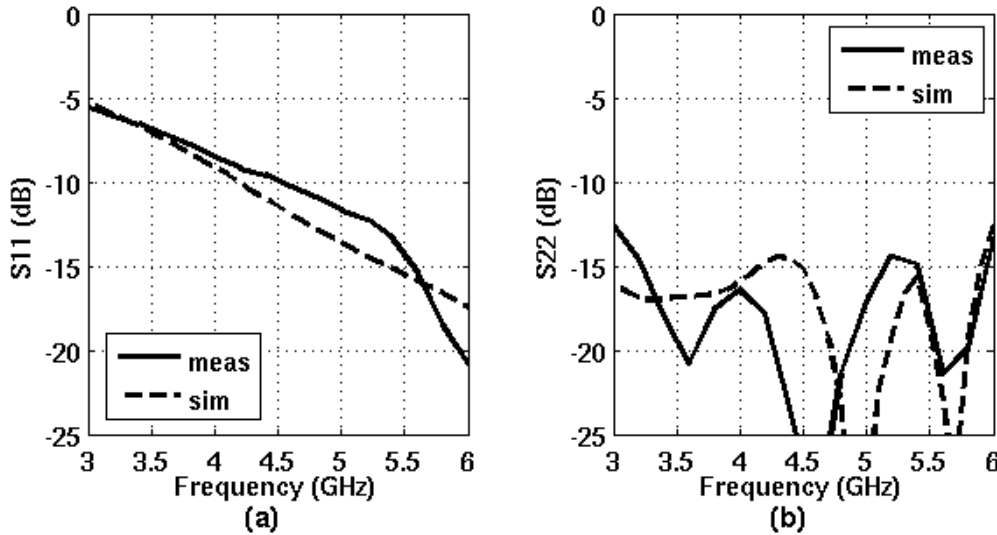
Figure 4.44: Measured RX (a) S11 and (b) S22 performance across frequency.

Fig. 4.45 shows the LNA's linearity performance by plotting RX output power as a function of its input power. The RX buffer is in bypass mode. The input-referred P1dB is -6.7dBm.
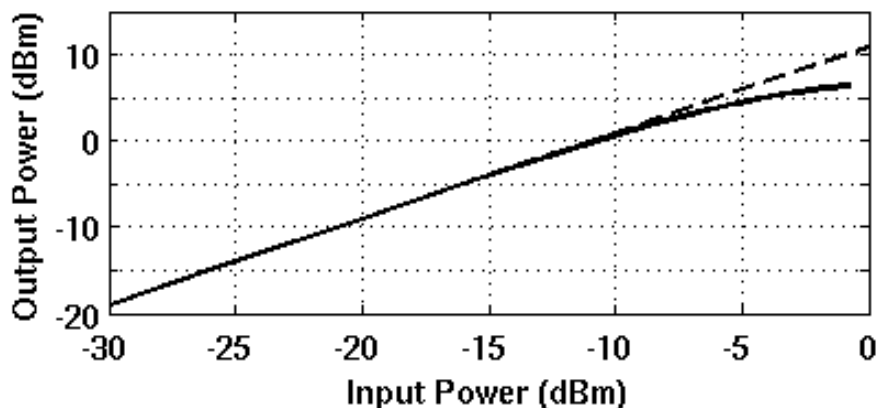
Figure 4.45: Measured RX output power as a function of input power.
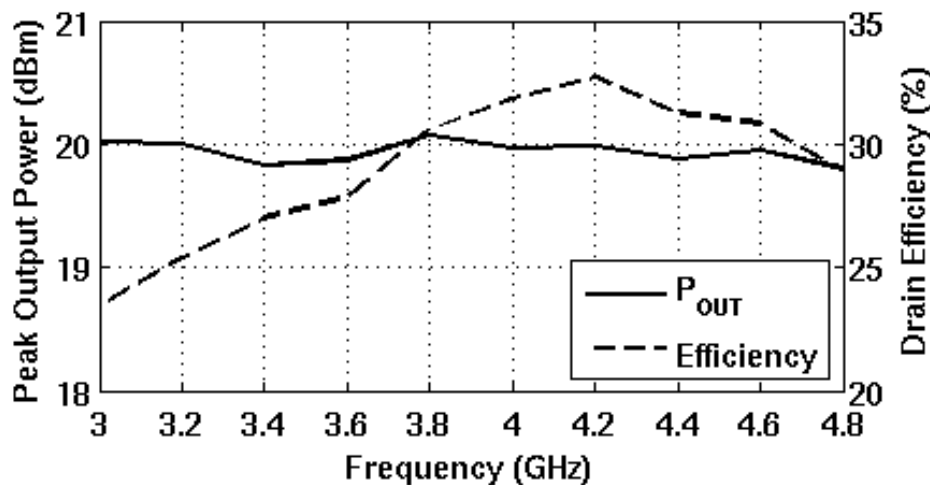
Figure 4.46: Measured peak PA output power and peak drain efficiency across frequency.

