Fall November 2014

Signal Processing in Wireless Communications: Device Fingerprinting and Wide-Band Interference Rejection

Adam C. Polak
University of Massachusetts Amherst

SIGNAL PROCESSING IN WIRELESS COMMUNICATIONS: DEVICE FINGERPRINTING AND WIDE-BAND INTERFERENCE REJECTION

A Dissertation Presented

by

ADAM C. POLAK

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2014

Electrical and Computer Engineering

SIGNAL PROCESSING IN WIRELESS COMMUNICATIONS: DEVICE FINGERPRINTING AND WIDE-BAND INTERFERENCE REJECTION

A Dissertation Presented

by

ADAM C. POLAK

Approved as to style and content by:

Dennis L. Goeckel, Chair

Prof. Marco F. Duarte, Member

Prof. Robert W. Jackson, Member

Prof. Brian N. Levine, Member

Prof. C. V. Hollot, Department Chair
Electrical and Computer Engineering

To my parents for everything that they taught me.
To my sisters for their great friendship.

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor Professor Dennis Goeckel. It has been a great honor to be his Ph.D. student. I appreciate all his contributions of time and ideas that made my Ph.D. experience productive and exciting. His great research enthusiasm was contagious and very motivational. I am thankful for his continuous availability, encouragement and a great example of excellent work ethic that he provided me with during the entire program. Thank you for the knowledge and inspiration that you gave me as an academic advisor, and also for everything that I had chance to learn from you on a non academic-level.

Thank you to my dissertation committee, Professors Marco Duarte, Robert Jackson and Brian Levine for their valuable advice and feedback on my dissertation research. I am grateful to Professor Robert Jackson for his ideas that motivated a big part of research reported in Chapters 5, 6 and 7; to Professor Marco Duarte for his great technical supervision while working on Chapters 5, 6 and 7; and to Professor Brian Levine for sharing his perspective on the topics of security and privacy.

I would like to thank Professor Kristian Kroschel for being an inspirational teacher and a great academic advisor during my undergraduate studies. I would also like to thank Dr. Güenter Nill (Agilent Technologies) for his great mentorship during my undergraduate studies. I am grateful to all my supervisors that I had a chance to collaborate with during my industrial internships at both the undergraduate and graduate level. Thank you for sharing your great expertise with me and for the exposure to practical challenges of electrical engineering. Thank you Brad Doerr, Brian Schipke, Tim Figgie, Nick Fernandez (Agilent Technologies), Samel Celebi, Luca Blessent, Dan Filipovic, Paul Draxler, Phil Coan (Qualcomm Inc.) for your

great guidance, genuine friendship, for all the precious advice that I received from you and for the interesting and challenging tasks and great summers of 2008, 2009, 2011, 2012 and 2013. All these internships were very motivational.

I also thank my lab-mates for a great collaborative working environment that we were able to create for each other here at the University of Massachusetts Amherst.

Finally, I would like to deeply thank my parents and my sisters for their great friendship and constant support that they provided me with during the entire program.

ABSTRACT

SIGNAL PROCESSING IN WIRELESS COMMUNICATIONS: DEVICE FINGERPRINTING AND WIDE-BAND INTERFERENCE REJECTION

SEPTEMBER 2014

ADAM C. POLAK

M.Sc., GDANSK UNIVERSITY OF TECHNOLOGY

Dipl.Ing., KARLSRUHE INSTITUTE OF TECHNOLOGY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Dennis L. Goeckel

The rapid progress of wireless communication technologies that has taken place in recent years has significantly improved the quality of everyday life. However with this expansion of wireless communication systems come significant security threats and significant technological challenges, both of which are due to the fact that the communication medium is shared. The ubiquity of open wireless Internet access networks creates a new avenue for cyber-criminals to impersonate and act in an unauthorized way. The increasing number of deployed wide-band wireless communication systems entails technological challenges for effective utilization of the shared medium, which implies the need for advanced interference rejection methods. Wireless security and interference rejection in wide-band wireless communications are therefore often considered as the two main challenges in wireless network's design and research. Important aspects of these challenges are illuminated and addressed in this dissertation.

This dissertation considers signal processing approaches for exploiting or mitigating the effects of non-ideal components in wireless communication systems. In the first part of the dissertation, we introduce and study a novel, model-based approach to wireless device identification that exploits imperfections in the transmitter caused by manufacturing process nonidealities. Previous approaches to device identification based on hardware imperfections vary from transient analysis to machine learning but have not provided verifiable accuracy. Here, we detail a model-based approach, that uses statistical models of RF transmitter components: digital-to-analog converter, power amplifier and RF oscillator, which are amenable for analysis. Our proposed approach examines the key device characteristics that cause anonymity loss, countermeasures that can be applied by the nodes to regain the anonymity, and ways of thwarting such countermeasures. We develop identification algorithms based on statistical signal processing methods and address the challenging scenario when the units that need to be distinguished from one another are of the same model and from the same manufacturer. Using simulations and measurements of components that are commonly used in commercial communications systems, we show that our anonymity breaking techniques are effective.

In the second part of the dissertation, we consider innovative approaches for the acquisition of frequency-sparse signals with wide-band receivers when a weak signal of interest is received in the presence of a very strong interference, and the effects of the nonlinearities in the low-noise amplifier at the receiver must be mitigated. All samples with amplitude above a given threshold, dictated by the linear input range of the receiver, are discarded to avoid the distortion caused by saturation of the low noise amplifier. Such a sampling scheme, while avoiding nonlinear distortion that cannot be corrected in the digital domain, poses challenges for signal reconstruction techniques, as the samples are taken non-uniformly, but also non-randomly. The considered approaches fall into the field of compressive sensing (CS); however, what

differentiates them from conventional CS is that a structure is forced upon the measurement scheme. Such a structure causes a violation of the core CS assumption of the measurements' randomness. We consider two different types of structured acquisition: signal independent and signal dependent structured acquisition. For the first case, we derive bounds on the number of samples needed for successful CS recovery when samples are drawn at random in predefined groups. For the second case, we consider enhancements of CS recovery methods when only small-amplitude samples of the signal that needs to be recovered are available for the recovery. Finally, we address a problem of spectral leakage due to the limited processing block size of block processing, wide-band receivers and propose an adaptive block size adjustment method, which leads to significant dynamic range improvements.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

The expansion of wireless communication technologies that has taken place in recent years entails inherent security threats and technological challenges due to the shared transmission medium. In this dissertation, we consider signal processing methods that exploit longstanding non-idealities of wireless *transmitters*, allowing for effective device identification. We also consider signal processing methods that mitigate the effects of the longstanding non-idealities of wireless *receivers*, allowing for an efficient medium utilization.

In particular, in the first part of this dissertation (Chapters 2, 3, and 4) we study signal processing methods that allow for the extraction of *fingerprints* of wireless devices from slight impurities present in the transmitter's components, such as the digital-to-analog converter, power amplifier and RF oscillator. These *fingerprints* can then be used as unique identification tags instead of conventional device tags such as, e.g. the Media Access Control (MAC) address or Electronic Serial Number (ESN), which can be easily modified by the users. We introduce novel, *model-based* identification methods that use statistical models of RF transmitter components and lead to effective device identification.

In the second part of the dissertation (Chapters 5 and 6), we consider the problem of interference rejection in wide-band communication systems for the challenging case when the in-band interferer has power orders of magnitude higher than the power of the message of interest and is causing saturation and nonlinear operation of the receiver's front-end. Motivated by a novel sampling approach of Jackson [5]

designed to mitigate the effects of this nonlinearity, we study new recovery methods. In particular, we derive bounds on the number of samples required for successful compressive sensing (CS) recovery of sparse signals when the samples are taken at random in pre-defined groups. We also consider methods to enhance the recovery of frequency-sparse signals from signal-dependent, low-amplitude samples. In addition, in Chapter 7, we address the problem of spectral leakage due to limited processing block size in wide-band receivers and propose an adaptive block size adjustment method that leads to significant dynamic range improvements.

1.1 Model-Based Approach for Mobile Device Identification

1.1.1 Motivation and Related Work

A significant increase in the number of crimes, such as the sexual exploitation of children [6], software piracy [7], intellectual property and identity theft [8], financial fraud [9], committed via the Internet, as well as the increase of financial losses caused by these crimes, have been reported in recent years. The Internet Crime Complaint Center (IC3): a multi-agency task force consisting of the Federal Bureau of Investigation, the National White Collar Crime Center, and the Bureau of Justice Assistance, receives about 300,000 complaints yearly from victims of crimes committed using the Internet. This is a significant increase when compared to tens of thousands of complaints reported in the early 2000s (Figure 1.1). The total financial loss caused by the Internet crime reported to IC3 in 2012 was estimated as $525,441,110.00 [2].

The ubiquity of wireless Internet access networks and mobile computing has revolutionized cyber-violations as crimes have become easier and cheaper to commit. Moreover, the cyber-criminals gain de facto anonymity when exploiting ubiquitous *open* wireless access points (APs), be it at airports, in shopping malls, public libraries, or coffee shops. Fortunately, the use of computers by the offenders typically results in digital evidence that can be used for the device and eventual user identification.

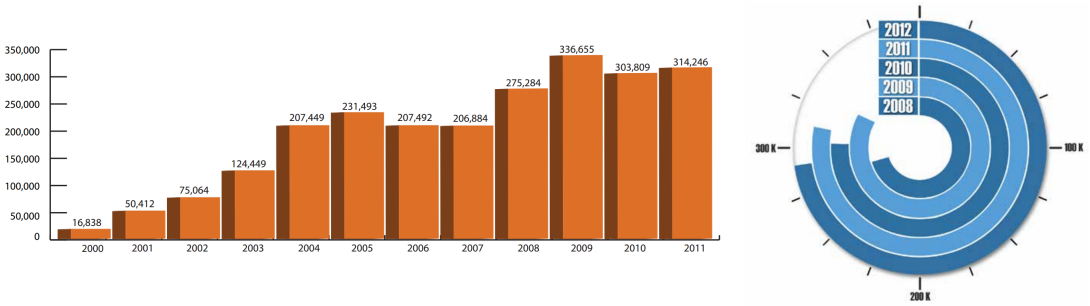

Figure 1.1. Number of crimes committed using the Internet reported to the Internet Crime Complaint Center in recent years [2].

Most of the well established techniques for device identification focus on desktop systems on the wired Internet. The primary artifact used in the investigations of crimes involving the wired Internet is the IP address of the suspect's computer. A consistent IP address assigned by an Internet service provider tags all outgoing and incoming traffic. With the introduction of mobile access systems come, however, new significant challenges for the crime investigators. In particular, in the case of wireless networks, IP addresses cannot be fully relied on by the investigators as unique and consistent tags, as they are often assigned to the users dynamically. Therefore, the ubiquity of open wireless APs has created a new avenue for persons to act with anonymity. The MAC address of a network interface card, which is globally unique, is sometimes used by the investigators to identify a specific piece of equipment. MAC addresses are, however, unreliable tags in both wireless and wired access networks as they can be easily reconfigured by the users. Similarly, numbers uniquely identifying mobile phones such as: Electronic Serial Number (ESN), currently mainly used with CDMA phones; and, International Mobile Station Equipment Identity (IMEI) number, used with all GSM phones, can be reconfigured by reprogramming erasable programmable read only memory (EPROM) cells where these numbers are stored [10].

All of these factors imply an urgent need for the development of new techniques that will focus on an extraction of characteristics of mobile devices that are consistent,

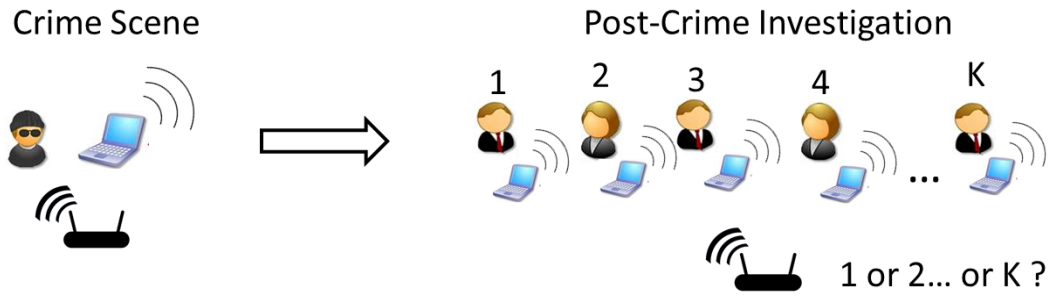

Figure 1.2. Post-crime device identification scenario.

trackable and hard to alter by the users. These characteristics become unique *fingerprints* that can be used to identify wireless transmitters. Identifying the source of an emitted signal is a long existing research topic, especially in military applications, where finding the source of a radar signal is of high interest [11], [12]. Recently, rapid progress of wireless communication technologies, with their inherent security threats due to the shared medium, has amplified the importance of radiometric identification of wireless transmitters.

Although the application space of wireless device identification is broad, as discussed later, here we focus in on its application in criminal investigations. In particular, unique radiometric *fingerprints* of wireless transmitters can be used by law enforcements after a crime to determine the device used in the commission of the crime, thereby significantly decreasing the anonymity level of mobile cyber-cryminals and aiding efficient identification of the offenders.

In particular the identification methods that we develop can be applied to test devices from a pool of suspects in order to decide which one was most likely used while the crime was committed when high-layer identification mechanisms fail or are not implemented. The only print from the crime scene is a signal record captured from the wireless transmitter by an access point. Having this record and records from a group of devices that have been potentially used to commit the crime, the

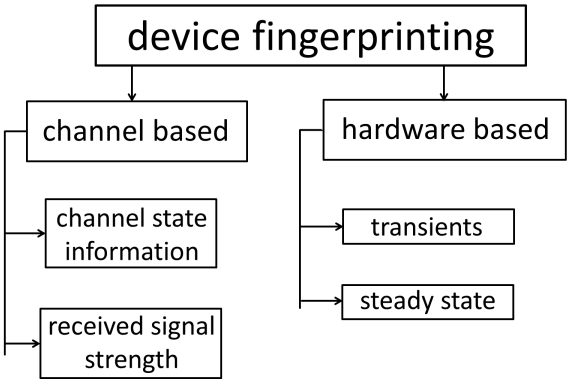

Figure 1.3. Simplified classification of wireless devices fingerprinting methods.

proposed methods can be used to successfully indicate the offender's device with a given probability.

Figure 1.2 presents the considered post-crime device identification scenario. A signal record is captured by a receiver from a criminal's device at the crime scene. After the crime is committed, using the same receiver, records are captured from a group of devices that might have been used to commit the crime. The goal is then to tie transmissions from the crime scene to other transmissions from that same device. Being able to indicate a device that, with a given probably, was used to commit the crime can then allow law enforcement to reduce the size of the original group of suspect devices and to justify applications for warrants, which can then lead to final identification decisions and possible arrests based on the digital content of the devices. The reduction of the size of the original pool of suspect devices can lead to a significant reduction of investigation cost and time. The considered identification scenario is analogous to classical ballistics testing used for crimes involving the discharge of a weapon.

There have been a number of wireless device fingerprinting efforts over the years, a simplified classification of which is visualized in Figure 1.3. They can roughly

be divided into two main groups. First there are approaches that exploit channel information. In a rich multi-path environment, because of rapid path decorrelation, users can be almost uniquely characterized by their channel conditions. This property allows for grouping transmissions from stationary wireless users [13–16]. Another channel-based fingerprinting technique uses received signal strength information to distinguish transmitters [17, 18]. Both of these techniques make a strong assumption on users' stationarity, which makes them unapplicable in many practical scenarios. Moreover all channel-based methods are temporal in nature.

The second class of techniques includes fingerprinting approaches that exploit hardware imperfections of RF devices. At the physical layer, despite decades of significant efforts by the microwave circuits community, there still exist longstanding imperfections in the RF portion of the wireless transmitter, which can become manifest by the use of the appropriate signal processing of the signal received. Furthermore, since these cannot be altered by the user without significant effort, they can be exploited to group together signals from one radio. There have been a number of efforts over the years to utilize hardware imperfections for the purpose of distinguishing wireless devices. Much of the work has focused on the detection and analysis of transients [19–26]. A transient is a brief radio emission produced while the power of the output of an RF transmitter goes from zero to the level required for data communication. Transient durations range from a few microseconds to tens of milliseconds. Their nature is such that they are difficult to detect and to describe in a succinct way. Moreover, to extract device fingerprints from the transient emissions, they need to be digitized at extremely high sampling rates. This is necessary to provide the granularity of amplitude information required for the transient feature extraction algorithms. For example a 5 GSamples/s sampling rate is reported by Serinken et al. [20, 22], and, a 500 MSamples/s sampling rate is used by Hall et al. [21]. Therefore, in addition to transients analysis, a significant amount of research has been conducted on device

identification based on hardware fingerprints extracted from steady-state transmissions (protocol-controlled transmissions after the effects of turn-on transients die off). Kohno et al. [27] proposed extraction of clock skews as unique parameters characterizing physical devices of Internet users from steady-state transmissions. Tomko et al. [28] built histograms of features, such as received power and frequency error, over a set of packets from each network user. Then they extracted user fingerprints as parameters used to fit the histograms to a Gaussian model and used these parameters for device identification. Gerdes et al. [29] showed that the matched filtering of received steady-state transmissions can be reliably used to build signal profiles that can be used to discriminate between Ethernet cards of different models. Brik et al. [30] and later Candore et al. [31] followed the hardware fingerprinting approach and used machine learning techniques on collected steady-state modulation data to train data-agnostic classifiers that are then able to distinguish wireless cards, even when produced by the same vendor [30]. Nguyen et al. [10] followed the approach of Brik, but in contrast to Brik et al. employed unsupervised learning techniques, which do not require the training phase. In [32] Kennedy et. al considered similar classifiers to the ones used by Brik et al., but exploited spectral features of the steady-state signals received.

Our fingerprinting approach, similarly to [10, 27, 28, 30–32], falls in the field of methods that exploit differences in transmitters' hardware extracted from steady-state transmissions. A simplified block diagram of a wireless transmitter is shown in Figure 1.4. Each of the components of the transmitter chain demonstrates imperfections caused by nonidealities of the production processes. MOS transistors, which the components' circuits are built of, exhibit variations in major device parameters (e.g. channel length, channel doping concentration, oxide thickness) among production lots. These variations may occur for many reasons, such as minor changes in the humidity or temperature in the clean-room, or due to the position of the die relative to the

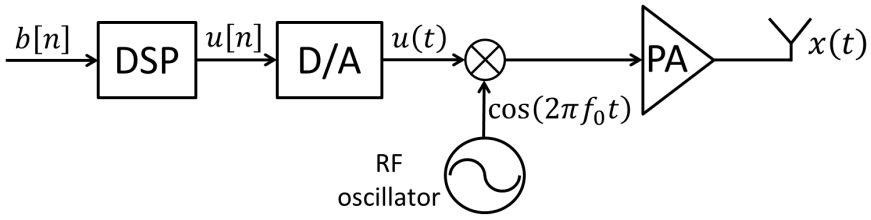

Figure 1.4. Basic components of a wireless transmitter, the imperfections of which can be exploited for user identification. $b[n]$ is the sequence of bits to be transmitted, $u[n]$ and $u(t)$ are the digital and analog baseband waveforms, respectively, and $x(t)$ is the transmitted signal up-converted to a carrier frequency f_c.

center of the wafer. Changes of the parameters influence transistors switching speed and thereby components' characteristics. Similarly, parameters of passive electronic devices, rather than taking a constant specified value, follow distributions caused by production inaccuracies. Despite technological advancements, constant market push for low-price, high-volume products results in variations among individual devices caused by the production imperfections. These variations, while being small enough for the devices to meet the specifications of communication standards, are significant enough to allow for unique characterization of these devices via RF fingerprints.

In this dissertation we consider two attacker models. In Chapter 2 we propose identification of devices, for which the adversary user is capable of modification of the higher layer tags such as the IP address, MAC address, ESN, IMEI number. Such an attack is further referred to as a *weak adversary* attack. In Chapter 3 we introduce device identification method that allows for successful identification of devices that are in use by *strong adversaries*, who in addition to modification of higher layer tags are capable of injecting slight digital distortions to physical data symbols, before the digital signal is exposed to transmitter's nonlinearities, which while allowing for reliable data transmission, changes the character of the total distortion observable at the receiver. In Chapter 4 we study an identification method based on unique signatures extracted from output of RF oscillators. Because of its data-independent

character, the method from Chapter 4 can be applied to identify devices of *strong adversaries*.

1.1.2 Innovation

In this dissertation, we focus on device identification based on hardware imperfections extracted from steady-state transmissions and propose a novel approach to device modeling, algorithm design and anonymity analysis that is significantly different from prior efforts on wireless device identification. In particular, in contrast to the recent empirical classification results of Brik et al. on commercial 802.11 cards, our proposed approach focuses on a comprehensive understanding of the phenomena being exploited for the node identification. Our work can be viewed as advancing the understanding of how mobile communications exposes the degree to which information can be collected or inferred about individuals and can play an important role in understanding the limits of personal privacy when interacting with digital devices. We consider individual components of the transmitter chain to gain an understanding about which of the components can play a dominant role in device identification. In Chapter 2, for power levels specified by the manufacturers as compliant with spectral masks of commercial communication standards, we establish that variations among power amplifiers (PAs) are significant enough for successful device identification at practical signal-to-noise ratio levels and dominant when compared to the variations among digital-to-analog converters (DACs). In Chapter 4, we propose an identification method based on RF oscillator imperfections. This is motivated by the fact that, in contrast to the PAs, which for transmit power controlled applications might be switching power modes over time, characteristics of the RF oscillators are power level independent and thus can be used as unique device tags in systems with implemented transmit power control mechanisms.

The identification methods proposed in this dissertation are based on variations of features of the transmitter chain components that are difficult to modify by the user, such as the integral nonlinearity of the DACs, the nonlinear gain of the PAs, and the phase noise of the RF oscillators. This makes the proposed identification methods more robust against potential attacks of sophisticated users than, for example, the methods of Brik et al. [30]. In particular, in [33] Danev et al. report thwarting the identification methods of Brik et al. [30] with a success rate close to 100% with a simple adjustment of the carrier frequency of the masquerading device and with digital modifications of constellation symbols. Our proposed identification methods, which exploit the production variations present in the components of wireless devices, are based on statistical models amenable for analysis, which allow us to identify the key device characteristics that cause anonymity loss, develop countermeasures that can be applied by the nodes to regain the anonymity, and then thwart such countermeasures. In particular the model-based approach introduced in this dissertation allows for development of spectral identification methods that allow for separation of the two possible sources of distortion: modification of the digital data symbols by sophisticated users and inherent transmitter nonlinearities. These methods, which are described in Chapter 3, allow for successful identification of the devices even if the digital symbols are actively modified by the strong adversaries in order to fake the device's fingerprint, as described by Danev et al. [33].