Fig. 4.46 shows measured P_{SAT} and drain efficiency across frequency. Once again, cables and PCB traces have been de-embedded. With 1.2V supply voltage, the PA achieves 32.7% peak drain efficiency at 20.0dBm output power at 4.2GHz. Peak total efficiency accounting for the entire transmitter's power consumption is 22.0%.

The PA had originally been intended to be operated up to 1.5V supply to target 23dBm

P_{SAT}, but the higher supply could not be applied due to robustness issues.[18] From both measurements and simulation, it is estimated that approximately 2dB in output power is lost due to using a 1.2V supply instead of 1.5V. Accounting for this supply voltage difference, measured PA output power differs from simulation by about 1dB. In addition, performance greater than 5GHz unfortunately could not be measured due to difficulties of propagating a robust >10GHz 2*LO clock to the chip. However, existing results indicate wideband PA frequency response similar to that of the LNA.

Fig. 4.47 shows the AM-AM and AM-PM characteristics of the transmitter across the full range of all amplitude codes, measured with 4.2GHz carrier. The compressive AM-AM behavior and the AM-PM distortion are characteristic of digital PAs and can be linearized with pre-distortion. A measured AM-AM response after pre-distortion is also shown in Fig. 4.47. All predistortion is calculated using look-up tables based on measured results.
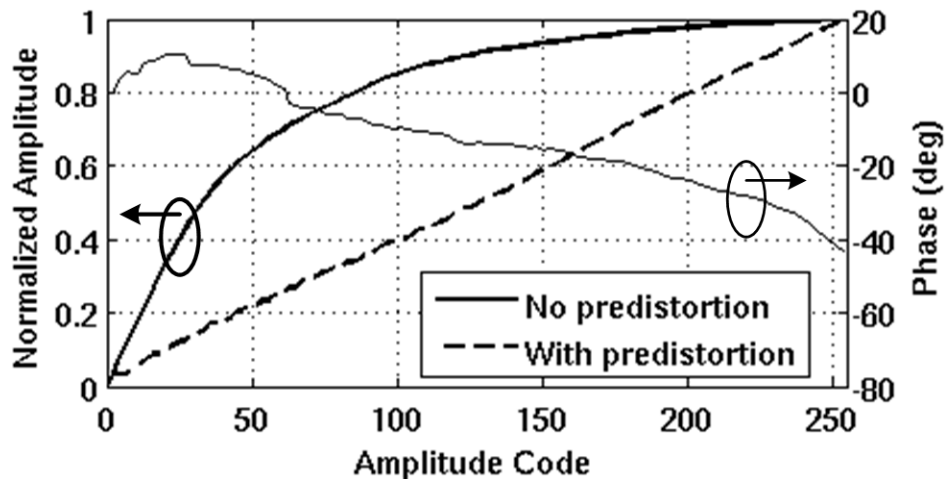
Figure 4.47: Measured transmitter AM-AM and AM-PM performance across AM codes.

Fig. 4.48 shows the PM-PM and PM-AM characteristics of the transmitter across the full range of phase codes, also measured with 4.2GHz carrier. The 4-quadrant curved PM-PM response is due to I/Q phase interpolation and can also be linearized with predistortion as shown in Fig. 4.48. Fig. 4.49 plots DNL of the measured PM-PM response both with and without pre-distortion to present a magnified view of the effect of phase linearization. PM-AM distortion is negligible.

[18]When the PA is operated at 1.5V supply and at peak output power, there results a permanent DC leakage current from the PA supply of up to ~25mA. Although there are no accompanying degradations to PA output power, it nevertheless seems that device oxide breakdown has occurred somewhere in the system. The higher supply was aggressive, and for future work, thin-oxide devices should be operated at 1.2V supply or below.
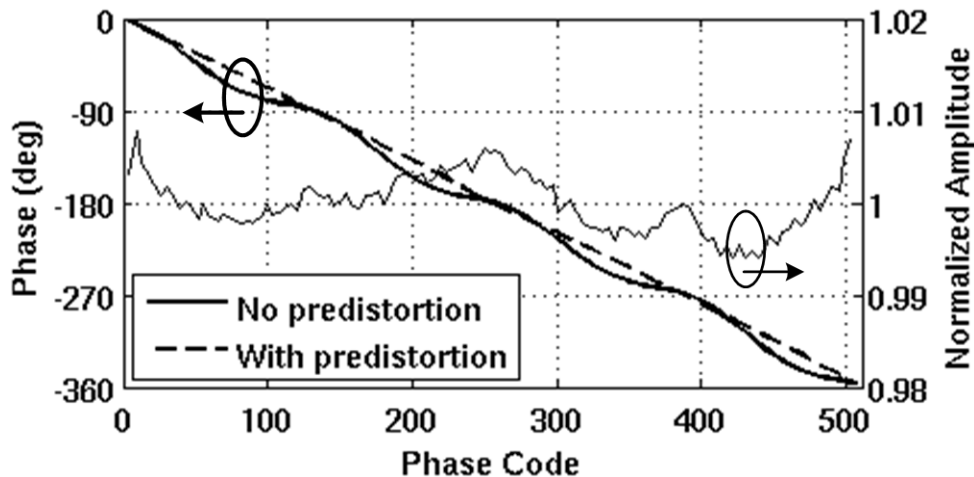
Figure 4.48: Measured transmitter PM-PM and PM-AM performance across PM codes.
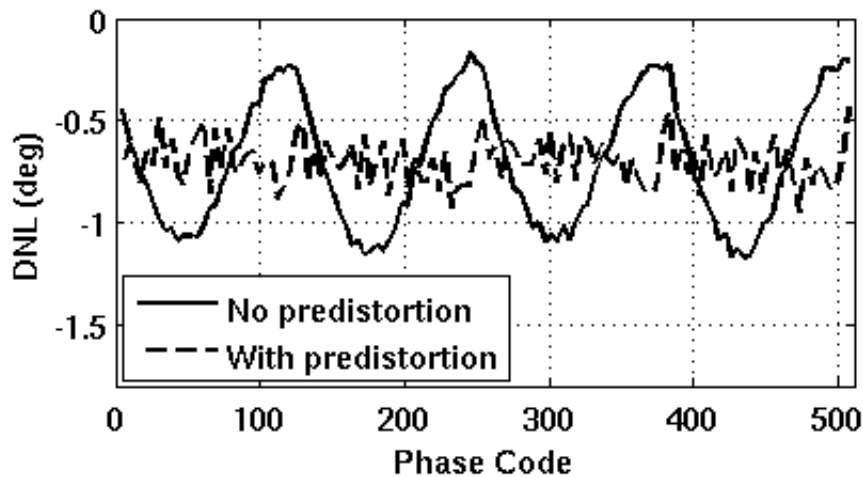
Figure 4.49: DNL of measured PM-PM performance with and without pre-distortion.

Fig. 4.50 shows a QAM16 constellation measurement at 4.2GHz and with 16.4dBm average P_{OUT}. Along with the AM and PM code measurements, it has been demonstrated that the implemented transmitter is fully functional and capable of operating with modulated signals.