1.2 Interference Robust Wide-Band Receiver

1.2.1 Motivation and Related Work

The bandwidth of wireless sensing and communication systems is increasing rapidly to support emerging applications, including cognitive and software radio [34], environmental sensing [35], vehicle surveillance [36], and multi-function radios [37]. Hence, future receivers will need to process an enormous amount of bandwidth to be effective.

Wide-band cognitive radio (CR) is a wireless communication concept that aims for an efficient use of spectral resources through dynamic spectrum management [34]. Unlike conventional wireless transceivers, operating in pre-allocated sub-bands, wide-band CRs need to support any momentarily unoccupied sub-bands in a wide frequency range of interest. As an example of a practical application of CR, consider the recently developed IEEE 802.22 standard [38] that was aimed at using CR for opportunistic transmissions in a 50-700MHz frequency band that became sparsely allocated [39] after the switchover to digital television in the United States in June 2009. In the conventional approach to receiver design, the RF circuitry downconverts the signal to baseband, the analog-to-digital converter (ADC) converts it to quantized samples, and then a digital signal processor (DSP) extracts the desired information. However, an ideal architecture for an opportunistic cognitive radio receiver would be a wide-band low-noise-amplifier (LNA) and an analog-to-digital converter directly following the receiving antenna. A digital processor would then process the output of the ADC in order to extract information from a dynamically assigned channel of interest. Such an architecture, while appealing with its flexibility, poses significant implementation challenges. First, the complexity and power consumption of analog-to-digital converters increases significantly at large bandwidths (and certainly at extreme bandwidths), as, in accordance to the Nyquist-Shannon sampling theorem, the sampling rate needed for signal recovery needs to be at least twice the bandwidth of the considered band. Secondly, because of the large bandwidths employed, in-band interference is nearly always present and can severely limit the ability of the receiver to resolve small signals of interest. In particular, when a small signal of interest is received with a large in-band interferer that is orders of magnitude larger, the receiver's RF front-end might be forced into a nonlinear range. Nonlinear distortion in the first stages of the hardware causes large interferers to corrupt lower-level signals before the interferers can be de-selected, even if the interferers occupy frequencies different than

the message frequencies. Conventional filtering after the front-end cannot remove this distortion due to the presence of the nonlinearity. Hence, interference rejection in wide-band receivers has received a significant interest from both the circuits and systems communities.

Due to the mentioned hardware complexity and high power consumption at large bandwidths, a number of innovative approaches for signal acquisition for wide-band radios have recently emerged, including a class based on compressive sensing (CS) [40–42]. Of particular interest is a family of 'analog-to-information' converters [43–55]. Compressive sensing allows for acquisition of sparse signals at sampling rates significantly lower than the Nyquist rate required for bandlimited signals. In CS approaches, the full signal bandwidth is not converted, hence avoiding the costly hardware; rather, prior knowledge of a concise signal model allows the recovery to focus only on signal aspects relevant to feature extraction. In particular, if there exists a basis in which a signal of interest can be represented sparsely (i.e., it can be fully characterized with a small number of coefficients), then it is possible to obtain all information needed for successful reconstruction of the signal from a relatively small number of randomized incoherent measurements [56]. This number is often much smaller than the number of samples implied by the Nyquist sampling rate for representation of all bandlimited signals.

Compressive sensing has the potential to transform wide-band receiver design by significantly simplifying analog-to-digital conversion in the presence of background noise. However, the reception of weak signals in the presence of in-band interference has so far been largely ignored. A few recent efforts in the compressive sensing community have acknowledged the problem and proposed solutions [48, 57–61], but these projection-based techniques are based on unrealistic linearity and dynamic range assumptions in the front-end. In particular, when a small signal of interest is received with a large in-band interferer orders of magnitude larger, and when the receiver's

RF front-end is forced into the nonlinear range, the models of [48, 57–61] become inadequate. Hence, current compressive sensing solutions leave a critical aspect of robust wide-band receiver design unresolved.

1.2.2 Innovation

In this dissertation, motivated by the observations of Jackson [5] made on circuits (devices are linear for small amplitudes) and compressive sensing (randomized sub-sampling is an effective acquisition approach for frequency-sparse signals), we study recovery methods for wide-band receivers when the acquisition process only selects samples that preserve the receiver's linearity. Such selective sampling causes a violation of the core assumption of CS: the randomness of the sampling times. This violation implies a need for new recovery guarantees on the number of measurements needed for successful recovery of sparse signals. Most CS contributions assume independent randomness in the measurement projections, and this is exploited to derive bounds on the number of projections required. In Chapter 5, we study the requirements on the number of measurements needed for successful recovery of sparse signals when, instead of individually random measurements, the measurements are taken uniformly at random in pre-defined, non-overlapping *groups* of equal size. For such a measurement scheme we derive bounds on the number of measurements needed for successful recovery, similar to the performance bounds of conventional compressive sensing [56]. In particular we introduce a metric that upper bounds the multiplicative penalty on the number of required measurements introduced by grouping with respect to the conventional CS. While the introduced metric cannot currently be evaluated in a closed form, we employ a computationally feasible method that provides lower and upper bounds on its value. We also evaluate via simulations the penalty predicted by the proposed metric.

One can relate the random, *grouped* measurement scheme to the application of the interference-robust, wide-band receiver by considering an in-band, powerful and known interferer that saturates the receiver's front-end and is uncorrelated with the message of interest that needs to be recovered. In such a scenario undistorted samples can only be taken at times when the interferer's amplitude values are small. Therefore the structure of the sampling scheme used for undistorted recovery of the message of interest is dictated by the powerful uncorrelated interferer. Depending on the character of the interference (modulation technique, periodicity), these interference-driven, constrained sampling can lead to a *grouped* sampling scheme, which is explained in more detail in Chapter 5 (Section 5.1).

Other applications, for which the samples can be drawn in pre-defined groups include medical imaging and remote sensing applications, where it might be difficult and costly to move the sensors randomly to different positions. For such applications it is common for the sensors to follow pre-defined trajectories during the acquisition process. Using such sampling trajectories clearly introduces structure into the measurement process and hence violates a key assumption underlying the standard analysis of CS schemes. Application of *grouped* sampling for medical imaging and remote sensing is discussed in more detail in the Problem Statement of Chapter 5 (Section 5.1)

In Chapter 5 of this dissertation, we study the recovery of sparse signals from limited sample sets, where the constraints on the sampling scheme are independent from the signal that needs to be recovered (the constraints are dictated either by an uncorrelated powerful interferer or by physical constraints of the acquisition systems). In contrast, in Chapter 6 we study the compressive sensing recovery of frequency-sparse signals from irregular, *signal dependent* samples. In particular, we attempt to recover the signal using samples with values within a given amplitude range $[-\tau, \tau]$. The motivation behind such a study, similar to the study of Chapter 5, is enhancement of the

dynamic range of receivers in a wide-band wireless communications systems. In the case of the study of Chapter 6, in contrast to Chapter 5, both the powerful interferer that saturates the receiver and the weak message of interest need to be recovered from a subset of samples. In Chapter 6 we introduce and characterize through numerical simulations three approaches for the improvement of the performance of recovery of frequency-sparse signals from amplitude-limited sample sets.

In addition to the receiver's nonlinearities, time-truncation of the processed signal records causes degradation of the dynamic range of wide-band receivers. In particular, the spectral content of signals processed in discrete blocks of equidistant samples can be misinterpreted when projected onto a discrete set of equidistant frequencies via a discrete Fourier transform (DFT) operation. Signal components with frequencies off the discrete frequency grid cannot be represented with a single DFT component, which leads to a leakage of the energy of the off-grid signal components among multiple DFT components (see Section 7.1). The misinterpretation of spectral content of a powerful interferer can lead to significant contamination of a message of interest occupying nearby frequency bands and cause degradation of the wide-band receiver's performance.

The effects of spectral leakage can be mitigated by increasing the length of the record block and hence increasing the resolution of the discrete frequency grid. Instead of increasing the block length, which can lead to high complexity of the receiver, in Chapter 7 we propose a method that is based on adaptive partial removal of the cyclic prefix. The cyclic prefix is a copy of the end of the signal block attached in front of the block in order to avoid inter-block interference (IBI) and to enable frequency domain equalization (FDE) of the channel. Because of the dynamic character of wireless channels, in practical designs the length of the CP is usually chosen with a safety margin. Therefore, in many cases, IBI can be avoided even if the CP is partially removed at the receiver. As will be shown in Chapter 7, adaptive, non-complete

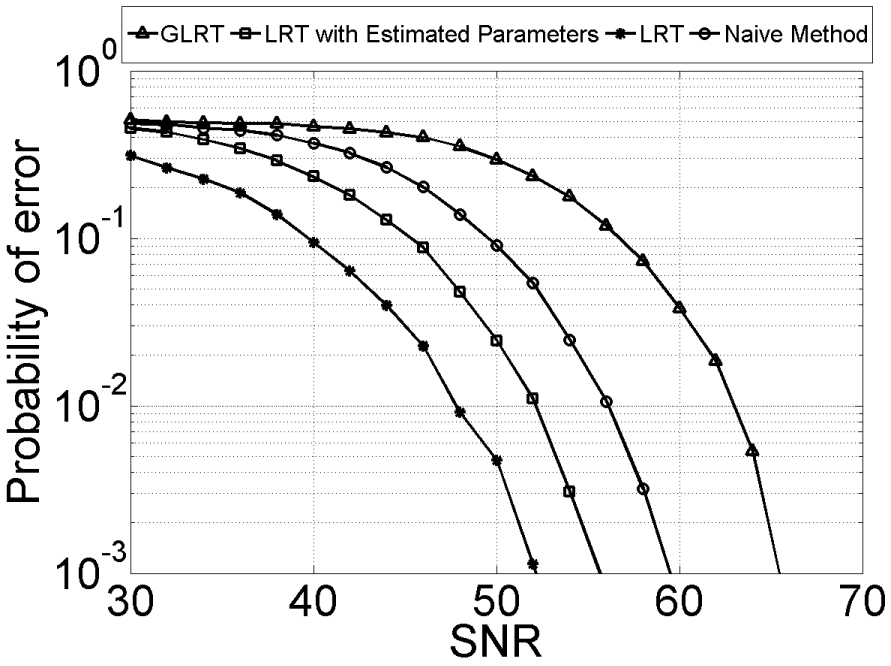

Figure 1.5. Probability of device identification error versus signal-to-noise ratio (controlled with the noise power level), averaged over 100 DAC pairs and over 150 input vectors of size $M = 2500$. Elements of the input vectors were normally distributed with mean value equal to half of the DAC's input range and standard deviation equal to $\frac{1}{6}$ of the DAC's input range (which resulted in 99% of the input values within the input range, values outside the input range were ignored). Standard deviation of normally distributed individual DAC sources was set to $\sigma_S = 0.02$. The first eight eigenfunctions of the Brownian Bridge random process were used for representation of the DAC's integrated nonlinearity, exploited for identification. As visualized in this plot high signal-to-noise ratio levels were needed to achieve a probability of error of 10^{-3}.

removal of the CP allows for choices of block length that can lead to a significant reduction of the spectral leakage into the frequencies occupied by a message of interest. This leads to significant improvement of the dynamic range of the wide-band receivers.

1.3 Proposed Contributions

The following contributions to the existing literature are claimed in this dissertation:

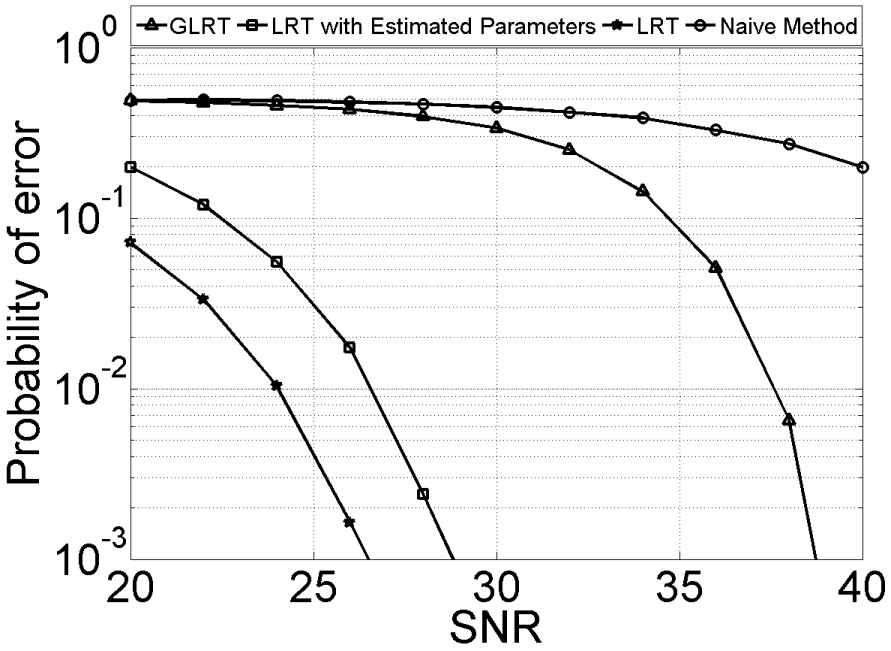

Figure 1.6. Probability of device identification error for a measured pair of MAXIM MAX2242 amplifiers versus signal-to-noise ratio (controlled with the noise power level), averaged over 25000 input vectors of size $M = 2500$, with standard deviation σ_x of the normally distributed elements of input vectors equal to 0.055. For $\sigma_x = 0.055$ the probability that the input signal exceeded the upper level of linear range of the considered amplifiers [3] was only 1%. Whenever the input signal exceeded the the linear range, it was clipped to its upper level. As visualized in this plot, relatively low signal-to-noise ratio levels were needed for a probability of error of 10^{-3}. Such levels are common in practical WLAN deployments.

- Development of model-based device identification methods that provide insight on how individual components of the transmitter chain (digital to analog converter (DAC), power amplifier (PA) and RF oscillator) contribute to the anonymity loss of wireless devices, due to the hardware imperfections caused by production process non-idealities. Figures 1.5 and 1.6 and Table 1.1 show the performance of the proposed identification methods when imperfections of, respectively, DACs, PAs and RF oscillators were exploited for breaking device's anonymity.

PLL	1	2	3	4	5	6	7	8
1	-	0.000	0.000	0.000	0.000	0.000	0.016	0.000
2	0.000	-	0.000	0.164	0.000	0.000	0.000	0.000
3	0.000	0.000	-	0.000	0.216	0.000	0.000	0.000
4	0.000	0.228	0.000	-	0.00	0.00	0.00	0.00
5	0.000	0.000	0.228	0.000	-	0.000	0.000	0.020
6	0.000	0.000	0.000	0.000	0.000	-	0.000	0.008
7	0.008	0.000	0.000	0.000	0.000	0.000	-	0.000
8	0.000	0.000	0.000	0.000	0.028	0.036	0.000	-

Table 1.1. Probability of device identification error averaged over 250 trials for all possible pairs from the group of 8 measured oscillators for the test (4.20) at $SNR = 15$dB (lower left, below the diagonal) and at $SNR = 35$dB (upper right, above the diagonal), when all 50 captured records of length $12.5 \cdot 10^6$ samples were used to extract the fingerprints of the devices from the pool of suspects, and a single record from the crime scene, randomly chosen from the group of all 50 captured records, was used for identification. For example the probability that the device 8 is mistaken for the device 5 is 0.02 when $SNR = 35dB$. Because of the differences between the identification methods employed to exploit imperfections of RF oscillators and methods employed to exploit imperfections of the PA's and DAC's (compare Sections 4.3 and 2.3), the processing of much longer vectors is possible here; hence, the very good signal-to-noise ratio (controlled with the noise power level) performance reported with this table is possible.

- Design and evaluation of countermeasures that can be applied by strong adversaries to regain the anonymity despite the hardware imperfections and the development of identification methods that can thwart such countermeasures. Figure 1.7 shows the performance of a proposed spectral identification test, when the masquerading users are: *i*) not faking their RF signatures; *ii*) faking their RF signatures by distorting digital data symbols; *iii*) faking their RF signatures by distorting digital data symbols, but the proposed identity fake thwarting identification methods from Chapter 3 are applied. In addition, for comparison Figure 1.7 shows the performance of the time-domain based methods from Chapter 2.

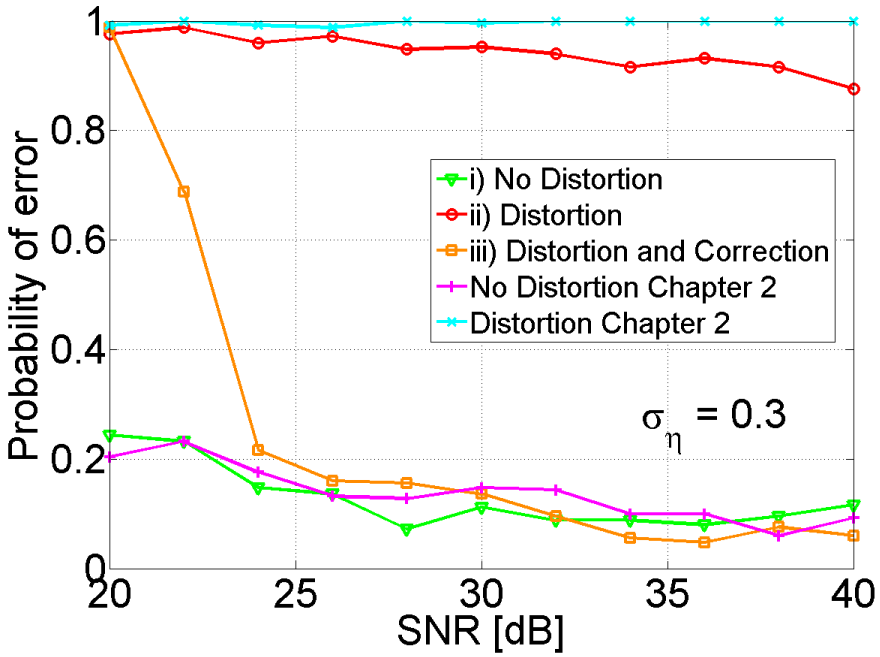

Figure 1.7. Probability of erroneous identification decision (3.14), calculated over 250 randomly generated groups of 3 power amplifiers and input signals, as a function of SNR (controlled with the noise power level), for standard deviation $\sigma_\eta = 0.3$ of the zero-mean, normal random variable η (3.16), for 50 signal records of length 1024 symbols, captured from the device used to commit a crime, and for 500 signal records of length 1024 undistorted symbols captured from the three suspected devices, for the three cases **i)**, **ii)** and **iii)**, described in Section 3.4, together with the performance of the time domain based methods of Chapter 2. As visualized in this plot, the spectral identification method of Section 3.3 allows for identification of devices of sophisticated users with effectiveness similar to the effectiveness of identification of devices of users that do not apply countermeasures to regain their anonymity.

- Derivation of bounds on the number of compressive measurements needed for successful recovery of sparse signals when the random projection measurements are structured into pre-defined groups. We introduce a metric γ (5.4) that bounds from above the multiplicative penalty on the number of required measurements introduced by grouping, with respect to conventional compressive sensing acquisition employing independent random measurement selection.

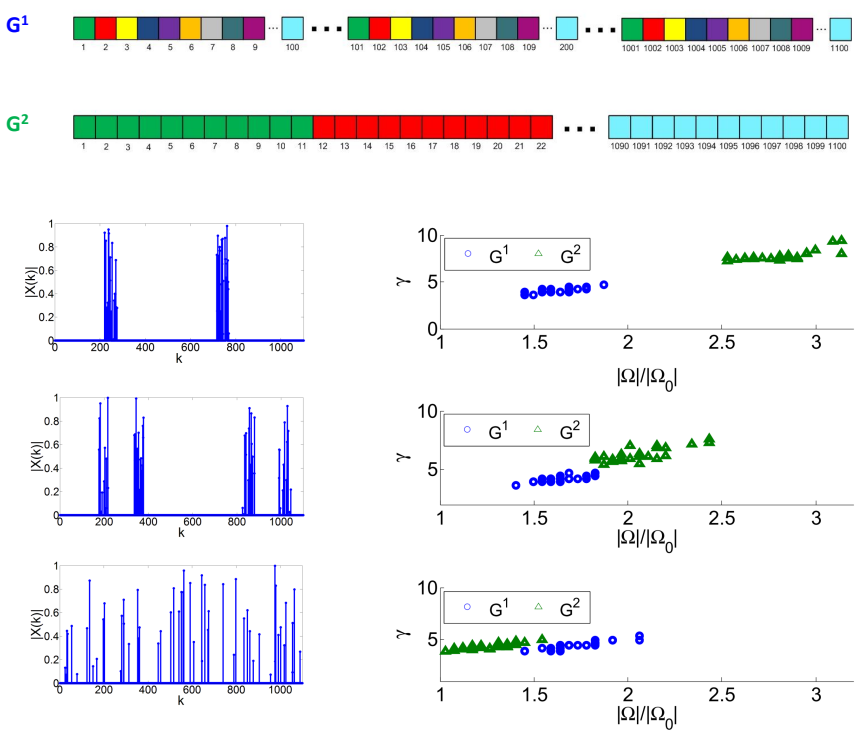

Figure 1.8. γ versus M/M_0 for group structures G^1 and G^2 when Fourier coefficients of 5% sparse signal s were concentrated within (top): a sub-band built out of two 5%-wide channels, (middle): a sub-band built out of four 5%-wide channels, (bottom): the entire band. As visualized in this figure, the introduced penalty parameter γ can be a good performance indicator for many practical scenarios, which are further discussed in Section 5.3.

Figure 1.8, visualizes two exemplary group structures for grouped sampling of a one-dimensional signal. For both structures, groups are represented by different colors. For the two group structures, for three different classes of signal supports, Figure 1.8 shows the relationship between the derived penalty factor γ (5.4) and the ratio of the number of grouped measurements M required for successful recovery to the number of measurements M_0 required for successful recovery when individual measurements are not taken in groups, but uniformly at random.

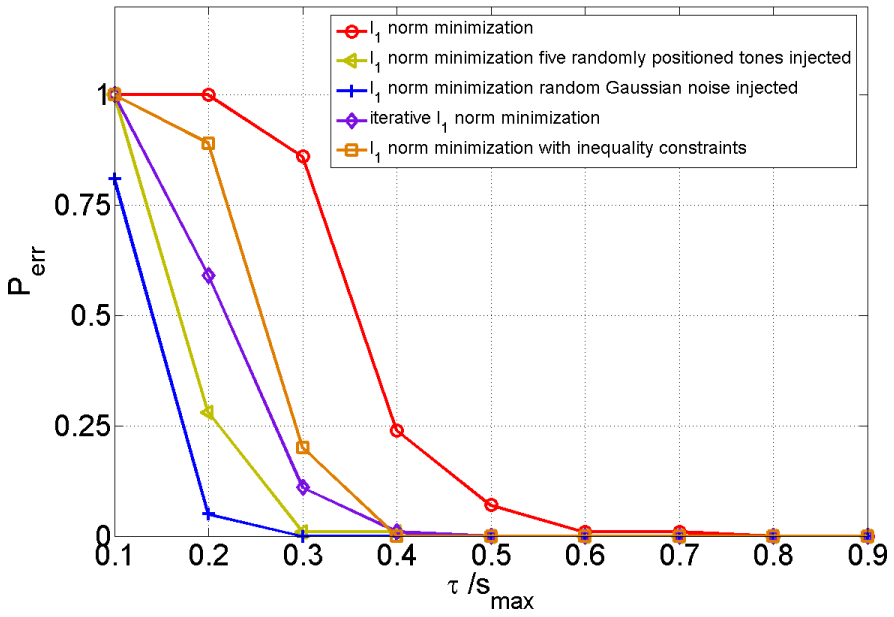

Figure 1.9. Probability of signal recovery error, defined as the probability that the normalized recovery error (6.2) is above 3%, for ℓ_1-norm minimization (6.1), for iterative ℓ_1-norm minimization (6.6) with five iterations, for constrained ℓ_1-norm minimization (6.4) as well as probability of recovery error for ℓ_1-norm minimization (6.1) for different types of known injected interferers, as a function of the threshold τ normalized to the maximal amplitude of the signal s_{max}. Only samples with amplitudes below τ are used for signal recovery. As visualized in this plot, the proposed reconstruction methods allow for significant reduction of the value of τ, when compared to conventional compressive sensing, which can then lead to significant reduction of the nonlinear distortion of the signal received.

- Development of methods for reconstruction of a frequency sparse signal from small-amplitude samples. Such samples that preserve linearity of the receivers' front-end are obtained by only sampling when the amplitude is below a given threshold τ. The potentially nonlinearly distorted samples with amplitude values that exceed the threshold are discarded. We consider techniques that substantially improve recovery performance when compared to conventional recovery methods of compressive sensing. Figure 1.9 shows recovery performance improvement for the proposed methods.

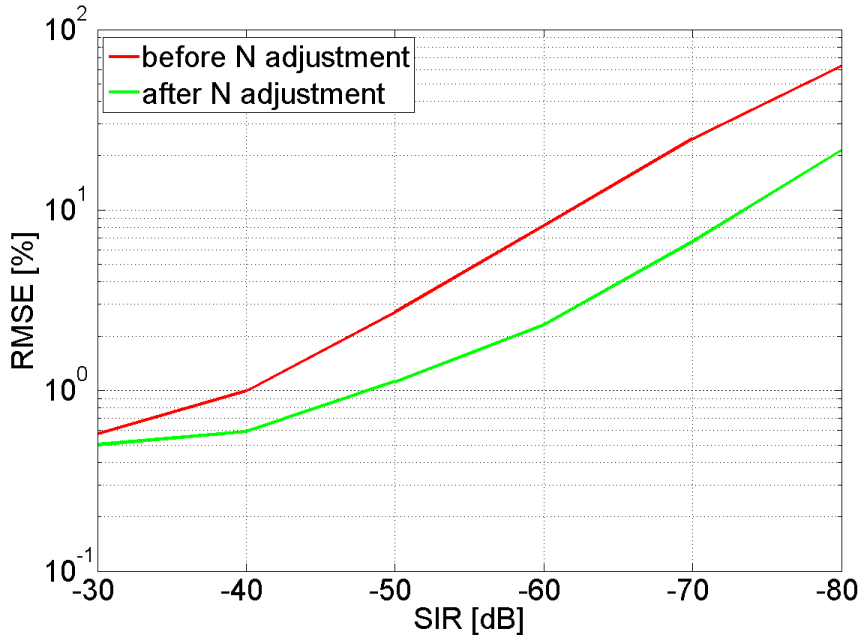

Figure 1.10. RMSE of the recovered QPSK message symbols before and after the adjustment of the processing block size N with (7.8) as a function of the SIR for a fixed guard bandwidth from (7.9) W_G=50MHz. As visualized in this plot, dynamic range improvement of over $10dB$ can be achieved with adaptive adjustment of the block processing size proposed and studied in Chapter 7.

- Development of an approach for mitigation of spectral leakage, caused by time-truncation of processed signal records, via an adaptive choice of processing window size for block processing, single-carrier CR receivers. The developed method is based on an adaptive partial removal of cyclic prefix, which is a copy of the end of the signal block attached in front of the block in order to avoid inter-block interference. The method can be applied in environments with maximal channel delay paths shorter than the length of the cyclic prefix. It does not require any structural changes to the receiver and allows for significant dynamic range improvements (over 10 dB), which is shown in Figure 1.10.

The mobile device identification study presented in this dissertation, mostly for the purposes of digital forensics and post cyber-crime investigations, can also help to

inform the recently emerged debate about the level of personal privacy. The existence of global digital surveillance practices, origins of which can be traced back to the middle of 20^{th} century [62], have not been widely acknowledged until the recent unveiling of active global surveillance programs run by U.S. government with a cooperation of telecommunication companies and European governments [63]. The uncovering of these global surveillance programs triggered a vigorous world-wide discussion about the level of personal privacy. It raised concerns about potential misuse of the surveillance practices and questions about surveillance methods that would allow modern intelligence agencies to achieve their aims without violating privacy. The study conducted and reported in this dissertation allows for a better understanding of the extent to which personal privacy can be decreased by interaction with commercial communication devices. In the post cyber-crime investigation application considered here, the transmitter for the criminal may be quite sophisticated, as significant physical and technology resources may be available. However, when considering the potential loss of privacy of many users with cost-constrained mass-produced consumer products, the ability of such users to actively thwart the techniques described here is likely quite limited. Hence, the work presented here can play an important role when trying to answer questions about the degree to which hardware-based RF fingerprinting can compromise individuals' privacy.

CHAPTER 2

IDENTIFYING WIRELESS USERS VIA TRANSMITTER IMPERFECTIONS

2.1 Problem Statement

In this chapter, we study model-based methods for identification of wireless devices. In particular we concentrate our attention on two components of the transmitter's chain from Figure 1.4: the digital-to-analog converter and the power amplifier. Exploitation of nonidealities of the RF oscillators for device identification is studied in Chapter 4.

Consider the basic block diagram of a wireless transmitter shown in Figure 1.4 of Chapter 1. A digital (discrete-time, discrete-amplitude) baseband signal $u[n]$ that carries the information bits is generated by a digital signal processor (DSP) and converted to an analog signal $u(t)$ by a digital-to-analog converter. Then it is translated to the desired carrier frequency by the mixer and amplified by the power amplifier. In an ideal system, the transmitted signal would be given by

$$x(t) = Au(t)\cos(2\pi f_0 t + \Theta), \tag{2.1}$$

where A is the gain of the power amplifier, $u(t)$ is the ideal analog form of $u[n]$ (i.e. the $sinc(\cdot)$-interpolated version of $u[n]$), f_0 is the desired carrier frequency, and Θ is the (constant) phase of the oscillator. However, digital-to-analog converters suffer from the finite precision of the digital input and more importantly, particularly for device identification work, demonstrate nonlinearities, which can vary significantly across

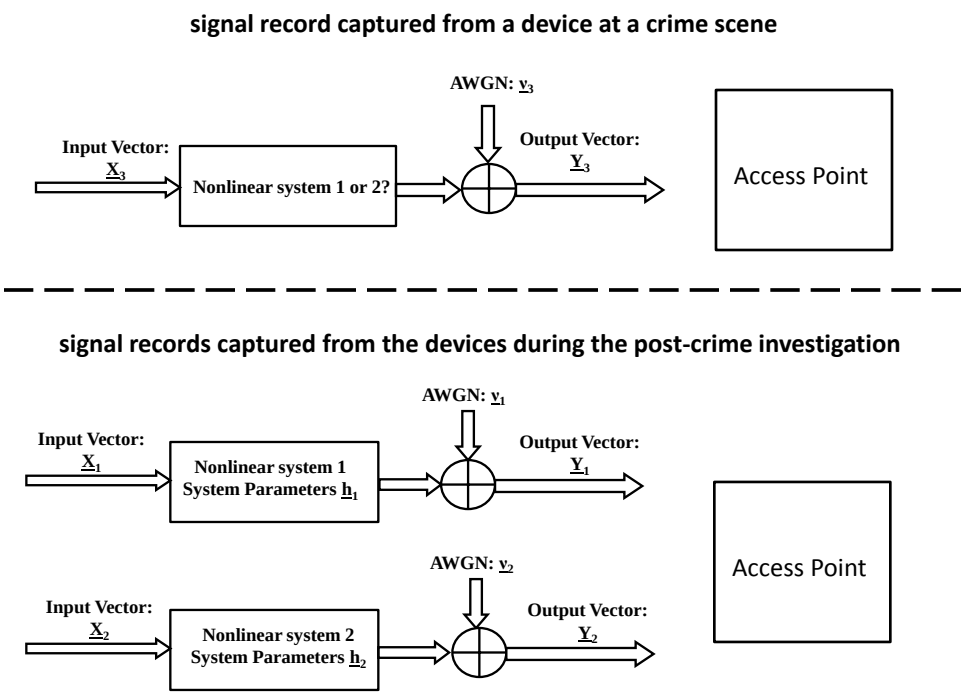

Figure 2.1. The two-system identification scenario: record captured at a crime scene from a the criminal's device (upper part); records captured from two devices during the post-crime investigation (lower part). An additive white Gaussian noise (AWGN) channel model is assumed.

individual units. Similarly, PAs, which seek to have linear characteristics such that the response to input $u(t)$ is $Au(t)$, are often quite nonlinear even with significant compensation. As in case of the DACs, the nonlinearity variations of PAs can be significant across the devices. Because of these variations, each device in a multiple device network can be characterized with a group of parameters that uniquely describe input/output (I/O) characteristics of its transmitter components. These parameters can then be used by access points for the purpose of device identification.

As described in Chapter 1, the goal of the proposed identification methods is to tie incident transmissions of an adversary's device to other, post-incident transmissions of that same device. Under the assumption that the criminal employs his/her true identity during the post-incident investigation, this allows to identify the of-

fending party. To simplify the exposition, this chapter considers a setup shown in Figure 2.1 for the two-device case, but the generalization to the case of n devices is straightforward (using n-hypothesis testing techniques). Each of the users' devices in Figure 2.1 is characterized with a parameter vector $\underline{\boldsymbol{h}}_k$, $k = 1, 2$. Because access points in wireless networks perform inverse operations to all operations performed by the transmitters of wireless devices (demodulation, A/D conversion, decoding, etc.), it is reasonable to assume that input samples (which can be reconstructed from the decoded data) and corresponding noise-corrupted output samples are accessible for all elements of the transmitter chain, if the rest of the chain is assumed linear. In particular, for the DACs the input samples are the elements of $u[n]$ from (2.1) and for the PAs the input samples are samples of $u(t)$ from (2.1). The input vectors of a considered component of device one and device two are further denoted as $\underline{\boldsymbol{X}}_1$ and $\underline{\boldsymbol{X}}_2$, respectively, and their respective output vectors are $\underline{\boldsymbol{Y}}_1$ and $\underline{\boldsymbol{Y}}_2$. These correspond to the case when the users are employing their true device identities, which they assume at some time when they are not committing a crime. Now, at some point in the past, one of the devices was used to commit a crime. The binary hypothesis problem is to identify this device given access to $\underline{\boldsymbol{X}}_k$ and $\underline{\boldsymbol{Y}}_k$, $k = 1, 2, 3$. In other words, it needs to be found out which of the two transmitters vector $\underline{\boldsymbol{X}}_3$ passed through, so that it resulted in $\underline{\boldsymbol{Y}}_3$.

2.2 Modeling Transmitter Components

Both transmitter components considered in this chapter display nonlinearities of their I/O characteristics. In general the I/O characteristic of a given transmitter component, for user k, can be described with a matrix equation of the form

$$\underline{\boldsymbol{Y}}_k = P_k \underline{\boldsymbol{h}}_k + \underline{\boldsymbol{\nu}}_k \quad ; \quad k = 1, 2, \tag{2.2}$$

where P_k is a matrix, elements of which are nonlinear functions of elements of the input vector $\underline{\boldsymbol{X}}_k$. These functions are determined by the model adopted for a given *type* of transmitter component (i.e. DAC or PA) and are the same for all devices of that type (i.e. they do not vary across DACs or across PAs), which we describe in the successive subsections. The $\underline{\boldsymbol{h}}_k$ is a column vector that contains the unique component parameters of user k and $\underline{\boldsymbol{\nu}}_k$ is an additive white Gausssian noise (AWGN) vector with elements $\sim \mathcal{N}(0, \sigma_\nu^2)$. In the next two subsections, we introduce mathematical models that allow for the construction of matrices P_k for both DACs and PAs.

2.2.1 DAC

An N-bit DAC converts an N-bit input word x to one of $n = 2^N - 1$ analog output values. One of the most important parameters of the DAC is the integral nonlinearity (INL). The INL specifies the deviation of the actual DAC's output level for a given input word from the ideal output level and is defined as

$$INL_x = \frac{I_{out,x} - x \cdot I_{LSB}}{I_{LSB}}, \tag{2.3}$$

where $I_{out,x}$ is the output level generated by an input word x, and I_{LSB} is the maximal output level divided by the number of all input words:

$$I_{LSB} = \frac{I_{out,(2^N-1)}}{2^N - 1}. \tag{2.4}$$

The I/O relation of the DAC can thus be expressed as

$$I_{out,x} = (INL_x + x) \cdot I_{LSB}. \tag{2.5}$$

The INL is caused by production inaccuracies that cause output levels of individual analog sources of the DAC to vary around their nominal values. If individual DAC

analog sources are modeled as independent normally distributed random variables with standard deviation σ_S, then the INL of a thermometer-coded DAC, for which all analog sources have identical nominal values and for which each increase of the input word by one causes activation of an additional source, can be modeled with a discrete Brownian Bridge random process BB [64]

$$INL_x = \sigma_s \sqrt{n} \cdot BB\left(\frac{x}{n}\right), \tag{2.6}$$

where recall x is an input word and n is the number of all input words. For a high number of bits N, $n = 2^{N-1}$ becomes very large and the discrete Brownian Bridge random process from (2.6) can be approximated with its continuous counterpart. A continuous Brownian Bridge random process is defined as

$$BB(t) = W(t) - t \cdot W(1); \quad t \in (0, 1), \tag{2.7}$$

where $W(t)$ is a Wiener random process [65]. A Wiener random process takes value equal to zero for $t = 0$ and its increments are normally distributed random variables with variance equal to the argument difference. In other words

$$W(0) = 0$$

$$W(t_2) - W(t_1) \sim \mathcal{N}(0, t_2 - t_1). \tag{2.8}$$

Using the Karhunen-Loeve theorem, a continuous Brownian Bridge random process can be represented with its eigenfunctions and eigenvalues found as solutions of the integral equation [65]

$$\{\varphi_j(t) : j = 1, 2, ...\} \qquad \{\lambda_j : j = 1, 2, ...\}$$

$$\int_0^1 R_{BB}(t,\tau)\cdot\varphi_j(\tau)d\tau=\lambda_j\varphi_j(t);\quad t\in(0,1), \tag{2.9}$$

where R_{BB} is the autocorrelation function of the Brownian Bridge random process. Solutions of this equation are

$$\varphi_j(t)=\frac{\sqrt{2}}{\pi j}\sin(\pi jt)\quad;\quad\lambda_j=1$$

$$j=1,\ldots,\infty \tag{2.10}$$

and the Karhunen-Loeve expansion of the the continuous Brownian Bridge is

$$BB\left(t\right)=\lim_{J\to\infty}\sum_{j=1}^{J}\frac{\Lambda_j}{\pi j}\cdot\sqrt{2}sin\left(\pi\cdot j\cdot t\right), \tag{2.11}$$

where Λ_j are i.i.d. normal random variables $\sim\mathcal{N}(0,\lambda_j)$. After replacing the continuous argument t with x/n, the I/O characteristic of the k^{th} DAC can be described with the matrix equation

$$\frac{\underline{\boldsymbol{I}}_{k,out}-\underline{\boldsymbol{X}}_k\cdot I_{LSB,k}}{I_{LSB,k}}=BB\left(\underline{\boldsymbol{X}}_k/n\right)+\frac{\underline{\boldsymbol{\nu}}_k}{I_{LSB,k}}, \tag{2.12}$$

which, when using the J first eigenfunctions to approximate the process, has form

$$\frac{\underline{\boldsymbol{I}}_{k,out}-\underline{\boldsymbol{X}}_k\cdot I_{LSB,k}}{I_{LSB,k}}$$

$$=\sqrt{2}\begin{bmatrix}\sin(\pi\cdot1\cdot\underline{\boldsymbol{X}}_k(1)/n) & \sin(\pi\cdot2\cdot\underline{\boldsymbol{X}}_k(1)/n) & \cdots & \sin(\pi\cdot J\cdot\underline{\boldsymbol{X}}_k(1)/n)\\ \sin(\pi\cdot1\cdot\underline{\boldsymbol{X}}_k(2)/n) & \sin(\pi\cdot2\cdot\underline{\boldsymbol{X}}_k(2)/n) & \cdots & \sin(\pi\cdot J\cdot\underline{\boldsymbol{X}}_k(2)/n)\\ \vdots & \vdots & \ddots & \vdots\\ \sin(\pi\cdot1\cdot\underline{\boldsymbol{X}}_k(M)/n) & \sin(\pi\cdot2\cdot\underline{\boldsymbol{X}}_k(M))/n & \cdots & \sin(\pi\cdot J\cdot\underline{\boldsymbol{X}}_k(M)/n)\end{bmatrix}\cdot\begin{bmatrix}h_k(1)\\ h_k(2)\\ \vdots\\ h_k(J)\end{bmatrix}+\frac{\underline{\boldsymbol{\nu}}_k}{I_{LSB,k}}, \tag{2.13}$$

where M is the length of the input sequence. Eq. (2.13) with

$$\underline{\boldsymbol{Y}}_k=\frac{\underline{\boldsymbol{I}}_{k,out}-\underline{\boldsymbol{X}}_k\cdot I_{LSB,k}}{I_{LSB,k}} \tag{2.14}$$

and

$$P_k = \sqrt{2}\begin{bmatrix} sin(\pi \cdot 1 \cdot \underline{\boldsymbol{X}}_k(1)/n) & sin(\pi \cdot 2 \cdot \underline{\boldsymbol{X}}_k(1)/n) & \cdots & sin(\pi \cdot J \cdot \underline{\boldsymbol{X}}_k(1)/n) \\ sin(\pi \cdot 1 \cdot \underline{\boldsymbol{X}}_k(2)/n) & sin(\pi \cdot 2 \cdot \underline{\boldsymbol{X}}_k(2)/n) & \cdots & sin(\pi \cdot J \cdot \underline{\boldsymbol{X}}_k(2)/n) \\ \vdots & \vdots & \ddots & \vdots \\ sin(\pi \cdot 1 \cdot \underline{\boldsymbol{X}}_k(M)/n) & sin(\pi \cdot 2 \cdot \underline{\boldsymbol{X}}_k(M))/n & \cdots & sin(\pi \cdot J \cdot \underline{\boldsymbol{X}}_k(M)/n) \end{bmatrix} \quad (2.15)$$

is the I/O equation introduced with (2.2). Elements of the vector $\underline{\boldsymbol{h}}_k$ are realizations of J random variables $\frac{\Lambda_j}{\pi j}$ (eigenvalues of the Brownian Bridge random process) that uniquely describe the single INL path of user k. Because of the assumption that the component's input and output signals $\underline{\boldsymbol{X}}_k$ and $\underline{\boldsymbol{Y}}_k$ are known and because of the existence of a fixed model for the I/O relation, the only parameters that the receiver needs to estimate to build the digital signature of a device k are elements of the vector $\underline{\boldsymbol{h}}_k$. Algorithms and numerical results for the adopted models of the transmitter components are reserved for Sections 2.3 and 2.4, respectively.

2.2.2 Power Amplifier

Imperfections of power amplifiers are attractive features for device identification purposes in that the PAs are the last elements of the transmitter chain and thus are the most difficult for a user to modify via software or even baseband control. In this work, the nonlinear characteristics of power amplifiers are modeled with Volterra series representations, as well-established in the microwave literature (see Chapter 4 of Wambacq and Sansen [66]). For the sake of exposition, we use a Volterra series representation with a memory of one and an order of two (a linear quadratic system), but the extension to higher memory and orders is straightforward. Hence the I/O relation is

$$\begin{aligned}\underline{\boldsymbol{Y}}_k(n) &= \sum_{l_1=0}^{1} h_{k,1}(l_1)\underline{\boldsymbol{X}}_k(n-l_1) + \sum_{l_1=0}^{1}\sum_{l_2=0}^{1} h_{k,2}(l_1,l_2)\underline{\boldsymbol{X}}_k(n-l_1)\underline{\boldsymbol{X}}_k(n-l_2) + \underline{\boldsymbol{\nu}}_k(n) \\ &= h_{k,1}(0)\underline{\boldsymbol{X}}_k(n) + h_{k,1}(1)\underline{\boldsymbol{X}}_k(n-1) + h_{k,2}(0,0)\underline{\boldsymbol{X}}_k^2(n) \\ &+ h_{k,2}(1,1)\underline{\boldsymbol{X}}_k^2(n-1) + h_{k,2}(0,1)\underline{\boldsymbol{X}}_k(n)\underline{\boldsymbol{X}}_k(n-1) + \underline{\boldsymbol{\nu}}_k(n).\end{aligned} \tag{2.16}$$

Similarly as in case of the DAC, the I/O relation of the PA can be characterized with the matrix equation of the form (2.2), with the matrix P_k being built out of the nonlinear functions of the inputs required in (2.16) for an input vector of length M (i.e., the kernels of the Volterra representation) and with the vector $\underline{\boldsymbol{h}}_k$ built out of the Volterra coefficients. Hence the I/O relation has form

$$\underline{\boldsymbol{Y}}_k = \begin{bmatrix} \underline{\boldsymbol{X}}_k(n) & \underline{\boldsymbol{X}}_k(n-1) & \cdots & \underline{\boldsymbol{X}}_k(n-M+1) \\ \underline{\boldsymbol{X}}_k(n-1) & \underline{\boldsymbol{X}}_k(n-2) & \cdots & \underline{\boldsymbol{X}}_k(n-M) \\ \underline{\boldsymbol{X}}_k^2(n) & \underline{\boldsymbol{X}}_k^2(n-1) & \ddots & \vdots \\ \underline{\boldsymbol{X}}_k^2(n-1) & \underline{\boldsymbol{X}}_k^2(n-2) & \ddots & \vdots \\ \underline{\boldsymbol{X}}_k(n)\underline{\boldsymbol{X}}_k(n-1) & \underline{\boldsymbol{X}}_k(n-1)\underline{\boldsymbol{X}}_k(n-2) & \cdots & \underline{\boldsymbol{X}}_k(n-M+1)\underline{\boldsymbol{X}}_k(n-M) \end{bmatrix}^T \cdot \underline{\boldsymbol{h}}_k + \underline{\boldsymbol{\nu}}_k. \tag{2.17}$$

Recall that the *model* is common to all PAs, and hence is known to the receiver. Therefore, with assumed knowledge of the inputs $\underline{\boldsymbol{X}}_k(n), \underline{\boldsymbol{X}}_k(n-1), ..., \underline{\boldsymbol{X}}_k(n-M)$ at the access point (recall that the access point is decoding the data packets of the user), the matrix P_k containing known nonlinear combinations of known inputs is known by the receiver. All that is to be estimated to build the digital signature of device k are the elements of the vector $\underline{\boldsymbol{h}}_k$.