As proof-of-concept for our proposed T/R switching technique, we have achieved re-use of the PA into a reasonable LNA while maintaining PA functionality. Simultaneously, we have presented a wideband TDD front-end with integrated T/R switching and balun and without frequency tuning, which has not previously been reported. TDD co-existence has been achieved with no series RF switches in the signal path, and with only low-frequency mode control switches. This work thus contributes a key innovation towards greater front-end integration and reconfigurability in multi-band radios.
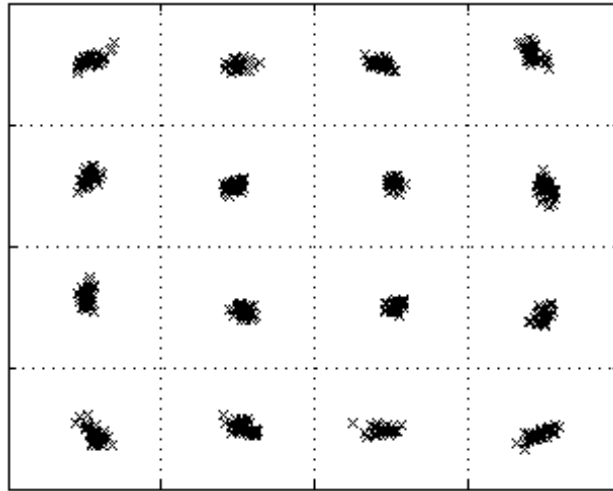
Figure 4.50: Sample measured QAM16 constellation with $P_{OUT} = 16.4$dBm.

Chapter 5

Conclusion

With the growing proliferation of smartphones and other internet-connected mobile devices, there have been an exponential increase in mobile data usage. This is a trend that is expected to continue in the near future. To accommodate this growth, there needs to be an accompanying increase in energy efficiency and spectrum efficiency of wireless networks. Reconfigurable radios have the potential enable these efficiency gains through its characteristics of front-end integration, wideband transceiver architectures, and spectrum awareness. In this thesis, we presented two chip implementations of systems that advanced the state-of-the-art in the area of reconfigurable radios.

First, we presented a spectrum sensing receiver for cognitive radio applications in the UHF TV band. This system uses known characteristics of the target signal to produce an energy efficient architecture employing subsampling downconversion and digital-analog hybrid correlation. Fine and coarse detection modes enable both rapid sensing and the detection of very weak signals. For a 6MHz TV band channel, the system achieves -91dBm sensitivity with rapid energy detection, -104dBm sensitivity with correlation, and 84dB of full-scale dynamic range, while consuming only 28mW of power. This work was the first integrated spectrum sensing receiver to achieve sensitivity greater than conventional data receivers, and this work achieved the highest sensitivity per bandwidth while consuming the lowest power.

While the spectrum sensing receiver was targeted narrowly at a particular standard and band, it employed a variety of techniques for high sensitivity and low power that can be generalized to other applications. For example, dual-mode sensing enabling both fast detection and high sensitivity has not been previously demonstrated. Relatedly, digital-analog hybrid correlation, as a method of achieving energy-efficient correlation functionality, had only been previously used for audio applications. This work pioneered its use in low-power wireless and spectrum sensing applications. In addition, this work incurred significant power savings by both taking advantage of known characteristics of the target signal and ignoring the signal's data content, focusing instead on detecting its energy content. Previous works, in contrast, have either attempted to build a higher sensitivity conventional data receiver, which

is unnecessarily high power, or to build a high resolution spectrum analyzer, which is blind to signal characteristics and adds unnecessary complexity.

There are two possible directions for future work on spectrum sensing receivers for reconfigurable radios. The first is to develop from this work a more general purpose, multi-band spectrum sensing receiver front-end. Techniques can be developed to efficiently detect not only one, but a variety of possible signal characteristics and modulations. Future work could also target a truly integrated and blocker-resilient front-end that does not rely on SAW filtering for attenuation of out-of-band interference. Furthermore, improvements can be made to the receiver's linearity and dynamic range, as greater linearity would be necessary in a multi-band detector that must accommodate higher powered signals and blockers than those within the TV band.