2.3 Algorithms

Having modeled the two nonlinear transmitter components considered in this chapter, next we develop algorithms for solving the hypothesis testing problem stated in Section 2.1. First, we consider the case when the parameter vectors $\underline{\boldsymbol{h}}_1$ and $\underline{\boldsymbol{h}}_2$ of

devices 1 and 2, respectively, are exactly known in order to find an upper bound on the identifiability in the noisy channel. This assumption is also practically reasonable when the devices can be observed over a long period of time that allows for a very accurate parameter estimation or when the parameters are obtained and saved before the transmitters are available on the market. Under this assumption, well-defined, optimal methods are known for solving the hypothesis testing problem. Next, we consider a scenario when the parameters are unknown and only short observations: $\underline{\boldsymbol{X}}_k, \underline{\boldsymbol{Y}}_k$, $k = 1, 2, 3$ are available. In this case, the optimal method is not straightforward and multiple approaches are considered.

2.3.1 Likelihood Ratio Test with Known Parameter Vectors

2.3.1.1 Decision Rule

If the parameter vectors describing the nonlinear aspects of the device's transmitters are exactly known, then, in the case of uniform costs, the probability of error of the receiver is minimized by a likelihood ratio test (LRT). Formally, define the following hypotheses: $\mathcal{H}_1$- the device used to commit the crime is device number 1; $\mathcal{H}_2$- the device used during to commit the crime is device number 2. For equally probable hypotheses, the decision rule for solving the problem presented in Section 2.1 is then

$$\Lambda(\underline{\boldsymbol{Y}}_3) \triangleq \frac{P_{\underline{\boldsymbol{Y}}_3|\underline{\boldsymbol{h}}_1,\underline{\boldsymbol{X}}_3}(\underline{\boldsymbol{Y}}_3|\underline{\boldsymbol{h}}_1,\underline{\boldsymbol{X}}_3)}{P_{\underline{\boldsymbol{Y}}_3|\underline{\boldsymbol{h}}_2,\underline{\boldsymbol{X}}_3}(\underline{\boldsymbol{Y}}_3|\underline{\boldsymbol{h}}_2,\underline{\boldsymbol{X}}_3)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} 1, \tag{2.18}$$

where $P_{\underline{\boldsymbol{Y}}_3|\underline{\boldsymbol{h}}_i,\underline{\boldsymbol{X}}_3}(\underline{\boldsymbol{Y}}_3|\underline{\boldsymbol{h}}_i,\underline{\boldsymbol{X}}_3)$, $i = 1, 2$ are the conditional probability density functions. In the AWGN channel

$$P(\underline{\boldsymbol{Y}}_3 \,|\, \underline{\boldsymbol{h}}_i, \underline{\boldsymbol{X}}_3) = \frac{1}{(\sqrt{2\pi}\sigma_\nu^2)^M} \exp\left\{ -\frac{(\underline{\boldsymbol{Y}}_3 - P_3 \cdot \underline{\boldsymbol{h}}_i)^H (\underline{\boldsymbol{Y}}_3 - P_3 \cdot \underline{\boldsymbol{h}}_i)}{2\sigma_\nu^2} \right\}$$

$$i = 1, 2, \tag{2.19}$$

which allows us to simplify the decision rule (2.18) to

$$||(\underline{\boldsymbol{Y}}_3 - P_3\underline{\boldsymbol{h}}_1)|| \underset{\mathcal{H}_1}{\overset{\mathcal{H}_2}{\gtrless}} ||(\underline{\boldsymbol{Y}}_3 - P_3\underline{\boldsymbol{h}}_2)||. \tag{2.20}$$

2.3.1.2 Algorithm Performance

The probability of error is the probability that the algorithm decides for a different device than the one that was used to commit a crime. In the case of the two-device scenario from Figure 2.1, the probability of error can be expressed as

$$P_e = Pr\{\mathcal{H}_1\} \cdot Pr\{test\ results\ in\ \mathcal{H}_2|\mathcal{H}_1\} + Pr\{\mathcal{H}_2\} \cdot Pr\{test\ results\ in\ \mathcal{H}_1|\mathcal{H}_2\}, \tag{2.21}$$

which, with the assumption of equally probable hypotheses and the symmetry of the problem reduces to

$$P_e = Pr\{test\ results\ in\ \mathcal{H}_2|\mathcal{H}_1\}. \tag{2.22}$$

With (2.20), P_e can be expressed as

$$P_e = Pr\{(\underline{\boldsymbol{Y}}_3 - P_3\underline{\boldsymbol{h}}_1)^H(\underline{\boldsymbol{Y}}_3 - P_3\underline{\boldsymbol{h}}_1) > (\underline{\boldsymbol{Y}}_3 - P_3\underline{\boldsymbol{h}}_2)^H(\underline{\boldsymbol{Y}}_3 - P_3\underline{\boldsymbol{h}}_2)\},$$

which, with

$$\underline{\boldsymbol{Y}}_3 = P_3\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{\nu}}_3 \tag{2.23}$$

under $\mathcal{H}_1$ simplifies to

$$P_e = Pr\left\{\underline{\boldsymbol{d}}^H P_3^H P_3 \underline{\boldsymbol{d}} - \underline{\boldsymbol{d}}^H(P_3^H\underline{\boldsymbol{\nu}}_3) - (P_3^H\underline{\boldsymbol{\nu}}_3)^H\underline{\boldsymbol{d}} < 0\right\}, \tag{2.24}$$

where $\underline{\boldsymbol{d}} = \underline{\boldsymbol{h}}_2 - \underline{\boldsymbol{h}}_1$.

2.3.2 Likelihood Ratio Test with Estimated Parameters

2.3.2.1 Decision Rule

In practical applications, when parameters of the components are unknown and only short input and output vectors $\underline{\boldsymbol{X}}_k$ and $\underline{\boldsymbol{Y}}_k$, $k = 1, 2, 3$ are available, the decision rule from (2.20) with the true parameter values replaced with their estimates is no longer optimal. However, it is still reasonable to use such a rule with the estimated parameters replacing the true values, since the result converges to the optimal rule when the parameter estimates become more and more accurate. The decision rule is then

$$||(\underline{\boldsymbol{Y}}_3 - P_3\hat{\underline{\boldsymbol{h}}}_1)|| \underset{\mathcal{H}_1}{\overset{\mathcal{H}_2}{\gtrless}} ||(\underline{\boldsymbol{Y}}_3 - P_3\hat{\underline{\boldsymbol{h}}}_2)|| \tag{2.25}$$

and the probability of error P_e is

$$P_e = Pr\left\{\hat{\underline{\boldsymbol{d}}}^H P_3^H P_3 \hat{\underline{\boldsymbol{d}}} - \hat{\underline{\boldsymbol{d}}}^H (P_3^H \underline{\boldsymbol{\nu}}_3) - (P_3^H \underline{\boldsymbol{\nu}}_3)^H \hat{\underline{\boldsymbol{d}}} < 0\right\}, \tag{2.26}$$

where $\hat{\underline{\boldsymbol{d}}} = \hat{\underline{\boldsymbol{h}}}_2 - \hat{\underline{\boldsymbol{h}}}_1$ and $\hat{\underline{\boldsymbol{h}}}_1$ and $\hat{\underline{\boldsymbol{h}}}_2$ are estimates of the parameter vectors $\underline{\boldsymbol{h}}_1$ and $\underline{\boldsymbol{h}}_2$.

Interestingly, the decision rule from (2.25) can also be derived in a different way, which further motivates its usage. In particular the receiver can first estimate the parameters of the transmitters ($\hat{\underline{\boldsymbol{h}}}_1$ and $\hat{\underline{\boldsymbol{h}}}_2$) and the parameters of the unit used to commit the crime ($\hat{\underline{\boldsymbol{h}}}_3$). Next the receiver can compare the probability density functions of the estimate $\hat{\underline{\boldsymbol{h}}}_3$ under hypothesis $\mathcal{H}_1$ (parameter vector of the device used to commit the crime is $\hat{\underline{\boldsymbol{h}}}_1$) and $\mathcal{H}_2$ (parameter vector of the device used to commit the crime is $\hat{\underline{\boldsymbol{h}}}_2$) and make a decision based on this comparison. For equally probable hypotheses, the decision rule can be expressed as

$$\Lambda(\hat{\underline{\boldsymbol{h}}}_3) = \frac{P_{\hat{\underline{\boldsymbol{h}}}_3|\hat{\underline{\boldsymbol{h}}}_1}(\hat{\underline{\boldsymbol{h}}}_3|\hat{\underline{\boldsymbol{h}}}_1)}{P_{\hat{\underline{\boldsymbol{h}}}_3|\hat{\underline{\boldsymbol{h}}}_2}(\hat{\underline{\boldsymbol{h}}}_3|\hat{\underline{\boldsymbol{h}}}_2)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{2.27}$$

The estimates $\hat{\underline{\boldsymbol{h}}}_k$ that minimize the squared error

$$e_k = ||\underline{\boldsymbol{Y}}_k - P_k \cdot \underline{\boldsymbol{h}}_k||^2 \tag{2.28}$$

can be found using standard Least Squares (LS)

$$\hat{\underline{\boldsymbol{h}}}_k = (P_k^H P_k)^{-1} P_k^H \underline{\boldsymbol{Y}}_k^H, \quad k = 1, 2, 3. \tag{2.29}$$

Further $\hat{\underline{\boldsymbol{h}}}_3$ given $\hat{\underline{\boldsymbol{h}}}_k$ is

$$\hat{\underline{\boldsymbol{h}}}_3 | \hat{\underline{\boldsymbol{h}}}_k = (P_3^H P_3)^{-1} P_3^H (P_3 \hat{\underline{\boldsymbol{h}}}_k + \underline{\boldsymbol{\nu}}_3) = \hat{\underline{\boldsymbol{h}}}_k + (P_3^H P_3)^{-1} P_3^H \underline{\boldsymbol{\nu}}_3$$

$$k = 1, 2. \tag{2.30}$$

Because $\underline{\boldsymbol{\nu}}_3$ is a Gaussian random vector, so is $\hat{\underline{\boldsymbol{h}}}_3 | \hat{\underline{\boldsymbol{h}}}_k$. Also

$$\underline{\boldsymbol{X}} \sim \mathcal{N}(\underline{\boldsymbol{m}}, C) \quad \Rightarrow \quad A\underline{\boldsymbol{X}} + \underline{\boldsymbol{b}} \sim \mathcal{N}(A\underline{\boldsymbol{m}} + \underline{\boldsymbol{b}}, ACA^H).$$

Thus

$$P_{\hat{\underline{\boldsymbol{h}}}_3 | \hat{\underline{\boldsymbol{h}}}_k}(\hat{\underline{\boldsymbol{h}}}_3 | \hat{\underline{\boldsymbol{h}}}_k) = \frac{1}{\sqrt{det(C)}(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_k)^H C^{-1} (\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_k)}, \tag{2.31}$$

where the covariance matrix C is

$$C = ((P_3^H P_3)^{-1} P_3^H) \cdot \sigma_\nu^2 I_{MXM} \cdot ((P_3^H P_3)^{-1} P_3^H)^H = \sigma_\nu^2 (P_3^H P_3)^{-1}. \tag{2.32}$$

This yields

$$P_{\hat{\underline{\boldsymbol{h}}}_3 | \hat{\underline{\boldsymbol{h}}}_k}(\hat{\underline{\boldsymbol{h}}}_3 | \hat{\underline{\boldsymbol{h}}}_k) = \frac{1}{\sqrt{det(C)}(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma_\nu^2}(\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_k)^H (P_3^H P_3) (\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_k)} \tag{2.33}$$

and (2.27) can then be rewritten as

$$(\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_1)^H (P_3^H P_3)(\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_1) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_2}{\gtrless}} (\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_2)^H (P_3^H P_3)(\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_2), \tag{2.34}$$

which, after substituting $\hat{\underline{\boldsymbol{h}}}_3$ with $(P_3^H P_3)^{-1} P_3^H \underline{\boldsymbol{Y}}_3^H$, is exactly (2.25).

The expected value of the left side of the inequality from (2.26) is

$$\begin{aligned} &E\{\hat{\underline{\boldsymbol{d}}}^H P_3^H P_3 \hat{\underline{\boldsymbol{d}}} - \hat{\underline{\boldsymbol{d}}}^H (P_3^H \underline{\boldsymbol{\nu}}_3) - (P_3^H \underline{\boldsymbol{\nu}}_3)^H \hat{\underline{\boldsymbol{d}}}\} \\ &= \hat{\underline{\boldsymbol{d}}}^H E\{P_3^H P_3\} \hat{\underline{\boldsymbol{d}}} = \hat{\underline{\boldsymbol{d}}}^H R_{P_3} \hat{\underline{\boldsymbol{d}}} = \hat{\underline{\boldsymbol{d}}}^H U \Theta U^H \hat{\underline{\boldsymbol{d}}} = M \sum_{j=1}^{J} \theta_j ||u_j^H \hat{\underline{\boldsymbol{d}}}||^2, \end{aligned} \tag{2.35}$$

where Θ is a diagonal matrix built out of the eigenvalues $\theta_j, j = 1, 2, ..., J$ of matrix R_{P_3} (the covariance matrix of P_3) and U is a matrix built out of the corresponding eigenvectors $u_j, j = 1, 2, ..., J$. Motivated by the transmitted signal in orthogonal frequency division multiplexing (OFDM) systems, elements of the input vectors are assumed to be realizations of independent, zero-mean, normal random variables with standard deviation σ_x. With this assumption, in the case of the considered Volterra representation of nonlinear power amplifiers, described with (2.16),

$$\begin{aligned} \sum_{j=1}^{J} \theta_j ||u_j^H \hat{\underline{\boldsymbol{d}}}||^2 &= \hat{\underline{\boldsymbol{d}}}(1)^2 \cdot \sigma_x^2 + \hat{\underline{\boldsymbol{d}}}(2)^2 \cdot \sigma_x^2 + \hat{\underline{\boldsymbol{d}}}(3)^2 \cdot 2\sigma_x^4 + \hat{\underline{\boldsymbol{d}}}(4)^2 \cdot 2\sigma_x^4 \\ &+ (\hat{\underline{\boldsymbol{d}}}(3) + \hat{\underline{\boldsymbol{d}}}(4))^2 \cdot \sigma_x^4 + \hat{\underline{\boldsymbol{d}}}(5)^2 \cdot \sigma_x^4 = WS(\hat{\underline{\boldsymbol{d}}}). \end{aligned} \tag{2.36}$$

$WS(\hat{\underline{\boldsymbol{d}}})$ is a weighted sum of the components of the distance vector $\hat{\underline{\boldsymbol{d}}}$. Eq. (2.36) shows how the importance of different Volterra coefficients changes with the standard deviation of the elements of the input vectors. For large values of σ_x, the elements of the Volterra representation that model nonlinearities are more important. This is

intuitively correct since the increase of the input power beyond the linear range of the PAs should allow for better exploitation of the differences in the nonlinearities of the considered units.

2.3.3 Generalized Likelihood Ratio Test

Another algorithm that can be used to solve the hypothesis testing problem from Section 2.1, when the parameter vectors of the users are unknown, is based on the Generalized Likelihood Ratio Test (GLRT). In the case of the GLRT, the receiver does not estimate the parameters, but rather builds and compares the maxima of the likelihood functions over the unknown parameter vectors.

2.3.3.1 Decision Rule

For equally probable hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$, the decision rule of the GLRT can be expressed as

$$\Lambda(\underline{\boldsymbol{Y}}_3) \triangleq \frac{\max_{\underline{\boldsymbol{h}}_1}\{P(\underline{\boldsymbol{Y}}_1, \underline{\boldsymbol{Y}}_3 \,|\, \underline{\boldsymbol{h}}_1, \underline{\boldsymbol{X}}_1, \underline{\boldsymbol{X}}_3)\}}{\max_{\underline{\boldsymbol{h}}_2}\{P(\underline{\boldsymbol{Y}}_2, \underline{\boldsymbol{Y}}_3 \,|\, \underline{\boldsymbol{h}}_2, \underline{\boldsymbol{X}}_2, \underline{\boldsymbol{X}}_3)\}} \begin{array}{c} \mathcal{H}_1 \\ \gtrless \\ \mathcal{H}_2 \end{array} 1. \tag{2.37}$$

In the AWGN channel

$$\begin{aligned} & P(\underline{\boldsymbol{Y}}_k, \underline{\boldsymbol{Y}}_3 \,|\, \underline{\boldsymbol{h}}_k, \underline{\boldsymbol{X}}_k, \underline{\boldsymbol{X}}_3) \\ = & \frac{1}{(\sqrt{2\pi}\sigma_\nu)^{2M}} \cdot \exp\left\{ -\frac{(\underline{\boldsymbol{Y}}_{k3} - P_{k3} \cdot \underline{\boldsymbol{h}}_k)^H (\underline{\boldsymbol{Y}}_{k3} - P_{k3} \cdot \underline{\boldsymbol{h}}_k)}{2\sigma_\nu^2} \right\}, \end{aligned} \tag{2.38}$$

where P_{k3} and $\underline{\boldsymbol{Y}}_{k3}$ are obtained by stacking matrices P_k, P_3 and vectors $\underline{\boldsymbol{Y}}_k$, $\underline{\boldsymbol{Y}}_3$ respectively

$$P_{k3} = \begin{bmatrix} P_k \\ P_3 \end{bmatrix} \qquad k = 1, 2 \tag{2.39}$$

$$\underline{\boldsymbol{Y}}_{k3} = \begin{bmatrix} \underline{\boldsymbol{Y}}_k \\ \underline{\boldsymbol{Y}}_3 \end{bmatrix} \qquad k = 1, 2 \tag{2.40}$$

and where M is the size of the input vector. After substitution of the corresponding probability density functions into (2.37) the decision rule can be rewritten as

$$\min_{\underline{\boldsymbol{h}}_1}\{(\underline{\boldsymbol{Y}}_{13} - P_{13}\underline{\boldsymbol{h}}_1)^H(\underline{\boldsymbol{Y}}_{13} - P_{13}\underline{\boldsymbol{h}}_1)\} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_2}{\gtrless}} \min_{\underline{\boldsymbol{h}}_2}\{(\underline{\boldsymbol{Y}}_{23} - P_{23}\underline{\boldsymbol{h}}_2)^H(\underline{\boldsymbol{Y}}_{23} - P_{23}\underline{\boldsymbol{h}}_2)\}.$$

Since

$$(\underline{\boldsymbol{Y}}_{k3} - P_{k3}\underline{\boldsymbol{h}}_k)^H(\underline{\boldsymbol{Y}}_{k3} - P_{k3}\underline{\boldsymbol{h}}_k) = ||(\underline{\boldsymbol{Y}}_{k3} - P_{k3}\underline{\boldsymbol{h}}_k)||^2, \tag{2.41}$$

the minimizations on each side of (2.41) are typical LS problems. Vectors minimizing the squared error are

$$\hat{\underline{\boldsymbol{h}}}_k = (P_{k3}^H P_{k3})^{-1} P_{k3}^H \underline{\boldsymbol{Y}}_{k3}^H, \quad k = 1, 2 \tag{2.42}$$

and

$$\min_{\underline{\boldsymbol{h}}_k} ||(\underline{\boldsymbol{Y}}_{k3} - P_{k3}\cdot\underline{\boldsymbol{h}}_k)||^2 = ||(\underline{\boldsymbol{Y}}_{k3} - P_{k3}\cdot\hat{\underline{\boldsymbol{h}}}_k)||^2 = \underline{\boldsymbol{Y}}_{k3}^H(I_{2M\times 2M} - P_{k3}(P_{k3}^H P_{k3})^{-1}P_{k3}^H)\underline{\boldsymbol{Y}}_{k3}. \tag{2.43}$$

With (2.43), the decision rule (2.41) can be written as

$$\underline{\boldsymbol{Y}}_{13}^H(I_{2M\times 2M} - P_{13}(P_{13}^H P_{13})^{-1}P_{13}^H)\underline{\boldsymbol{Y}}_{13} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_2}{\gtrless}} \underline{\boldsymbol{Y}}_{23}^H(I_{2M\times 2M} - P_{23}(P_{23}^H P_{23})^{-1}P_{23}^H)\underline{\boldsymbol{Y}}_{23}.$$

2.3.3.2 Algorithm Performance

Eq. (2.22) together with the decision rule (2.44) yields

$$P_e = Pr\{(\underline{\boldsymbol{Y}}_{13}^H(I_{2M\times 2M} - P_{13}(P_{13}^H P_{13})^{-1}P_{13}^H)\underline{\boldsymbol{Y}}_{13} > \underline{\boldsymbol{Y}}_{23}^H(I_{2M\times 2M} - P_{23}(P_{23}^H P_{23})^{-1}P_{23}^H)\underline{\boldsymbol{Y}}_{23})|\mathcal{H}_1\}. \quad (2.44)$$

Under $\mathcal{H}_1$ the third system is actually system 1 with parameters $\underline{\boldsymbol{h}}_1$. Thus vectors $\underline{\boldsymbol{Y}}_{13}$ and $\underline{\boldsymbol{Y}}_{23}$ in (2.44) can be replaced with

$$\underline{\boldsymbol{Y}}_{13} = \begin{bmatrix} \underline{\boldsymbol{Y}}_1 \\ \underline{\boldsymbol{Y}}_3 \end{bmatrix} = \begin{bmatrix} P_1\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{\nu}}_1 \\ P_3\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{\nu}}_3 \end{bmatrix} = P_{13}\underline{\boldsymbol{h}}_1 + \begin{bmatrix} \underline{\boldsymbol{\nu}}_1 \\ \underline{\boldsymbol{\nu}}_3 \end{bmatrix} \quad (2.45)$$

$$\underline{\boldsymbol{Y}}_{23} = \begin{bmatrix} \underline{\boldsymbol{Y}}_2 \\ \underline{\boldsymbol{Y}}_3 \end{bmatrix} = \begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 + \underline{\boldsymbol{\nu}}_2 \\ P_3\underline{\boldsymbol{h}}_1 + \underline{\boldsymbol{\nu}}_3 \end{bmatrix} = \begin{bmatrix} P_2\underline{\boldsymbol{h}}_2 \\ P_3\underline{\boldsymbol{h}}_1 \end{bmatrix} + \begin{bmatrix} \underline{\boldsymbol{\nu}}_1 \\ \underline{\boldsymbol{\nu}}_3 \end{bmatrix}. \quad (2.46)$$

With this substitution and after simple algebraic manipulations the probability of error from (2.44) can be finally put in the form

$$P_e = Pr\{(\underline{\boldsymbol{\nu}} + \underline{\boldsymbol{B}})^H \mathbf{P}\ (\underline{\boldsymbol{\nu}} + \underline{\boldsymbol{B}}) < 0\}, \quad (2.47)$$

where

$$\mathbf{P}_{(3M\times 3M)} = \begin{bmatrix} -(I - P_1X^*P_1^H) & 0 & P_1X^*P_3^H \\ 0 & (I - P_2XP_2) & -P_2XP_3^H \\ P_3X^*P_1^H & -P_3XP_2^H & -P_3(X - X^*)P_3^H \end{bmatrix}$$

$$\underline{\boldsymbol{B}}_{(3M\times 1)} = \begin{bmatrix} \underline{\mathbf{0}} \\ P_2.\underline{\boldsymbol{d}} \\ \underline{\mathbf{0}} \end{bmatrix} \ ; \ \underline{\boldsymbol{\nu}}_{(3M\times 1)} = \begin{bmatrix} \underline{\boldsymbol{\nu}}_1 \\ \underline{\boldsymbol{\nu}}_2 \\ \underline{\boldsymbol{\nu}}_3 \end{bmatrix}$$

$$X = \left(P_2^H P_2 + P_3^H P_3\right)^{-1} \quad ; \quad X^* = \left(P_1^H P_1 + P_3^H P_3\right)^{-1}$$

and

$$\underline{\boldsymbol{d}} = \underline{\boldsymbol{h}}_2 - \underline{\boldsymbol{h}}_1.$$

2.3.4 Naive Method

In addition to the algorithms introduced in the previous subsections, another algorithm termed the Naive Method is considered. In this algorithm, the detection system outputs the device number for which the estimated parameter vector $\hat{\underline{\boldsymbol{h}}}_k, k = 1, 2$, estimated with standard Least Squares, is closest to the estimated parameter vector $\hat{\underline{\boldsymbol{h}}}_3$ of the suspect's device under an L_2-norm criterion

$$||\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_2|| \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} ||\hat{\underline{\boldsymbol{h}}}_3 - \hat{\underline{\boldsymbol{h}}}_1||.$$

2.4 Measurements and Simulations

In this section, the performance of the methods from Section 2.3 is investigated. In particular, the influence of parameters such as the power of the input signal and the signal-to-noise ratio (SNR) on the probability of error is analyzed. This section also provides insight on the variations of components of transmitters used in practical applications, and, most importantly, demonstrates the utility of the approaches for such components even for short input sequences and practical $SNRs$.

Data sheets of digital-to-analog converters usually specify the maximal value of the integral nonlinearity: INL_{max}. In addition to this information, the data sheets often include exemplary INL paths [67]. Because of this, we found no need to perform measurements to examine the variations of the I/O characteristics among the DACs. In the case of the PAs, nonlinearity variations across individual devices are usually

not described in the data sheets. Thus measurements were performed on commercial RF PAs to analyze variations of the I/O characteristics among the PAs.

2.4.1 Exploitation of Digital-to-Analog Converter Nonlinearities for User Identification

The required DAC size for commercial OFDM-based communication applications varies from 6 to 18 bits and depends on the largest signal constellation and number of OFDM subcarriers [68]. For our simulations, we considered 10-bit DACs. We set the standard deviation σ_s of individual DAC sources to be 2% of their nominal value. The upper plot of Figure 2.2 shows 1000 exemplary INL paths of 10-bit, thermometer-coded DACs with $\sigma_s = 0.02$. The lower plot shows the INL_{max} histogram. For 10-bit DACs used in commercial communication transceivers, the value of INL_{max} is typically in the range: ± 1 LSB [69], which justifies the choice of 0.02 as a value for the σ_S.

We applied the algorithms introduced in Section 2.3 to identify devices based on their DAC INL paths. Figure 2.3 shows the probability of error as a function of SNR, averaged over 100 DAC pairs and over 150 input vectors of size $M = 2500$, the elements of which were chosen as realizations of normal random variables (rounded to an integer) with mean value equal to half of the the DAC input range and standard deviation chosen as one third of the half of the input range (which resulted in 99% of the input values within the input range, values outside the input range were ignored). The size M of the input vector was limited by computational capabilities of a computer used to run the simulations. The first eight eigenfunctions of the Brownian Bridge random process were used for INL representation. Note that a relatively high SNR was required for user identification at these short input lengths in this case.

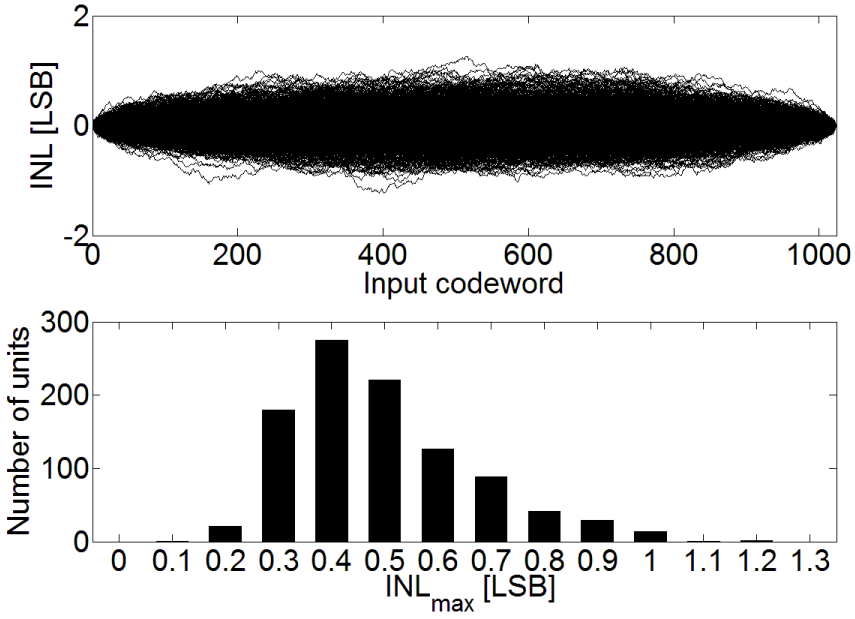

Figure 2.2. INL Brownian Bridge paths of one thousand 10-bit, thermometer-coded DACs with standard deviation of individual sources $\sigma_s = 0.02$ (upper plot) and INL_{max} histogram (lower plot).

2.4.2 Exploitation of Power Amplifier Nonlinearities for User Identification

Similarly, the performance of the considered methods was simulated when the nonlinearities of PAs were exploited for the device identification. We assumed the elements of the input vectors to be realizations of zero-mean normal random variable with standard deviation σ_x. First, to investigate the behavior of P_e as a function of increasing input power and increasing difference of the Volterra representations of considered units, we artificially generated the Volterra series representation of the amplifiers and standard deviation of the elements of the input vectors. Next, and most importantly, we obtained the Volterra series representations from measurements of

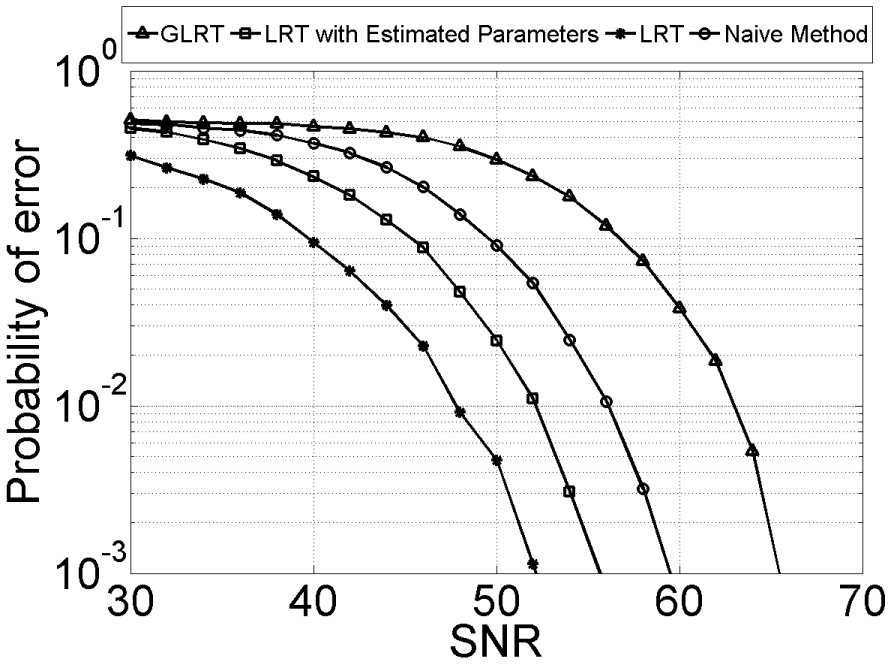

Figure 2.3. Probability of error versus signal-to-noise ratio (controlled with the noise power level), averaged over 100 DAC pairs with $\sigma_S = 0.02$ and over 150 input vectors of size $M = 2500$, with normally distributed elements with mean value equal to half of the DAC's input range and standard deviation equal to $\frac{1}{6}$ of the DAC's input range (which resulted in 99% of the input values within the input range, values outside the input range were ignored). The first eight eigenfunctions of the Brownian Bridge random process were used for INL representation.

actual RF amplifiers and we analyzed the performance of the methods at input power levels specified as linear or 802.11 standard compliant by the manufacturers.

Consider first the artificial generation of amplifier characteristics. Figure 2.4 shows the simulated probability of error of the considered methods, for $SNR = 30dB$, versus standard deviation σ_x of the elements of the input vectors, averaged over 200 different input vectors of size $M = 100$ and over 200 randomly generated Volterra vector pairs. For generation of Volterra vector pairs, random vectors with normally distributed elements $\sim \mathcal{N}(0, 10^{-6})$ were added to a mean value vector: [1 0.01 0.01 0.01 0.01].

Figure 2.5 also shows simulated P_e, but this time for the standard deviation of the elements of the input vectors kept constant ($\sigma_x = 100$) and for the standard deviation

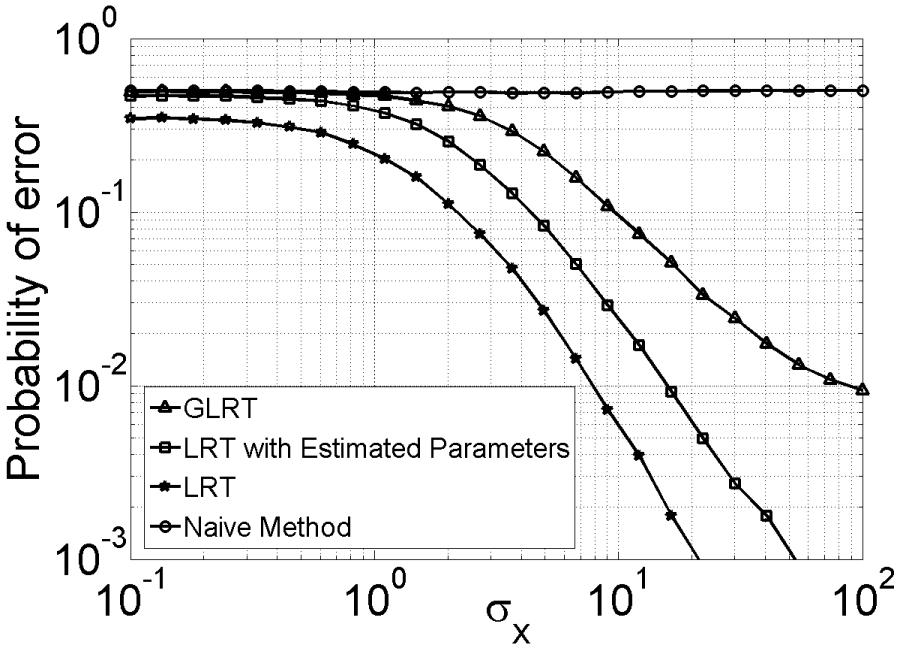

Figure 2.4. Probability of error versus the standard deviation of the elements of the input vectors averaged over 200 different input vectors of size $M = 100$ and over 200 randomly generated Volterra vector pairs, with standard deviation of elements $\sigma_h = 5 \cdot 10^{-3}$; $SNR = 30dB$.

σ_h of elements of random vectors added to the mean value vector [1 0.01 0.01 0.01 0.01] for Volterra vector pairs generation varied in the range $\sigma_h \in (0, 0.005)$. Each point of the curve was obtained as an average over 1000 different input vectors of size $M = 100$ and over 1000 randomly generated Volterra vector pairs. Similarly, as in case of Figure 2.4, the SNR was set to $30dB$.

As expected, Figures 2.4 and 2.5 demonstrate that the performance of the methods increases when the power of input signals increases and when the differences among amplifiers get larger. But speaking more precisely, the methods perform better when the value of the weighted sum from (2.36) increases. In particular the differences in the Volterra series representation of PAs should always be analyzed together with the input power for complete insight into performance of the methods. Figure 2.6 shows P_e as a function of weighted sum $WS(\underline{\boldsymbol{d}})$ (upper plot) and the L_2 distance $||\underline{\boldsymbol{d}}||$ (lower

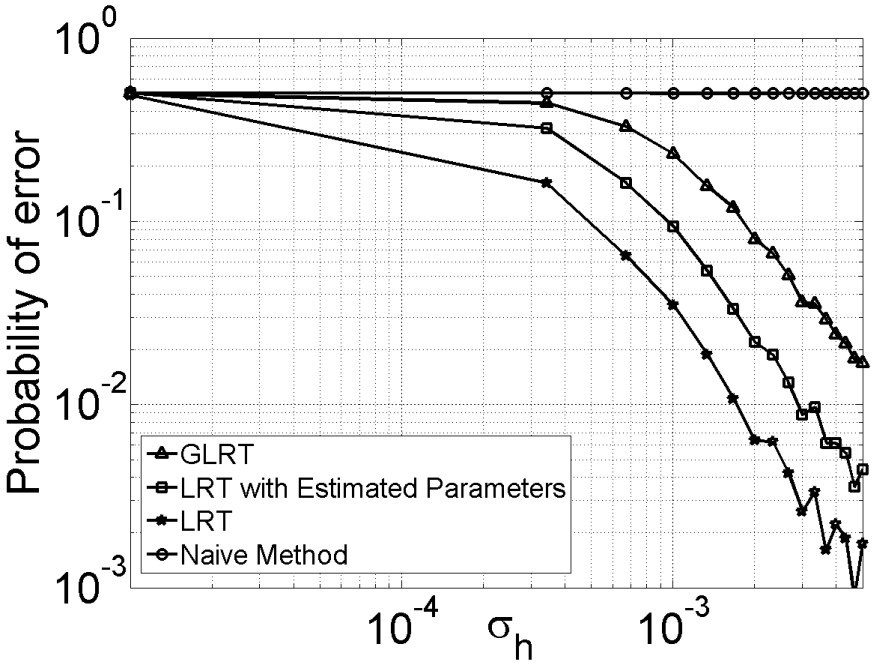

Figure 2.5. Probability of error versus the standard deviation of the Volterra coefficients averaged over 1000 randomly generated Volterra vector pairs and 1000 different input vectors of size $M = 100$, with standard deviation of the elements $\sigma_x = 100$; SNR=30dB.

plot) for 100 randomly generated amplifier pairs with $\sigma_h = 2.5 \cdot 10^{-4}$ averaged over 10000 input vectors of size $M = 100$ with the standard deviation of elements set to $\sigma_x = 100$. It can be seen that the weighted sum is a much more appropriate metric.

To be able to validate effectiveness of the presented anonymity breaking techniques when exploiting imperfections of PAs, we next consider how the nonlinearities of power amplifiers, even these of the same model and from the same manufacturer, differ in practice due to the production process inaccuracies. As mentioned previously nonlinearity variations across devices are usually not described in the data sheets of commercial RF PAs.

Measurements were performed on two different sets of power amplifier chips commercially used in WLAN transmitters. First, two amplifier evaluation boards (MAX2242 EVKIT) loaded with MAXIM MAX2242 [3] amplifiers were stimulated with a $2.45GHz$

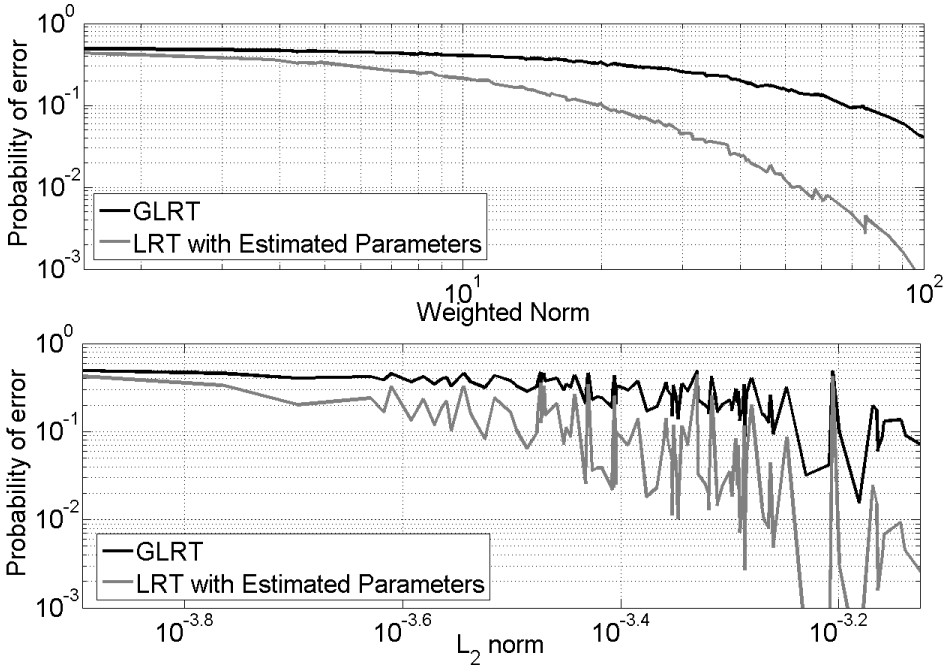

Figure 2.6. Probability of error versus weighted sum from (2.36)- metric that combines differences in Volterra coefficients and power of the input signal (upper plot) and versus L_2 norm of the vector $\underline{\boldsymbol{d}}$- metric that takes into account only differences in Volterra coefficients (lower plot).

sinusoidal signal, generated with Agilent Technologies $E8251A$ $PSG-A$ signal generator and the output was measured on a $12.5GHz$, $50GSa/s$ Tektronix $DPO71254B$ real time oscilloscope. A $20dB$ attenuator was connected to PAs' output to ensure a linear operation of the oscilloscope. We measured multiple points of the single-tone, amplitude I/O characteristics. The measurement points were normalized to the same value of linear gain, which was then used as a value of a linear, zero-memory coefficient of the Volterra representation for both of the amplifiers. The linear part was then subtracted from the normalized characteristics. The nonlinear coefficients of memoryless Volterra representation were then obtained via curve fitting of the remaining, nonlinear part of the measured, normalized characteristics. For a 4^{th} order memoryless model this resulted in the following parameter vectors:

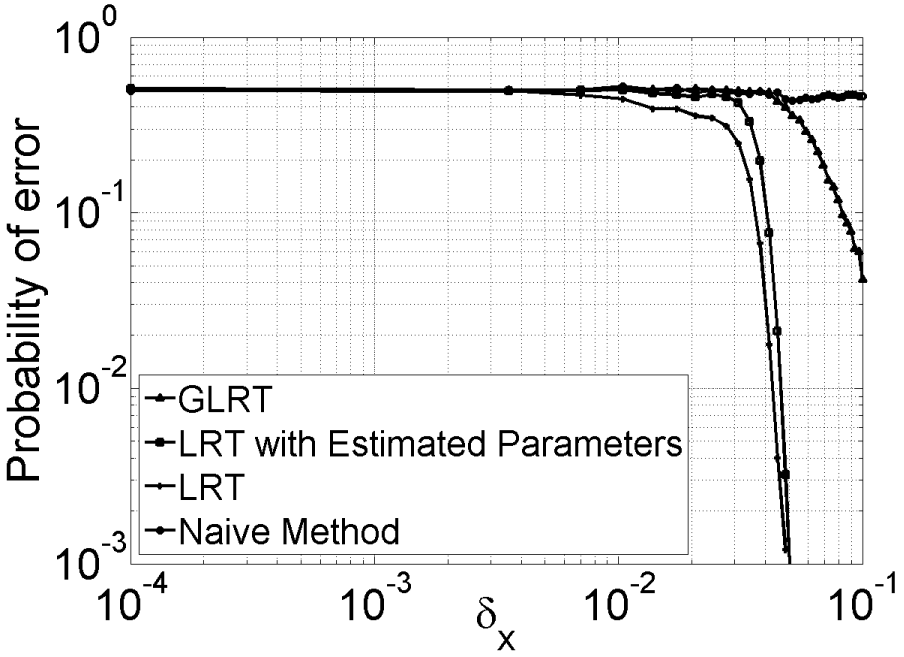

Figure 2.7. Probability of error for measured MAXIM MAX2242 amplifiers versus the standard deviation of the elements of input vectors σ_x, averaged over 2500 input vectors of size $M = 2500$; $SNR = 30dB$.

$$\underline{\boldsymbol{h}}_1 = [32.5462, 29.5342, -509.5277, 1311.5641]$$

$$\underline{\boldsymbol{h}}_2 = [32.5462, 29.4025, -479.4057, 928.3273]\,,$$

which were used for performance simulation results which are reported in Figures 2.7 and 2.8.