While the first possible future direction focuses on expanding front-end reconfigurability for spectrum sensing receivers, a second possible direction could instead focus on integration of the signal-processing backend. As a further proof-of-concept for the spectrum detection techniques presented in this work, there is value in the development of a complete system with on-chip energy detectors and integrated implementation of the hybrid correlation scheme. There is a large design space in this area to determine, for example, the best balance between analog, digital, and mixed-signal processing to achieve optimal energy efficiency while maintaining robust detection and low latency.

Next, in this thesis, we presented a wideband TDD front-end with PA re-use T/R switching technique. By re-using the PA as an LNA during receive mode, the system eliminates the conventional series T/R switch from the signal path and utilizes only DC mode control switches to enable TDD co-existence. With integrated front-end balun transformer, the full polar transmitter achieves 20dBm peak output power with 32.7% peak drain efficiency. In receive mode, the PA is reconfigured into a wideband 3.4GHz-5.4GHz LNA achieving -6.7dBm P1dB and 5.1dB NF.

The TDD front-end provided two main contributions to state-of-the-art. Firstly, it demonstrated functionality of a completely new integrated T/R switching technique that has not previously been proposed or attempted. The concepts of PA re-use and single block shared PA/LNA have not existed in any previous work on TDD co-existence or integrated T/R switches. We successfully implemented PA re-use scheme inside a fully functional polar transmitter, and we demonstrated a reasonable gigahertz wideband LNA can be built from PA transformation while maintaining transmitter functionality and performance.

Secondly, this work presents the only wideband integrated TDD front-end without frequency tuning and including full TX, while achieving comparable performance to narrowband systems. Furthermore, TDD co-existence at RF frequencies is achieved with no series RF switches in the signal path, relying only on low-frequency power and mode switches. In contrast, existing works have still depended on the very traditional series-shunt switch topology

to achieve wideband functionality since past innovations in integrated T/R switching have been narrowband.

While the TDD front-end demonstrated reasonable performance as a proof-of-concept, further performance improvements are needed before it can be truly competitive and usable in real-world mobile wireless applications. The primary areas for improvement and future work are LNA noise figure, LNA input impedance, and PA output power. While LNA noise figure is to some extent limited by its common gate topology, the front-end passive network is nevertheless a major contributor of noise due to both parasitics and insertion loss. Re-architecting the front-end transformer, possibly using different turns ratios or alternate stacking structures, could potentially reduce NF significantly while also providing improved input impedance matching for the LNA. Higher PA output power could easily be achieved with the use of thick-oxide cascodes and higher PA supply voltages. However, higher PA supplies and output power will bring about new challenges for the design of shared PA/LNA core and power and mode switches. Thus, extending PA re-use to higher and more commercially feasible PA output power levels is a robust area for future research as well.

Finally, while the two systems presented in this thesis contribute key innovations towards greater front-end integration and reconfigurability, they obviously represent only a small portion of the possible areas of research in the broader space of reconfigurable radios. To aim towards the theoretical ideal reconfigurable radio shown in Fig. 1.2, much works remains to be done.

Bibliography

[1] K. Walsh. (2010). RF Switches Guide Signals In Smart Phones, Skyworks, Inc., [Online]. Available: `http://www.skyworksinc.com/downloads/press_room/published_articles/Microwave_RF_092010.pdf`.

[2] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

[3] *Second Report and Order and Memorandum Opinion and Order*, ET Docket 08-260, U.S. FCC, 2008.

[4] D. B. Cabric and R. W. Brodersen, "Cognitive Radios: System Design Perspective," PhD thesis, EECS Department, University of California, Berkeley, 2007. [Online]. Available: `http://www.eecs.berkeley.edu/Pubs/TechRpts/2007/EECS-2007-156.html`.

[5] *Second Memorandum Opinion and Order*, ET Docket 10-174, U.S. FCC, 2010.

[6] *Evaluation of the Performance of Prototype TV-Band White Space Devices, Phase II*, OET 08-TR-1005, U.S. FCC, 2008.

[7] *ATSC Digital Television Standard, Part 2 - RF/Transmission System Characteristics (A/53, Part 2:2007)*, Advanced Television Systems Committee, 2007.