For the plots in Figures 2.7 and 2.8, the elements of the input vectors were chosen as the absolute value of realizations of a zero mean random variable $\sim \mathcal{N}(0, \sigma_x^2)$. Figure 2.7 shows how the P_e decreased as the standard deviation of the input went up (while the SNR was kept constant at a level of $30dB$ and length of the input vector was set to $M = 2500$). For standard deviation equal to $\sigma_x = 0.055$ the probability that power of the input signal exceeded -7dBm was only 1%. -7dBm input power corresponds to the upper level of the linear range of the considered MAXIM amplifiers (adjacent channel power ratio below -33dBc and below -55dBc for, respectively, 1^{st}

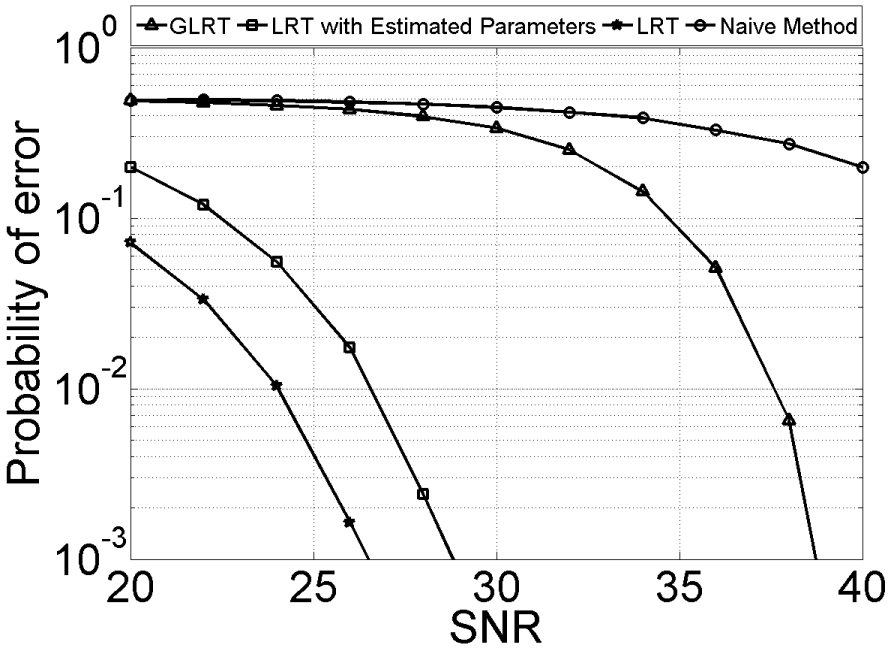

Figure 2.8. Probability of error for measured MAXIM MAX2242 amplifiers versus signal-to-noise ratio, (controlled with the noise power level), averaged over 25000 input vectors of size $M = 2500$, with standard deviation of the elements of input vectors $\sigma_x = 0.055$.

and 2^{nd} side lobe [3]). Whenever the input signal exceeded the the linear range, it was clipped to its upper level. This means that the amplifiers worked in the range specified as linear all the time. Figure 2.8 shows how the P_e behaved for a fixed $\sigma_x = 0.055$ as a function of SNR (again the input signal was clipped to the upper level of the linear region).

Motivated by the promising results obtained with the MAXIM evaluation boards, we next prepared a larger experiment using SKYWORKS SKY 65006-348LF [1] amplifiers. For cost reasons, this motivated the development of our own evaluation board. We executed a similar procedure as in the case of the MAXIM amplifiers to measure the single-tone, amplitude I/O characteristics and to obtain parameter vectors of the 4^{th} order polynomial for the SKYWORKS amplifiers. We chose elements of input vectors to be the absolute value of realizations of a zero-mean normal random vari-

# amplifier	1	2	3	4	5	6	7	8
1	-	0.245	0.000	0.011	0.322	0.000	0.331	0.003
2	0.000	-	0.081	0.308	0.478	0.099	0.135	0.000
3	0.000	0.000	-	0.389	0.030	0.506	0.000	0.000
4	0.000	0.000	0.002	-	0.199	0.428	0.001	0.000
5	0.000	0.226	0.000	0.000	-	0.046	0.153	0.000
6	0.000	0.000	0.318	0.017	0	-	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	-	0.003
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-

Table 2.1. Simulated probability of error of Generalized Likelihood Ratio Test (upper right part) and Likelihood Ratio Test with Estimated Parameters (lower left part) for all possible pairs of 8 SKYWORKS SKY65006-348LF WLAN amplifiers, averaged over 1000 input vectors of size M = 2500. The standard deviation of the components of the input vectors was chosen such that the output power exceeded 21dBm (for which, according to [1], the parts are still 802.11b mask-compliant) with probability equal to 1%. The input was clipped to the upper level of the 802.11b mask-compliant input range. Signal-to-noise ratio (controlled with the noise power level) was set to $35dB$. For SNR=$42dB$ and M=7500, no errors were observed for all possible pairs for the Likelihood Ratio Test with Estimated Parameters in 1000 trials.

able, with a standard deviation set to a value for which 99% of the time the input was within a range specified as 802.11b mask-compliant by the manufacturer; for the remaining 1% of the time, when the input exceeded the 802.11b mask-compliant range, it was clipped to its upper level. Table 2.4.2 shows the simulated probability of error for the Generalized Likelihood Ratio Test (upper right part) and the Likelihood Ratio Test with Estimated Parameters (lower left part) for all possible amplifier pairs from a group of 8 measured SKYWORKS amplifiers, averaged over 1000 input vectors of size $M = 2500$ for SNR set to $35dB$. Note that the identifiability of most of the pairs was very good, at the SNR of $35dB$ for the *very short* input sequence of only 2500 physical-layer symbols. For $SNR = 42dB$ and $M = 7500$ for the Likelihood Ratio Test with Estimated Parameters, we found no identification errors for all possible pairs in 1000 trials. In reality of course, due to very high transmission rates

employed, even short sessions will generally consist of hundreds or thousands of data packets that can be used for user identification.

One of the main concerns about RF fingerprinting approaches exploiting transmitter's hardware imperfections is that they can be negatively influenced by the variation of the performance of the transmitter components across the temperature. In particular, a meaningful variation of the performance of the power amplifiers can be observed as a function of temperature. This is particularly true in large base station amplifiers, where such temperature variation is the bane of designers attempting to linearize such. However power amplifier chips used on wireless cards of today's mobile devices are very small (usually in the range of $4 - 6mm^2$). These small chips achieve their normal operating temperature very quickly, and in fact, we have observed such stabilization of the chip temperature after tens of seconds. This is a very short time that is often needed for the mobile device to boot up. Therefore, we believe it is fair to ignore the temperature variations in this initial investigation. However, we do believe that it is an important consideration as we consider further refinement of our algorithms.

2.4.3 Evaluation of the Results

To our knowledge, model-based approaches similar to ours have not yet been investigated. Thus, it is to hard conduct a comparison of our results with results of previous work. In particular, it is not possible to conclusively compare our simulation and measurement results with the strictly empirical results Hall et al. [23] and Brik et al. [30], for which numerous parameters are not specified, including such basic ones as the operating signal-to-noise ratio. Hence, we are reduced to re-stating the experimental outcomes of Hall et al. [23] and Brik et al. [30] and comparing them quite roughly to our work. In [23], Hall et al. report an average success rate of 94-100%, while trying to distinguish among 14 802.11 transceivers. In [30], Brik et al. report

identification error rates equal to fractions of a percent (0.34% for their best scheme), while distinguishing among 138 802.11 transceivers. Methods introduced by Brik et al. can however be thwarted with a success rate close to 100% with a simple adjustment of the carrier frequency of the masquerading device and with digital modifications of constellation symbols, as reported by Danev et al. [33]. As we mentioned previously, for $SNR = 42dB$, even for short input sequences of $M = 7500$, no errors were observed during 1000 simulation trials, while trying to distinguish among 8 802.11b mask-compliant PAs from the same manufacturer. This suggests that our methods can outperform methods of Hall et al. and Brik et al. ([23] and [30]), when applied to the same setup, but, per above, we cannot make this statement conclusively. While methods of Hall et al. and Brik et al. are strictly experimental and their results hard to reproduce, one advantage of our model-based approach is that the results are easy to replicate for comparison to the performance of methods developed by others in the future.

2.5 Conclusions

In this chapter, we proposed a new approach for breaking user anonymity in wireless communication systems based on minute imperfections of different components of the transmitter hardware. The general models used to model the transmitter components allow for the determination of the probability of error of the decisions, which makes the proposed methods especially interesting for establishing probable cause and for use in court. Our simulations have shown that the nonlinear variations of digital-to-analog converters can only be exploited when the signal-to-noise ratio is very high. However, in the case of power amplifiers, measurements from commercially employed chips indicate that amplifiers can be easily identified at typical power levels, at practical $SNRs$ and with short observed sequences.

CHAPTER 3

IDENTIFICATION OF WIRELESS DEVICES OF USERS WHO ACTIVELY FAKE THEIR RF FINGERPRINTS WITH ARTIFICIAL DATA DISTORTION

3.1 Problem Statement

As discussed in Chapter 2, variations in the RF chain of radio transmitters caused by imperfections of manufacturing processes can be used as a signature to uniquely associate wireless devices with a given transmission. In Chapter 2 we proposed a model-based approach that allows for the identification of wireless devices based on signatures obtained with time domain analysis of steady state pairs of received and decoded signals. All of the RF fingerprinting techniques based on steady state signal analysis exploit the fact that nonideal transmitters cause signal distortions that, while being slight enough for the transmitters to meet requirements of the communication standards, are significant enough to make the distortions observable and able to be tied with an individual transmitter. Therefore, an *adversary* user, who is capable of modification of the higher-layer tags such as the IP address and MAC address can be successfully identified based on its physical-layer fingerprint. However, a *strong adversary*, aware of the fingerprinting methods, could inject slight distortions to the digital data signal, before the signal is exposed to transmitter's nonlinearities, based on which the RF fingerprint is extracted. This slight, artificial distortion while still allowing for reliable data transmission (see Section 3.4), would change the character of the total distortion observable at the receiver over time, and hence the character of the fingerprint which would significantly degrade performance of the steady state signal based methods from Chapter 2 and methods of Brik et al. [30]. In particular

in [33] authors report thwarting the identification methods of Brik et al. [30], with success rate close to 100%, with a simple adjustment of the carrier frequency of the masquerading device and with digital shrinking/expanding of the constellation symbols' positions.

The *strong adversary*, considered in this chapter, has the capability to intentionally distort the digital data by taking the digital symbols off the signal constellation grid, before they are pulse-shaped and exposed to the PA's nonlinearity. When such a distortion is employed, the correctly detected data symbols, chosen at the receiver as elements of the constellation, differ from the symbols at the transmitter directly before the pulse-shaping and the exposure to the PA nonlinearities. This intentional distortion, while causing only slight signal quality degradation (see Section 3.4), causes significant degradation of the identification methods from Chapter 2, which makes the attack very attractive for the *strong adversary*. In particular, in Chapter 2 we assumed that for the PA, which dominates the transmitter's nonlinearity, the input samples are accessible, as they can be reconstructed from the correctly decoded data at the receiver if the rest of the transmitter chain is assumed linear. We then used time-domain analysis of the decoded data and the received signal for the identification decisions. In this chapter, we address a scenario when the input to the transmitter's PA can be different from the input reconstructed at the receiver because of slight distortions that the *strong adversary* could have injected to the digital symbols, in addition to inherent waveform distortions caused by the PA impairments. Such injection might allow the *strong adversary* to fake the RF signature of its device, while still allowing for correct data decoding. Therefore, in this chapter we develop an identification method that allows for separation of the two possible sources of distortion: modification of the digital data symbols by a *strong adversary* and inherent transmitter nonlinearities, thus allowing for successful identification of the devices used by the *strong adversaries*.

The method proposed in this chapter is based on the observation that the nonlinearity of the radio frequency power amplifiers, which are the last elements of the transmitter chain and thus cannot be influenced by software modifications, cause slight in-band distortion and spectral regrowth of the signal that is dependent on the parameters of the amplifier's nonlinearity. We demonstrate that this distortion can be isolated from a potential, additional cause of spectral modification of the waveform: modification of the digital data symbols by the *strong adversary*. Hence, with oversampling of the captured signals at the receiver, the mobile devices can be identified even if the masquerading users fake their RF signatures by injecting artificial distortions to the data symbols while committing the crime.

It is important to stress here that, similarly to the Chapter 2, we only exploit the differences in the nonlinear character of the power amplifiers. Differences of values of the linear gain that could be masked by varying the distance between transmitter and receiver or fading effects of the channel are ignored. Thus, as in Chapter 2, we normalize all captured signals to the same gain value to prevent the exploitation of gain for identification. Moreover, we consider a scenario where *strong adversaries* are capable of changing the character of the distortion applied to each packet individually. Hence, only a single packet is available to be used to detect and characterize the distortion and to correct for it.

Figure 3.1 presents the identification scenario considered in this chapter. Similarly to Chapter 2 we study here a post-incident mobile device identification method, which involves testing of devices from a pool of suspects in order to decide which one was most likely used while the crime was committed, when high-layer identification mechanisms fail or are not implemented. The two-device scenario from Chapter 2 (Figure 2.1) is extended to a K-device scenario. A signal record from the criminal's device is captured by a receiver at the crime scene, when the *strong adversary* either does or does not distort the digital symbols in order to fake the device's RF signature.

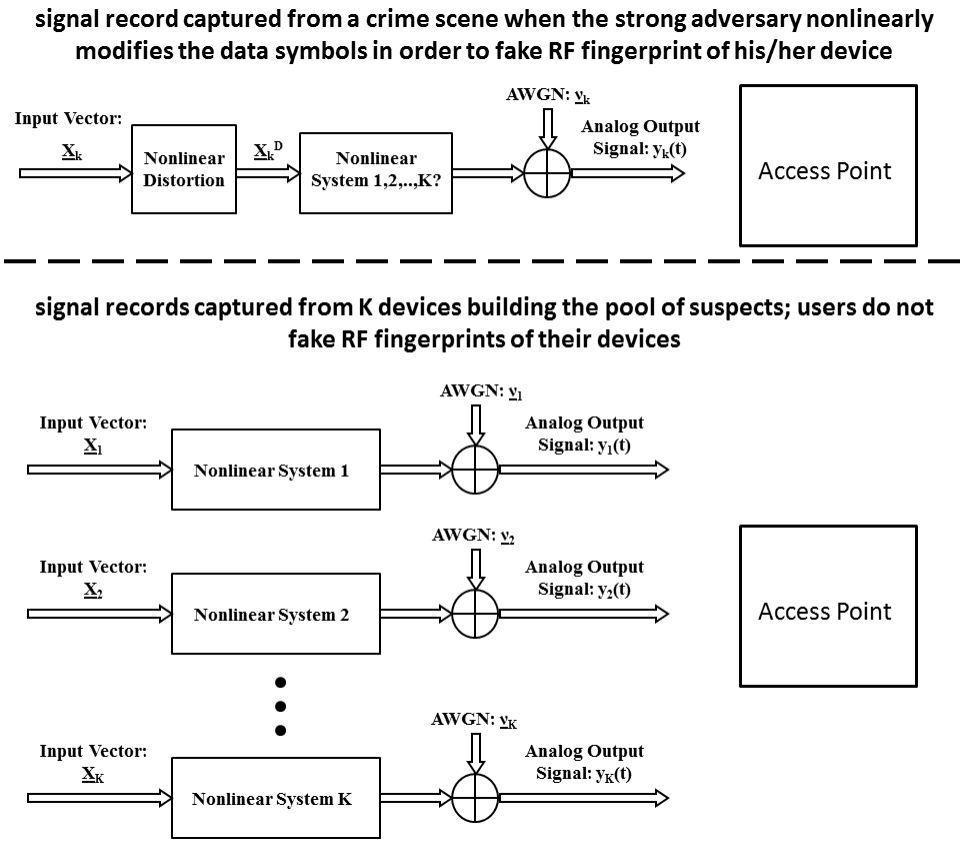

Figure 3.1. The K-device identification scenario: record captured at a crime scene from a device used by the *strong adversary* that is capable of artificial distortion of the data symbols applied in order to fake RF signature of the device (upper part); post-crime records captured from K devices building a pool of suspects (lower part). An additive white Gaussian noise (AWGN) channel model is assumed.

After the crime is committed, using the same receiver, records are captured from a group of devices that might have been used to commit the crime. The goal of this work is to tie transmissions from the crime scene to other transmissions from that same device. Being able to indicate a device that, with a given probably, was used to commit the crime, can then allow law enforcement to reduce the size of the original group of suspect devices and to justify issuing device confiscation warrants that can then lead to final identification decisions and possible arrests based on the digital content of the devices.

Our identification method is based on an assumption that only short transmissions are available from the crime scene captured at times when *strong adversaries* might or might have not injected distortions to their data symbols and that the records from the devices that are building the pool of suspects are captured at times when users do not distort the data signals, and that the devices can be observed for a long enough time to obtain relatively accurate estimates of their true RF signatures. We believe that this is a reasonable assumption, since the digital data distortion does cause some signal quality degradation and thus device performance degradation that the *strong adversary* is not expected to tolerate over extended period of time. This is analogous to a criminal taking an uncomfortable mask off some time after he masqueraded to commit the crime.

3.2 Modeling the Spectrum of the Output of the Nonlinear RF Power Amplifier

Behavioral modeling of RF power amplifiers is an extensive research area that concentrates on extraction of low complexity models for system simulations that accurately capture the performance impairments and distortions caused by circuit level effects. Among various models that have been considered to model the behaviour of the PAs are polynomial memoryless models, two-box Hammerstein and Wiener models, multiple-box models, Volterra series based models and neural networks based models. Piecewise modeling approaches have also been considered. A variety of survey papers and books provide a rich overview of the existing behavioral modeling approaches e.g. [70–72]. In this work we are considering low order, memoryless polynomial behavioral models, which as reported by Isaksson et al. in [71], allow for high modelling accuracy even for wide-band signals (up to 20MHz baseband). The extension to more complex models is possible, but leads to increase of complexity of the PA output spectrum model introduced in (3.7).

Consider a communication signal modeled with a random process

$$x(t) = \sum_{n=-\infty}^{\infty} a_n \cdot p(t - nT_s), \tag{3.1}$$

where $p(t)$ is an analog pulse and a_n's are data symbols, which for commonly used digital modulation schemes are modeled as identically distributed, uncorrelated, zero-mean complex random variables, for which the distribution depends on the digital modulation scheme. $x(t)$ is not a wide-sense stationary, but a cyclostationary random process, as its expected value is zero and its autocorrelation function is periodic in T_s. Thus, instead of using the Wiener−Khinchin theorem, we consider the more general definition of power spectral density (psd) in order to describe the spectrum of the signal captured from the device that needs to be identified.

Consider an arbitrary random process $x'(t)$. Its power spectral density can be expressed as

$$S_{x'}(f) = \lim_{T\to\infty} \frac{1}{T} E\left[\left|X'_T(f)\right|^2\right], \tag{3.2}$$

where

$$x'_T(t) = \begin{cases} x'(t), & t \in (-T/2, T/2) \\ 0, & \text{otherwise .} \end{cases} \tag{3.3}$$

and $X'_T(f)$ is the Fourier transform of $x'_T(t)$.

On the transmitter side the cyclostationary communication process $x(t)$ from (3.1) is amplified with a nonlinear amplifier, the characteristic of which can be accurately modeled with an odd-order polynomial with coefficients $\{h_{2p-1}, p = 1, 2, ..., P\}$ [71]. The resultant process $y(t)$ can be expressed as

$$y(t) = \sum_{p=1}^{P} h_{2p-1} \left(\sum_{n=-\infty}^{\infty} a_n \cdot p(t - nT_s) \right)^{(2p-1)} . \tag{3.4}$$

Define

$$y_N(t) = \begin{cases} y(t), & t \in \left(\frac{-N}{2}T_s, \frac{N}{2}T_s\right) \\ 0, & \text{otherwise} \end{cases} \tag{3.5}$$

and $Y_N(f) = \mathcal{F}\{y_N(t)\}$. With (3.2) the power spectral density of the random process $y(t)$ sampled at the receiver, after being sent through an additive white Gaussian noise (AWGN) channel, can then be expressed as

$$S_y(f) = \lim_{N\to\infty} \frac{1}{2N+1} \cdot E\left[Y_N(f) \cdot Y_N(f)^*\right] + \sigma_\nu^2, \tag{3.6}$$

where σ_ν^2 is a power spectral density of an AWGN process $\nu(t)$.

With (3.4), (3.5) and (3.6), and the linearity of expectation and the Fourier transform, the power spectral density of $y(t)$ can be expressed as

$$\begin{aligned} S_y(f) = \sigma_\nu^2 + \sum_{p_1=1}^{P} h_{2p_1-1} \sum_{p_2=1}^{P} h^*_{2p_2-1} \lim_{N\to\infty} \frac{1}{2N+1} \\ \sum_{n_1^1=-N}^{N} \cdots \sum_{n^1_{2p_1-1}=-N}^{N} \sum_{n_1^2=-N}^{N} \cdots \sum_{n^2_{2p_2-1}=-N}^{N} \\ E\left[a_{n_1^1} \cdot \ldots \cdot a_{n^1_{2p_1-1}} \cdot a^*_{n_1^2} \cdot \ldots \cdot a^*_{n^2_{2p_2-1}}\right] \\ \cdot\mathcal{F}\left\{p(t - n_1^1 T_s) \cdot \ldots \cdot p(t - n^1_{2p_1-1} T_s)\right\} \\ \cdot\mathcal{F}\left\{p(t - n_1^2 T_s) \cdot \ldots \cdot p(t - n^2_{2p_2-1} T_s)\right\}^*. \end{aligned} \tag{3.7}$$

Because the $a_n{}'s$ are identically distributed, uncorrelated and zero-mean random variables, (3.7) simplifies significantly, as the expected value

$$E\left[a_{k_1^1} \cdot \ldots \cdot a_{k^1_{2p_1-1}} \cdot a^*_{k_1^2} \cdot \ldots \cdot a^*_{k^2_{2p_2-1}}\right]$$

inside of the multiple sum takes non-zero values only when among all $(2p_1-1)+(2p_2-1)$ sum indices n, all subsets of indices that take the same values have size that is an

even number. In all other cases, because of the uncorrelated and zero-mean property of the $a_n's$, the expected value can be written as a product of factors, at least one of which is equal to zero. Therefore, for a known pulse-shaping filter, the power spectral density (3.7) can be simplified and expressed as a function of even-order central moments (pg. 86 of Section 2.4 [73]) of the, potentially digitally distorted, random variables a_n, and of the coefficients of the nonlinearity of the amplifier. For a 5^{th} order, real, odd polynomial representation of the PA's input/output (I/O) characteristic ($P = 3$ in (3.7)) (3.7) can be reduced to

$$\begin{aligned}
&S_y(f, \mu_2, \mu_4, \mu_6, \mu_8, \mu_{10}, h_1, h_3, h_5) \\
&= h_1^2 \cdot \mu_2 \cdot R_1(f) + h_1 h_3(\mu_4 \cdot R_2(f) + \mu_2^2 \cdot R_3(f)) \\
&+ h_1 h_5(\mu_6 \cdot R_4(f) + \mu_2 \cdot \mu_4 \cdot R_5(f) + \mu_2^3 \cdot R_6(f)) \\
&+ h_3^2(\mu_6 \cdot R_7(f) + \mu_2 \cdot \mu_4 \cdot R_8(f) + \mu_2^3 \cdot R_9(f)) \\
&+ h_3 h_5(\mu_8 \cdot R_{10}(f) + \mu_6 \cdot \mu_2 \cdot R_{11}(f) + \mu_4^2 \cdot R_{12}(f) \\
&+ \mu_4 \cdot \mu_2^2 \cdot R_{13}(f) + \mu_2^4 \cdot R_{14}(f)) + h_5^2(\mu_{10} \cdot R_{15}(f) \\
&+ \mu_6 \cdot \mu_4 \cdot R_{16}(f) + \mu_6 \cdot \mu_2^2 \cdot R_{17}(f) + \mu_4 \cdot \mu_2^3 \cdot R_{18}(f) \\
&+ \mu_2^5 \cdot R_{19}(f)) + \sigma_\nu^2
\end{aligned} \tag{3.8}$$

where μ_L is the L^{th} central moment of a_n, and the $R_l(f)'s$ are functions that only depend on the pulse $p(t)$ used for pulse-shaping and can be found as sums of products of Fourier transforms of products of time shifted pulses $p(t)$. For example

$$
\begin{aligned}
R_1(f) &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \mathcal{F}\left\{p(t-nT_s)\right\} \cdot \mathcal{F}\left\{p(t-nT_s)\right\}^* \\
R_2(f) &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} 2\Re\left\{\mathcal{F}\left\{p(t-nT_s)\right\} \cdot \mathcal{F}\left\{p^3(t-nT_s)\right\}^*\right\} \\
R_3(f) &= \lim_{N\to\infty} \frac{1}{2N+1}\, 3 \sum_{n_1=-N}^{N} \sum_{n_2=-N}^{N} 2\Re\left\{\mathcal{F}\left\{p(t-n_1T_s)\right\} \cdot \mathcal{F}\left\{p(t-n_1T_s)p^2(t-n_2T_s)\right\}^*\right\} \\
R_{11}(f) &= \lim_{N\to\infty} \frac{1}{2N+1}\, 3 \sum_{n_1=-N}^{N} \sum_{n_2=-N}^{N} 2\Re\left\{\mathcal{F}\left\{p(t-n_1T_s)p^2(t-n_2T_s)\right\} \cdot \mathcal{F}\left\{p^5(t-n_1T_s)\right\}^*\right\} \\
&\quad + 15 \sum_{n_1=-N}^{N} \sum_{n_2=-N}^{N} 2\Re\left\{\mathcal{F}\left\{p^2(t-n_1T_s)p(t-n_2T_s)\right\} \cdot \mathcal{F}\left\{p^4(t-n_1T_s)p(t-n_2T_s)\right\}^*\right\} \\
&\quad + 10 \sum_{n_1=-N}^{N} \sum_{n_2=-N}^{N} 2\Re\left\{\mathcal{F}\left\{p^3(t-n_1T_s)\right\} \cdot \mathcal{F}\left\{p^3(t-n_1T_s)p^2(t-n_2T_s)\right\}^*\right\}.
\end{aligned}
\tag{3.9}
$$

Although $N \to \infty$ in (3.7) and (3.9), in practice these sums are finite, because practical pulses have finite lengths and their shifted versions overlap only up to a given finite relative time shift. Figure 3.2 shows the $R_l(f)$ functions from (3.8) for $l = 1, 2, \ldots, 18$, for a raised-cosine pulse-shaping filter with roll-off factor $r = 0.5$. Function $R_{19}(f)$ has not been plotted because of the extensive time required for its calculation. Eq. (3.8) then shows how the psd of a received waveform changes with the change of the moments of the data symbols that can be caused by potential distortions intentionally injected to the data symbols by the strong adversary. As we will show in the numerical results of Section 3.4, very good performance of the proposed identification method is obtained even if the first 18 out of the 19 $R_l(f)$ functions are used to calculate the psd model (3.8), as the contribution of the $R_l(f)$ decreases with increasing index l. Model (3.8) was derived for a low-order polynomial PA model (3.4) with real coefficients h. The extension of (3.8) to low-order, complex polynomial PA models, which according to [71] allow for good accuracy of modeling practical RF amplifiers, as well as extension to other practical PA models [71] is possible at the expense of complexity of (3.8).

Algorithm 1 Post-Crime Device Identification Algorithm

for number of hypotheses (number of devices building the pool of suspects) **do**

for number of packets captured when users do not distort data symbols **do**

- Receive a packet oversampled with factor M
- Normalize the received oversampled packet to unitary power and calculate the periodogram estimate of the received signal's psd
- Decode the data symbols
- Estimate the first L even central moments of the decoded data symbols
- Downsample the captured packet and normalize the received data symbols to gain $G = 1$
- Based on the decoded data symbols and the received, normalized data symbols estimate P coefficients of an odd-order polynomial model of the PA's I/O characteristic

end for

- Average the central moment estimates, estimates of the coefficients of the PA's I/O polynomial model, and the estimates of the psd over all captured packets
- With a known range of the decoded data symbols and with the averaged estimates of the coefficients of the PA's I/O polynomial model, find the coefficients of a polynomial function inverse to the PA's I/O polynomial model

for number of packets captured from the the strong adversary's device at the crime scene **do**

- Capture a packet oversampled with factor M
- Normalize the captured oversampled packet to unitary power and calculate the periodogram estimate of the received signal's psd
- Downsample the captured packet and normalize the received data symbols to gain $G = 1$
- Apply the polynomial function inverse to the devices's PA's I/O polynomial model to the normalized received data symbols
- Estimate the first L even central moments of the normalized received data symbols after applying the inverse function
- Calculate the correction of the psd estimates with (3.15)

end for

- Average the estimate of the psd over the number of packets captured at the crime scene
- Average the calculated correction over the number of packets captured at the crime scene
- Calculate likelihood function (3.12) with the corrected psd estimate

end for

- With (3.14) find an index of a device with the maximal value of the likelihood function

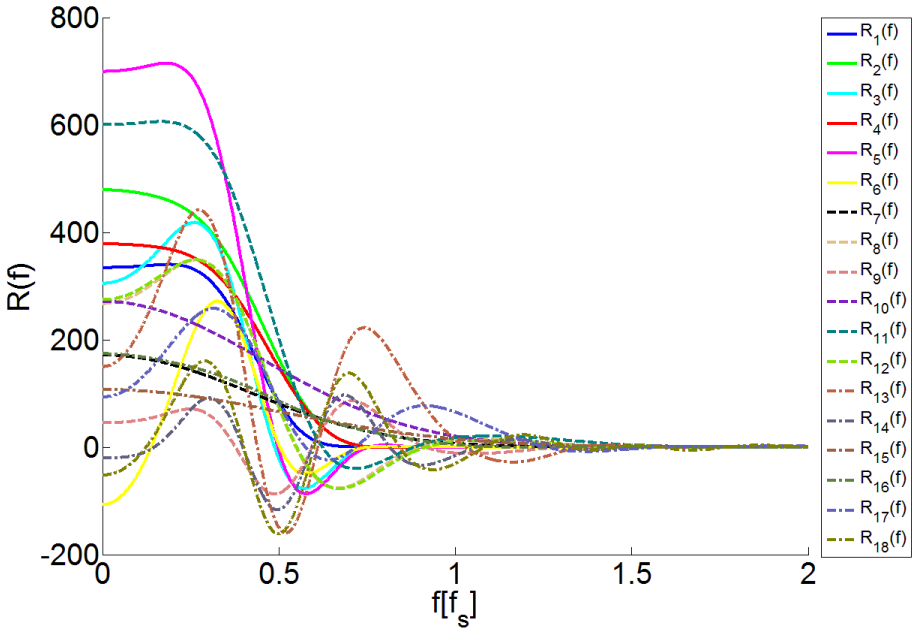

Figure 3.2. $R_l(f)$ functions, $l = 1, 2, \ldots, 18$, used in (3.8) calculated for a raised-cosine pulse-shaping filter with roll-off factor $r = 0.5$.

3.3 Proposed Identification Method

3.3.1 Hypothesis Test

Eq. (3.6) is a well-known formula for a periodogram spectral estimator (pg. 65 of Section 4.3 [74]). Because such an estimator relies on random data of limited length, the estimate of the psd at each frequency is a random variable itself. Although the mean value of the estimate goes to the true value as $N \to \infty$, the variance is unaffected by the length of the captured time sequence (pg. 66 of Section 4.3 [74]). The variance of the estimate can only be reduced by averaging the periodograms calculated over multiple data sequences. The values of the periodogram at each frequency asymptotically behave like independently-distributed Chi-square (for a non-averaged periodogram) and Gamma (for an averaged periodogram) random variables with mean value equal to the true value of the psd [75]. Hence a likelihood ratio test can be performed to reveal the identity of the wireless device.

Consider the two-device scenario. The two hypotheses of the test are $\mathcal{H}_1$: masquerading user uses device 1; $\mathcal{H}_2$: masquerading user uses device 2. Then the likelihood ratio test is

$$\Lambda = \frac{p_{S_\mathcal{Y}|\mathcal{H}_1}(S_\mathcal{Y}|\mathcal{H}_1)}{p_{S_\mathcal{Y}|\mathcal{H}_2}(S_\mathcal{Y}|\mathcal{H}_2)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \tau \tag{3.10}$$

Because the hypotheses are equally probable, for uniform Bayesian costs, a threshold that minimizes the risk of the test is $\tau = 1$ (pg. 26 of Section 2.2 [76]). The two-device scenario can easily be generalized to a K-device scenario, for which the identified device k is the device for which the likelihood function takes maximal value:

$$k_{opt} = \max_{k=1,\ldots,K} p_{S_\mathcal{Y}|\mathcal{H}_k}(S_\mathcal{Y}|\mathcal{H}_k). \tag{3.11}$$

For the more general case of the averaged periodogram, the likelihood functions are

$$p_{S_\mathcal{Y}|\mathcal{H}_k}(S_\mathcal{Y}|\mathcal{H}_k) = \prod_{n=1}^{N_{DFT}} S_\mathcal{Y}(f_n)^{\kappa-1} \frac{\exp\{-S_\mathcal{Y}(f_n)/\Theta_k(f_n)\}}{\Gamma(\kappa)\Theta_k(f_n)^\kappa} \tag{3.12}$$

where N_{DFT} is the length of the discrete Fourier transform (DFT) applied to calculate the discrete version of $Y(f)$ from (3.6), κ is a shape parameter of the Gamma distribution that is equal to the number of averaged periodograms, and $\Theta_k(f_n)$ is a scale parameter of the Gamma distribution at the frequency f_n. Under hypothesis $\mathcal{H}_k$, $\Theta_k(f_n)$ is equal to

$$\Theta_k(f_n) = \zeta_k(f_n)/\kappa, \tag{3.13}$$

where $\zeta_k(f_n)$ is the true value of the psd at frequency f_n. $\Gamma(\kappa)$ is the Gamma function, which for positive integers κ takes values $\Gamma(\kappa) = (\kappa-1)!$. With (3.12), and because of the monotonicity of the logarithm, the likelihood test (3.11) can be rewritten as

$$k_{opt} = \max_{k=1,\ldots,K} \sum_{n=1}^{N_{DFT}} \left\{ (\kappa-1)\ln\left(S_\mathcal{Y}(f_n)\right) - \frac{S_\mathcal{Y}(f_n)}{\Theta_k(f_n)} - \kappa \cdot \ln\left(\Theta_k(f_n)\right) \right\}. \tag{3.14}$$

Typically the true psd values $\zeta_k(f_n)$ from (3.13) are not available and their accurate estimates $\hat{\zeta}_k(f_n)$ are used to calculate estimates $\hat{\Theta}_k(f_n)$ of the scale parameters (3.13).

3.3.2 Correction of Hypothesis Test True Values for Identification of a Strong Adversary

The K-ary likelihood test (3.14) provides good performance only if the adversary does not inject distortions to its data symbols while committing the crime. If the masquerading user modifies moments of its data symbols, the performance of the test can degrade significantly. However, with the model from (3.8) it is possible to take into account changes of the psd caused by moments' modifications and accordingly correct the scale parameter vectors $\Theta_k(f_n)$ used in (3.14). For this, a knowledge of the moments of the data symbols for the captures from the devices building the pool of suspects and for capture from the device used to commit the crime is needed. For the devices from the pool of suspects, unmodified decoded input data symbols are easily accessible, and hence these moments can be estimated very accurately. However, for the capture from the crime scene, the correctly decoded data symbols are not necessarily identical to the data symbols generated at the transmitter, and we cannot estimate the moments of the data symbols based on the decoded data. Therefore for the purpose of identification, instead of decoding the received data from the crime scene, we apply functions that are the inverse of the nonlinear I/O characteristics of the amplifiers that under each hypothesis can be accurately estimated from the I/O data collected from devices building the pool of suspects. Moments of the data symbols obtained after applying these inverse functions can then be used to calculate the corrected estimates of the true values of the psd $\hat{\zeta}_k^C(f_n)$ and the corrected scale parameters $\hat{\Theta}_k^C(f_n) = \hat{\zeta}_k^C(f_n)/\kappa$.

$$\begin{aligned}\hat{\zeta}_k^C(f_n) = \hat{\zeta}_k(f_n) - S_{\mathcal{Y}}(f, \widehat{\mu}_{2,k}, \widehat{\mu}_{4,k}, \widehat{\mu}_{6,k}, \widehat{\mu}_{8,k}, \widehat{\mu}_{10,k}, \hat{h}_{1,k}, \hat{h}_{3,k}, \hat{h}_{5,k}) \\ + S_{\mathcal{Y}}(f, \widehat{\mu}'_{2,k}, \widehat{\mu}'_{4,k}, \widehat{\mu}'_{6,k}, \widehat{\mu}'_{8,k}, \widehat{\mu}'_{10,k}, \hat{h}_{1,k}, \hat{h}_{3,k}, \hat{h}_{5,k}).\end{aligned} \tag{3.15}$$

$\hat{\zeta}_k(f_n)$ in (3.15) is the psd estimated accurately for device k when the user was not distorting the data symbols. $S_{\mathcal{Y}}$ is the model from (3.8). $\widehat{\mu}_{L,k}$ are the L^{th} central moments of the undistorted data symbols accurately estimated for the device k. $\widehat{\mu}'_{L,k}$ are the L^{th} central moments of data symbols from the device that was used to commit the crime, obtained after applying functions inverse to the estimated nonlinearity of the amplifier under hypothesis $\mathcal{H}_k$. The $\hat{h}_{j,k}$ are estimated j^{th} coefficients of the odd 5^{th} order polynomial approximation of the I/O characteristic of the amplifier of the device k. Algorithm 1 summarizes the proposed scheme for post-incident identification of devices used by strong adversaries.

Since (3.8) is a nonlinear function of the central moments, the correction needs to be applied with the periodicity not lower than the periodicity of changes of the character of data symbol distortion of which the strong adversary is capable. In this work we assume a pessimistic scenario where the strong adversary can change the character of the distortion applied to the data symbols for each individual packet. Hence the correction needs to be applied to each received packet individually, and thus the moment estimates used for the correction are obtained from a single packet only. A lower periodicity of distortion character changes can only increase the accuracy of the moment estimates and thus increase the performance of the proposed identification method, reported in the following Section 3.4.

3.4 Numerical Results

To investigate the performance of the proposed identification device method, we first generate polynomial representations of PA I/O characteristics artificially. Next, and most importantly, we obtain polynomial representations for actual RF power

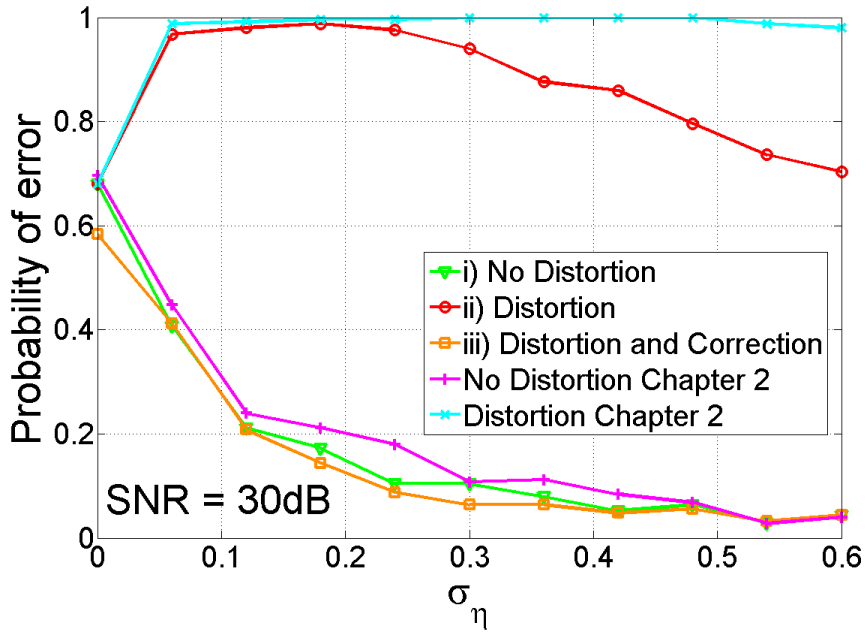

Figure 3.3. Probability of erroneous identification decision (3.14), calculated over 250 randomly generated groups of 3 power amplifiers and input signals, as a function of standard deviation σ_η of the zero-mean, normal random variable η (3.16), for $SNR = 30dB$, for 50 signal records of length 1024 symbols, captured from the device used to commit a crime, and for 500 signal records of length 1024 undistorted symbols captured from the three suspected devices, for the three cases **i)**, **ii)** and **iii)**, described in Section 3.4, together with the performance of the time domain based methods of Chapter 2.

amplifiers by measurements, and we analyze the performance of the identification method at input power levels specified as linear by the manufacturer.

Consider first the artificial generation of amplifier polynomial representations and a 3-hypotheses scenario, where each amplifier was modeled with a real 5^{th} order odd polynomial coefficient vector. The first of the three vectors was arbitrarily set to $\underline{h} = [1 \quad h_3 \quad h_5] = [1 \quad -0.1 \quad -431.71]$. The second and the third vectors were then generated as a sum of the vector $\underline{h}$ and respective random vectors

$$\underline{h}_{add_2} = [0 \quad -|\eta|h_3 \quad -|\eta|h_5], \qquad \underline{h}_{add_3} = [0 \quad |\eta|h_3 \quad |\eta|h_5], \tag{3.16}$$

where η was a zero-mean, normal random variable with standard deviation σ_η. The input data signal to the power amplifiers was modeled as a sequence of realizations of a zero-mean, normal random variable with standard deviation σ_x, clipped to the level C, and pulse-shaped with a raised-cosine pulse-shaping filter with a roll-off factor $r = 0.5$. The oversampling ratio $\mathcal{O}$ of the pulse-shaping filter was set to $\mathcal{O} = 4$. The clipping level C was set to the $1dB$ input compression point of the amplifier modeled with the coefficient vector $\underline{h}$. σ_x was chosen such that 99% of the data symbols were below the clipping level C. We used the first 18 out of the 19 R_l functions to calculate the psd model (3.8). N_{DFT} from (3.12) was set to $N_{DFT} = 4096$. We used the frequency interval $[0.76 \cdot f_s, 0.91 \cdot f_s]$ to calculate the likelihood functions (3.12), as the high variance of the periodogram at frequencies lower than $0.76 \cdot f_s$ caused degradation of the test performance. Also, for frequencies higher than $0.91 \cdot f_s$, for the considered pulse-shaping filter, the regrowth caused by the PA nonlinearity was very small and including these frequencies did not bring performance improvements.

Figure 3.3 shows the probability of error of the hypothesis test (3.14), calculated over 250 randomly generated groups of 3 power amplifiers and input signals, as a function of the standard deviation σ_η of the zero-mean, normal random variable η (3.16), for signal-to-noise ratio (SNR) equal to $30dB$, for 50 signal records of length 1024 symbols, captured from the device when the crime was committed, and for 500 signal records of length 1024 undistorted symbols captured from the three suspected devices, for three cases:

i) while committing the crime user 1 with I/O characteristic $\underline{h}$ was not modifying the data symbols to fake the device's RF signature

ii) while committing the crime user 1 with I/O characteristic $\underline{h}$ was faking the device's RF signature by applying a different 3^{rd} order odd polynomial functions to individual data packets. The polynomials' linear coefficients were fixed to 1 and the 3^{rd} order coefficients were chosen uniformly at random from range $[0.75 \cdot k_{3,1.25}, 1.25 \cdot k_{3,1.25}]$,

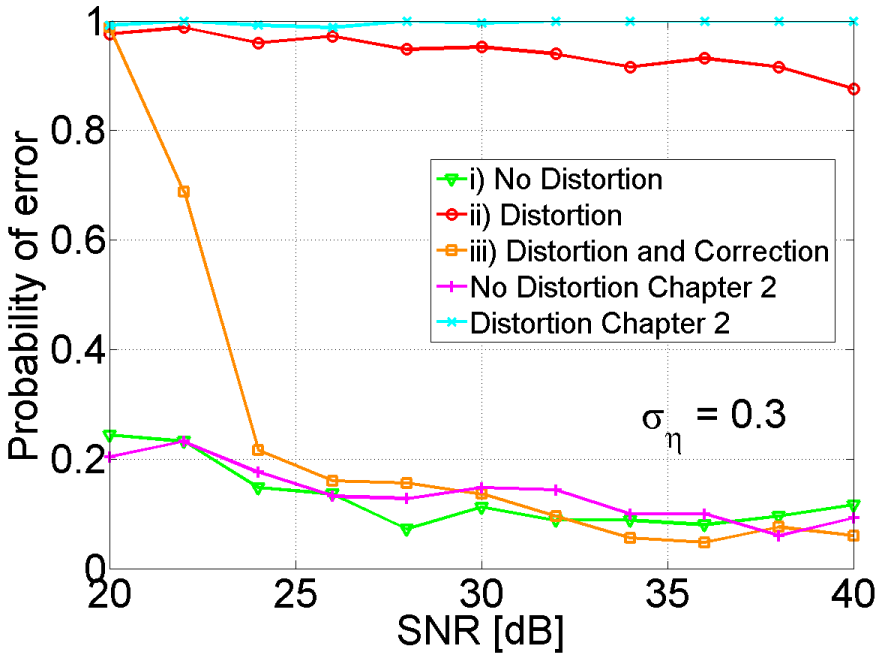

Figure 3.4. Probability of erroneous identification decision (3.14), calculated over 250 randomly generated groups of 3 power amplifiers and input signals, as a function of SNR (controlled with the noise power level), for standard deviation $\sigma_\eta = 0.3$ of the zero-mean, normal random variable η (3.16), for 50 signal records of length 1024 symbols, captured from the device used to commit a crime, and for 500 signal records of length 1024 undistorted symbols captured from the three suspected devices, for the three cases **i)**, **ii)** and **iii)**, described in Section 3.4, together with the performance of the time domain based methods of Chapter 2.

where $k_{3,1.25}$ was the 3^{rd} order coefficient of a polynomial with $1.25dB$ compression point for the data symbols

iii) while committing the crime user 1 with I/O characteristic $\underline{h}$ was faking its device's RF signature as in **ii)**, but the proposed algorithm using the corrected true values (3.15) was used to calculate the likelihood functions (3.12).