[8] S. J. Shellhammer, *Spectrum Sensing in 802.22*, Workshop on Cognitive Information Processing, 2008.

[9] *ATSC Recommended Practice: Receiver Performance Guidelines (A/74:2010)*, Advanced Television Systems Committee, 2010.

[10] R. Tandra, S. M. Mishra, and A. Sahai, "What is a Spectrum Hole and What Does it Take to Recognize One?" *Proceedings of the IEEE*, vol. 97, no. 5, pp. 824–848, 2009.

[11] J. Park, T. Song, J. Hur, S. M. Lee, J. Choi, K. Kim, K. Lim, C. H. Lee, H. Kim, and J. Laskar, "A Fully Integrated UHF-Band CMOS Receiver With Multi-Resolution Spectrum Sensing (MRSS) Functionality for IEEE 802.22 Cognitive Radio Applications," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 1, pp. 258–268, 2009.

[12] M. Kitsunezuka, H. Kodama, N. Oshima, K. Kunihiro, T. Maeda, and M. Fukaishi, "A 30-MHz-2.4-GHz CMOS Receiver With Integrated RF Filter and Dynamic-Range-Scalable Energy Detector for Cognitive Radio Systems," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 5, pp. 1084–1093, 2012.

[13] M. S. O. O. Alink, E. A. M. Klumperink, M. C. M. Soer, A. B. J. Kokkeler, and B. Nauta, "A 50Mhz-To-1.5Ghz Cross-Correlation CMOS Spectrum Analyzer for Cognitive Radio with 89dB SFDR in 1Mhz RBW," in *2010 IEEE Symposium on New Frontiers in Dynamic Spectrum*, 2010, pp. 1–6.

[14] M. S. O. Alink, E. A. M. Klumperink, A. B. J. Kokkeler, W. Cheng, Z. Ru, A. Ghaffari, G. J. M. Wienk, and B. Nauta, "A CMOS spectrum analyzer frontend for cognitive radio achieving +25dBm IIP3 and -169 dBm/Hz DANL," in *2012 IEEE Radio Frequency Integrated Circuits Symposium*, 2012, pp. 35–38.

[15] T. H. Yu, C. H. Yang, D. Cabric, and D. Markovic, "A 7.4mW 200MS/s wideband spectrum sensing digital baseband processor for cognitive radios," in *2011 Symposium on VLSI Circuits (VLSIC)*, 2011, pp. 254–255.

[16] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Experimental study of spectrum sensing based on energy detection and network cooperation," in *Proc. First Int'l. Workshop on Technology and Policy for Accessing Spectrum (TAPAS)*, 2006.

[17] E. Alon, V. Abramzon, B. Nezamfar, and M. Horowitz, "On-Die Power Supply Noise Measurement Techniques," *IEEE Transactions on Advanced Packaging*, vol. 32, no. 2, pp. 248–259, 2009.

[18] H. Pekau and J. W. Haslett, "A 2.4 GHz CMOS sub-sampling mixer with integrated filtering," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 11, pp. 2159–2166, 2005.

[19] S. Lerstaveesin, M. Gupta, D. Kang, and B. S. Song, "A 48-860 MHz CMOS Low-IF Direct-Conversion DTV Tuner," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 9, pp. 2013–2024, 2008.

[20] P. H. Dietz and L. R. Carley, "Analog/digital hybrid vlsi signal processing using single bit modulators," in *1993 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 1993, pp. 136–139.

[21] A. M. Abo and P. R. Gray, "A 1.5-V, 10-bit, 14.3-MS/s CMOS pipeline analog-to-digital converter," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 5, pp. 599–606, 1999.

[22] K. Kundert, *Simulating Switched-Capacitor Filters with SpectreRF*, The Designer's Guide Community, 2006.

[23] X. Xiao and B. Nikolic, "A dual-mode, correlation-based spectrum sensing receiver for TV white space applications achieving -104dBm sensitivity," in *2014 IEEE Radio Frequency Integrated Circuits Symposium*, 2014, pp. 317–320.

[24] (2015). E-UTRA: User Equipment (UE) Radio Transmission and Reception, 3GPP TS 36.101, [Online]. Available: `http://www.3gpp.org/DynaReport/36101.htm`.

[25] (2016). Product Brief: General Purpose RF Switches, Skyworks, Inc., [Online]. Available: `http://www.skyworksinc.com/uploads/documents/PB_RFSwitches_PB121_15B.pdf`.

[26] A. Cicalini, S. Aniruddhan, R. Apte, F. Bossu, O. Choksi, D. Filipovic, K. Godbole, T. P. Hung, C. Komninakis, D. Maldonado, C. Narathong, B. Nejati, D. O'Shea, X. Quan, R. Rangarajan, J. Sankaranarayanan, A. See, R. Sridhara, B. Sun, W. Su, K. van Zalinge, G. Zhang, and K. Sahota, "A 65nm CMOS SoC with embedded HSDPA/EDGE transceiver, digital baseband and multimedia processor," in *2011 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2011, pp. 368–370.

[27] M. He, R. Winoto, X. Gao, W. Loeb, D. Signoff, W. Lau, Y. Lu, D. Cui, K. S. Lee, S. W. Tam, P. Godoy, Y. Chen, S. Joo, C. Hu, A. A. Paramanandam, X. Wang, C. H. Lin, and L. Lin, "A 40nm dual-band 3-stream 802.11a/b/g/n/ac MIMO WLAN SoC with 1.1Gb/s over-the-air throughput," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 350–351.

[28] J. Moreira, S. Leuschner, N. Stevanovic, H. Pretl, P. Pfann, R. Thringer, M. Kastner, C. Prll, A. Schwarz, F. Mrugalla, J. Saporiti, U. Basaran, A. Langer, T. D. Werth, T. Gossmann, B. Kapfelsperger, and J. Pletzer, "A single-chip HSPA transceiver with fully integrated 3G CMOS power amplifiers," in *2015 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2015, pp. 1–3.

[29] N. Klemmer, S. Akhtar, V. Srinivasan, P. Litmanen, H. Arora, S. Uppathil, S. Kaylor, A. Akour, V. Wang, M. Fares, F. Dulger, A. Frank, D. Ghosh, S. Madhavapeddi, H. Safiri, J. Mehta, A. Jain, H. Choo, E. Zhang, C. Sestok, C. Fernando, R. K. A., S. Ramakrishnan, V. Sinari, and V. Baireddy, "A 45nm CMOS RF-to-Bits LTE/WCDMA FDD/TDD 2x2 MIMO base-station transceiver SoC with 200MHz RF bandwidth," in *2016 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2016, pp. 164–165.

[30] K. Yamamoto, T. Heima, A. Furukawa, M. Ono, Y. Hashizume, H. Komurasaki, S. Maeda, H. Sato, and N. Kato, "A 2.4-GHz-band 1.8-V operation single-chip Si-CMOS T/R-MMIC front-end with a low insertion loss switch," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 8, pp. 1186–1197, 2001.

[31] F.-J. Huang and K. O, "A 0.5- um CMOS T/R switch for 900-MHz wireless applications," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 3, pp. 486–492, 2001.

[32] S. Goswami, H. Kim, and J. L. Dawson, "A Frequency-Agile RF Frontend Architecture for Multi-Band TDD Applications," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2127–2140, 2014.

[33] A. A. Kidwai, C. T. Fu, J. C. Jensen, and S. S. Taylor, "A Fully Integrated Ultra-Low Insertion Loss T/R Switch for 802.11b/g/n Application in 90 nm CMOS Process," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 5, pp. 1352–1360, 2009.

[34] R. Winoto, M. He, Y. Lu, D. Signoff, E. Chan, C. H. Lin, W. Loeb, J. Park, and L. Lin, "A WLAN and bluetooth combo transceiver with integrated WLAN power amplifier, transmit-receive switch and WLAN/bluetooth shared low noise amplifier," in *2012 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2012, pp. 395–398.

[35] Y. Wang, H. Wang, C. Hull, and S. Ravid, "A Transformer-Based Broadband Front-End Combo in Standard CMOS," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 8, pp. 1810–1819, 2012.

[36] H. Xu and K. K. O, "A 31.3-dBm Bulk CMOS T/R Switch Using Stacked Transistors With Sub-Design-Rule Channel Length in Floated p-Wells," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 11, pp. 2528–2534, 2007.

[37] A. Madan, M. J. McPartlin, Z. F. Zhou, C. W. P. Huang, C. Masse, and J. D. Cressler, "Fully Integrated Switch-LNA Front-End IC Design in CMOS: A Systematic Approach for WLAN," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 11, pp. 2613–2622, 2011.

[38] Q. Li and Y. P. Zhang, "CMOS T/R Switch Design: Towards Ultra-Wideband and Higher Frequency," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 3, pp. 563–570, 2007.

[39] C. T. Fu, H. Lakdawala, S. S. Taylor, and K. Soumyanath, "A 2.5GHz 32nm 0.35mm^2 3.5dB NF -5dBm P1dB fully differential CMOS push-pull LNA with integrated 34dBm T/R switch and ESD protection," in *2011 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2011, pp. 56–58.

[40] H. S. Chen, H. Y. Tsai, L. X. Chuo, Y. K. Tsai, and L. H. Lu, "A 5.2-GHz full-integrated RF front-end by T/R switch, LNA, and PA co-design with 3.2-dB NF and +25.9-dBm output power," in *2015 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, 2015, pp. 1–4.

[41] R. Winoto, A. Olyaei, M. Hajirostam, W. Lau, X. Gao, A. Mitra, O. Carnu, P. Godoy, L. Tee, H. Li, E. Erdogan, A. Wong, Q. Zhu, T. Loo, F. Zhang, L. Sheng, D. Cui, A. Jha, X. Li, W. Wu, K. S. Lee, D. Cheung, K. W. Pang, H. Wang, J. Liu, X. Zhao, D. Gangopadhyay, D. Cousinard, A. A. Paramanandam, X. Li, N. Liu, W. Xu, Y. Fang, X. Wang, R. Tsang, and L. Lin, "A 2x2 WLAN and Bluetooth combo SoC in 28nm CMOS with on-chip WLAN digital power amplifier, integrated 2G/BT SP3T switch and BT pulling cancelation," in *2016 IEEE International Solid-State Circuits Conference (ISSCC)*, 2016, pp. 170–171.

[42] L. Ye, "Design and Analysis of Digitally Modulated Transmitters for Efficiency Enhancement," PhD thesis, EECS Department, University of California, Berkeley, 2013. [Online]. Available: `http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-99.html`.

[43] D. Chowdhury, L. Ye, E. Alon, and A. Niknejad, "An Efficient Mixed-Signal 2.4-GHz Polar Power Amplifier in 65-nm CMOS Technology," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 8, pp. 1796–1809, 2011.

[44] G. Liu, "Fully Integrated CMOS Power Amplifier," PhD thesis, EECS Department, University of California, Berkeley, 2006. [Online]. Available: `http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-162.html`.

[45] L. Ye, J. Chen, L. Kong, E. Alon, and A. Niknejad, "Design Considerations for a Direct Digitally Modulated WLAN Transmitter With Integrated Phase Path and Dynamic Impedance Modulation," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 12, pp. 3160–3177, 2013.

[46] P. Haldi, D. Chowdhury, P. Reynaert, G. Liu, and A. Niknejad, "A 5.8 GHz 1 V Linear Power Amplifier Using a Novel On-Chip Transformer Power Combiner in Standard 90 nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 5, pp. 1054–1063, 2008.

[47] A. M. Niknejad, *Electromagnetics for High-Speed Analog and Digital Communication Circuits*. Cambridge University Press, 2007.

[48] D. Chowdhury, "Efficient Transmitters for Wireless Communications in Nanoscale CMOS Technology," PhD thesis, EECS Department, University of California, Berkeley, 2010. [Online]. Available: `http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-168.html`.