The modification of data symbols as in **ii)** resulted in degradation of $EVM = \sqrt{\frac{P_{error}}{P_{reference}}} \cdot 100\%$ by 5.51% (averaged over 10000 trials). Such EVM degradation for common modulation schemes at common SNR levels should not lead to a significant

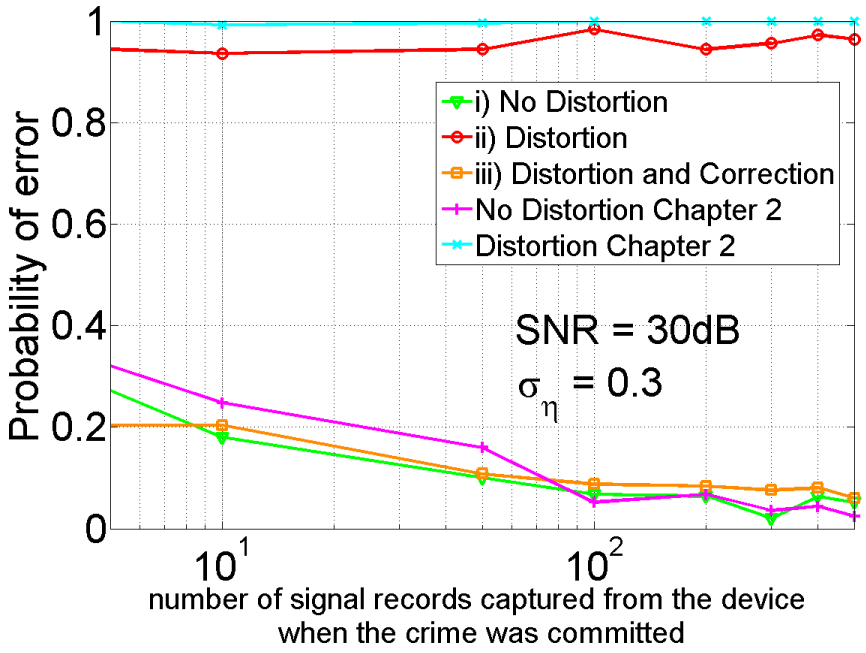

Figure 3.5. Probability of erroneous identification decision (3.14), calculated over 250 randomly generated groups of 3 power amplifiers and input signals, as a function of number of signal records of length 1024 symbols captured from the device used to commit a crime, for $SNR = 30dB$, for standard deviation $\sigma_\eta = 0.3$ of the zero-mean, normal random variable η (3.16), and for 100 signal records of length 1024 undistorted symbols captured from the three suspected devices, for the three cases **i)**, **ii)** and **iii)**, described in Section 3.4, together with the performance of the time domain based methods of Chapter 2.

increase of the bit error rates (Table IV [77]), which establishes the possibility and attractiveness of the strong adversary attack model introduced in this work.

Figure 3.4 shows the probability of error of the hypothesis test (3.14), calculated over 250 randomly generated groups of 3 power amplifiers and input signals, as a function of SNR, for standard deviation $\sigma_\eta = 0.3$ of the zero-mean, normal random variable η (3.16), for 50 signal records of length 1024 symbols, captured from the device when the crime was committed, and for 500 signal records of length 1024 undistorted symbols captured from the two suspected devices, for the three cases **i)**, **ii)** and **iii)** described above.

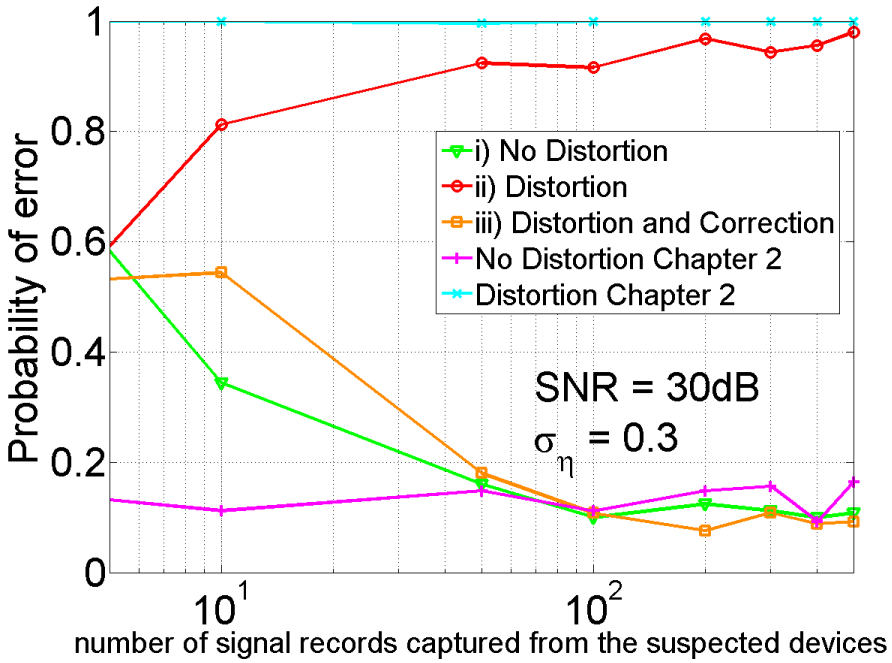

Figure 3.6. Probability of erroneous identification decision (3.14), calculated over 250 randomly generated power amplifiers and input signals, as a function of number of signal records of length 1024 undistorted symbols captured from the three suspected devices, for $SNR = 30dB$, for standard deviation $\sigma_\eta = 0.3$ of the zero-mean, normal random variable η (3.16), and for 50 signal records of length 1024 symbols captured from the device used to commit a crime, for the three cases **i)**, **ii)** and **iii)**, described in Section 3.4, together with the performance of the time domain based methods of Chapter 2.

Finally, Figures 3.5 and 3.6 show performance of the test (3.14) for fixed values of the SNR and σ_η as a function of, respectively, the number of records captured from the criminal's device and the number of records captured from the devices building the pool of suspects.

In addition, for comparison, the Figures 3.3, 3.4, 3.5 and 3.6 show performance of the time domain based methods of [78]. Figures 3.3, 3.4, 3.5 and 3.6 show that if the method proposed in this work is not employed, in the case of the considered 3-hypotheses scenario, the strong adversary is able to successfully fake the RF signature of its device by applying a simple nonlinear function to its data symbols. Spoofing

via data symbol distortion is, however, not a trivial task, as it is not straightforward to obtain the functions that, when applied to the data symbols, would modify the moments properly when the characteristics of the victims are unknown. The search for effective techniques that would give criminals the capability of faking the RF signatures of their devices via slight modifications of the data symbols is an interesting topic for future research. The method introduced in this work allows for the successful identification of the devices even if the adversary user had such a capability.

To be able to verify the utility of the proposed identification method, insight on the variations of the I/O characteristics of amplifiers used in practical applications is needed. To obtain such insight, similarly as in described in Section 2.4.2, we used a $12.5GHz$, $50GSa/s$ Tektronix $DPO71254B$ oscilloscope and Agilent Technologies $E8251A\ PSG-A$ signal generator and measured multiple points of the single-tone, amplitude I/O characteristics of eight commercial WLAN amplifiers of the same model: SKYWORKS SKY65006 [1] loaded on evaluation boards and operating at a frequency $f = 2.45GHz$. The obtained measurement points were used to approximate the amplitude I/O characteristics with 5^{th} order, odd polynomials. These approximated characteristics were then used to generate amplified data in MATLAB. Similarly as in the case of the artificially generated amplifiers, the input signal was modeled as a sequence of realizations of a zero-mean, normal random variable with standard deviation σ_x, clipped to the level C, and pulse-shaped with a raised-cosine pulse-shaping filter with a roll-off factor $r = 0.5$. The oversampling ratio $\mathcal{O}$ of the pulse-shaping filter was set to $\mathcal{O} = 4$. The clipping level $C = 0.1412$ of the input signal to the PAs was set to the upper boundary of the range specified as 802.11b frequency mask-compliant for the considered amplifiers [1]. The standard deviation σ_x was chosen such that 99% of the symbols were below the clipping level C ($\sigma_x = 0.055$). The number of undistorted signal records of length 4096 symbols captured from the devices building the pool of suspects was set to 10000. The num-

ber of signal records of length 4096 symbols captured from the device committing the crime was set to 500. N_{DFT} from (3.12) was set to 4096. The K-ary test (3.14) was used to identify each of the eight measured amplifiers. Again three cases similar to cases **i)**, **ii)** and **iii)**, described earlier, for the artificially generated amplifier pairs were considered.

Similarly, as in the case of the artificially generated amplifiers, we used the frequency interval: $[0.76 \cdot f_s, 0.91 \cdot f_s]$ to calculate the likelihood functions (3.12) and the first 18 out of the 19 R_l functions were used to calculate the psd model (3.8).

Tables 3.1, 3.2 and 3.3, for SNR values of $30dB$, $35dB$, and $40dB$ respectively, show the probability of erroneous identification decision (3.14) calculated over 250 trials for the three cases **i)**, **ii)** and **iii)**, for eight measured amplifiers. These tables show that while users of devices $3, 6, 7$ and 8 were not very successful at faking their devices' RF signatures by distorting the data symbols, the degradation of the performance of the uncorrected (3.14) was significant when users of devices $1, 2, 4$ or 5 were distorting their data, especially at high $SNRs$. Correction of the scale factors from the test (3.14) allowed for highly probable identification of these strong adversaries for the considered high SNR values, which based on measurements reported from existing WLAN deployments are reasonable for indoor short-range scenarios. In addition, for comparison, the Tables 3.1, 3.2 and 3.3 show the performance of the time domain based methods of [78].

The work presented in this paper is an extension of work presented in [78] to a scenario where the user fakes its RF signature with data symbol distortion. In [78] we provide a comparative overview of related steady state identification techniques. As we stress in [78], such a comparison is hard to conduct as approaches similar to the model based identification approach that we introduced have not yet been investigated as to the best of our knowledge. Hence in [78], we are necessarily limited to restating the experimental outcomes of related steady state identification studies,

SNR=30dB								
PA #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
i)	0.076	0.252	0.184	0.032	0.256	0.112	0.000	0.000
ii)	0.936	0.692	0.540	0.764	0.896	0.316	0.000	0.000
iii)	0.076	0.468	0.016	0.012	0.014	0.224	0.072	0.000
Ch. 2 no distortion	0.012	0.452	0.004	0.000	0.076	0.152	0.032	0.000
Ch. 2 distortion	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000

Table 3.1. Probability of erroneous identification decision (3.14) calculated over 250 trials for eight measured SKYWORKS amplifiers for $SNR = 30dB$ (controlled with the noise power level), for the three cases: **i)** user was not modifying the data symbols while committing crime; **ii)** user was distorting the data symbols while committing crime in order to fake device's RF signature; **iii)** user was distorting the data symbols while committing crime and the proposed algorithm using the corrected true values (3.15) was used to calculate the likelihood functions (3.12), together with the performance of the time domain based methods of Chapter 2.

SNR=35dB								
PA #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
i)	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000
ii)	0.996	0.764	0.156	0.944	1.000	0.012	0.000	0.000
iii)	0.000	0.240	0.000	0.000	0.000	0.052	0.000	0.000
Ch. 2 no distortion	0.000	0.064	0.012	0.000	0.092	0.000	0.000	0.000
Ch. 2 distortion	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000

Table 3.2. Probability of erroneous identification decision (3.14) calculated over 250 trials for eight measured SKYWORKS amplifiers for $SNR = 35dB$ (controlled with the noise power level), for the three cases: **i)** user was not modifying the data symbols while committing crime; **ii)** user was distorting the data symbols while committing crime in order to fake device's RF signature; **iii)** user was distorting the data symbols while committing crime and the proposed algorithm using the corrected true values (3.15) was used to calculate the likelihood functions (3.12), together with the performance of the time domain based methods of Chapter 2.

but, since these studies are largely empirical, we are only able to compare them quite roughly to the results from [78]. Thus, here we are forced to limit ourselves to the comparison to the identification methods from [78], and show significant performance improvements (Figures 3.3, 3.4, 3.5 and 3.6 and Tables 3.1, 3.2 and 3.3)

SNR=40dB								
PA #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
i)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ii)	1.000	0.676	0.008	0.908	1.000	0.000	0.000	0.000
iii)	0.000	0.252	0.000	0.000	0.000	0.000	0.000	0.000
Ch. 2 no distortion	0.000	0.020	0.000	0.000	0.008	0.000	0.000	0.000
Ch. 2 distortion	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000

Table 3.3. Probability of erroneous identification decision (3.14) calculated over 250 trials for eight measured SKYWORKS amplifiers for $SNR = 40dB$ (controlled with the noise power level), for the three cases: **i)** user was not modifying the data symbols while committing crime; **ii)** user was distorting the data symbols while committing crime in order to fake device's RF signature; **iii)** user was distorting the data symbols while committing crime and the proposed algorithm using the corrected true values (3.15) was used to calculate the likelihood functions (3.12), together with the performance of the time domain based methods of Chapter 2.

3.5 Conclusions

In this chapter, we considered the novel problem of wireless device identification for the case when strong adversaries actively fake their device's RF signature with artificial injection of a slight distortion to the data symbols. While this is unlikely for a standard adversary employing a wireless card, its potential use by strong adversaries motivates the consideration of techniques to address such. Our identification method does not require strict assumptions on the distribution of the data symbols. It is only assumed that elements of the data symbol stream are uncorrelated and have zero mean values. As shown with simulations based on parameters of commercially employed PAs, for practical SNR values the application of the proposed method allows for the prevention of the performance degradation caused by modification of the data symbols by the strong adversaries, as results are similar to those when adversaries are not sophisticated enough to modify the data. Because of the high data rates of modern communications networks, the data records that need to be captured to perform identification correspond to short observation times of the masquerading users.

Because of the fast stabilization of the operating temperature of the measured SKYWORKS PAs [1], temperature variations were ignored in the presented investigation. These variations should however be considered in future research for refinement of the proposed identification methods.

CHAPTER 4

WIRELESS DEVICE IDENTIFICATION BASED ON RF OSCILLATOR IMPERFECTIONS

4.1 Problem Statement

In Chapter 2, we considered two components of the transmitter chain: the digital-to-analog-converter (DAC) and the power amplifier (PA) and showed that exploitation of variations among PAs can lead to successful device identification at much lower SNR levels when compared to the DACs. Here, we analyze the degree to which a wireless device can be correctly identified from measurable non-idealities of RF oscillators, employed by wireless transmitters. Work presented in this chapter is motivated by the fact that, in contrast to the PAs, which in transmit power controlled applications might be switching modes over time, characteristics of the RF oscillators are power level independent, and thus can be used as unique device tags in systems with implemented transmit power control mechanisms.

In the mobile devices, RF oscillators are typically implemented as phased-locked-loops (PLLs). Figure 4.1 shows a basic block diagram of a PLL. An ideal PLL would generate a sinusoidal oscillation at a carrier frequency f_0. Instead, in practice, the PLL generates a signal of the form

$$y(t) = \cos\left(2\pi(f_{0_{PLL}} + \Delta_{PLL}(t))t + \Theta_{PLL}(t)\right), \tag{4.1}$$

where $\Delta_{PLL}(t)$ is the frequency offset and $\Theta_{PLL}(t)$ is the phase noise. The frequency offset $\Delta_{PLL}(t)$ is specific to a given PLL chip. However, it can be easily compromised

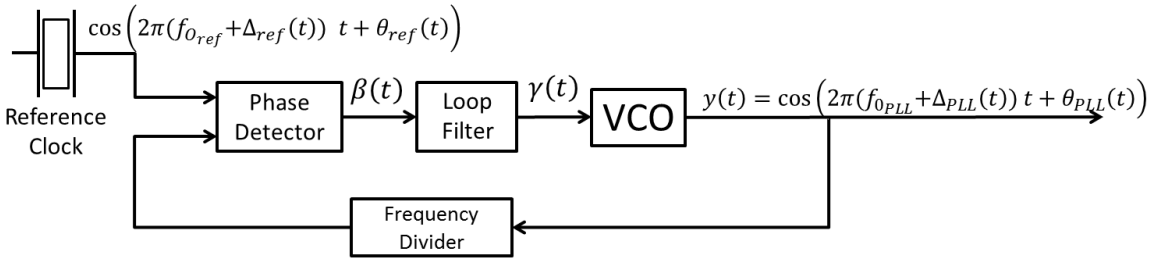

Figure 4.1. A basic PLL block diagram.

by strong adversaries via multiplication of the digital symbols with a time-varying factor, and it is also sensitive to chip temperature changes. Therefore, in this chapter we focus on the extraction of devices' RF fingerprints based on differences in the characteristics of the PLL's phase noise, which is caused by variations in the components that comprise the PLL circuit and cannot be easily modified by the user.

4.2 PLL Phase Noise Model and RF Fingerprint Extraction

For the case of free running (open loop) RF oscillators $\Theta(t)$ becomes a Wiener process as $t \to \infty$ [79]. The phase noise is then characterized with a single quality parameter that determines the width of the oscillator's spectrum, which exhibits a Lorenzian shape [80]. For PLLs, the analytic description of the phase noise is more complex. In [81], the PLL is modeled with a set of stochastic differential equations and the autocorrelation of the phase noise corrupted PLL's output $y(t)$ is found as

$$R_y(\tau) = \sum_{i=-\infty}^{\infty} X_i X_i^* \cdot \exp(-ji\omega_0\tau) \cdot \exp\left[-0.5\omega_0^2 i^2 \left[c_{ref}|\tau| + 2\sum_{l=1}^{n}(\nu_l+\mu_l)\left[1-\exp(-\lambda_l|\tau|)\right]\right]\right], \tag{4.2}$$

where X_i are coefficients of the Fourier series expansion of the PLL crystal's reference signal oscillating with nominal angular frequency ω_0, c_{ref} is a quality parameter of the crystal oscillator, and λ_l, ν_l and μ_l, $l = 1, 2, \ldots, n$, are parameters that depend on entries of matrices defining the set of differential stochastic equations modeling the PLL of order $n-1$ (Appendix of [81]). For the 1^{st} order charge pump loop filter with cut-off frequency ω_{cp} and transfer function $\frac{s+\omega_{cp}}{s}$, $n = 2$ and we find μ_l and ν_l, $l = 1, 2$:

$$\begin{aligned} \mu_1 &= c_{ref}\frac{-\lambda_2(\lambda_1-\omega_{cp})}{\omega_{cp}(\lambda_1-\lambda_2)\lambda_1} \\ \mu_2 &= c_{ref}\frac{-\lambda_1(\omega_{cp}-\lambda_2)}{\omega_{cp}(\lambda_1-\lambda_2)\lambda_2} \end{aligned} \tag{4.3}$$

$$\begin{aligned} \nu_1 &= \frac{c_{VCO}+c_{ref}}{(\lambda_1-\lambda_2)^2}\left(\frac{\lambda_2^2(\omega_{cp}-\lambda_1)^2}{2\omega_{cp}^2\lambda_1} - \frac{\frac{\lambda_1\lambda_2}{\omega_{cp}^2}\left(-\omega_{cp}^2+\omega_{cp}(\lambda_1+\lambda_2)-\lambda_1\lambda_2\right)}{2(\lambda_1+\lambda_2)}\right) \\ \nu_2 &= \frac{c_{VCO}+c_{ref}}{(\lambda_1-\lambda_2)^2}\left(\frac{\lambda_1^2(\omega_{cp}-\lambda_2)^2}{2\omega_{cp}^2\lambda_2} - \frac{\frac{\lambda_1\lambda_2}{\omega_{cp}^2}\left(-\omega_{cp}^2+\omega_{cp}(\lambda_1+\lambda_2)-\lambda_1\lambda_2\right)}{2(\lambda_1+\lambda_2)}\right). \end{aligned} \tag{4.4}$$

Typically $\lambda_1 = \lambda_2^*$, which further implies that $\mu_1 = \mu_2^*$ and $\nu_1 = \nu_2^*$. If the reference signal is generated with a high quality crystal oscillator, its Fourier series expansion can be accurately approximated with a single non-zero element. This allows for simplification of (4.2). Let $p = \mu_1 + \nu_1$ and $\lambda = \lambda_1$, then

$$\begin{aligned} R_y(\tau) &= \exp(-j\omega_0\tau)\cdot\exp\left[-0.5\omega_0^2 c_{ref}|\tau|\right] \\ &\cdot\exp\left[-\omega_0^2\left[p\cdot(1-\exp[-\lambda|\tau|]) + p^*\cdot(1-\exp[-\lambda^*|\tau|])\right]\right] \\ &= \exp(-j\omega_0\tau)\cdot\exp\left[-0.5\omega_0^2 c_{ref}|\tau|\right] \\ &\cdot\exp\left[-\omega_0^2\left[2\Re\{p\} - p\cdot\exp[-\lambda|\tau|] - p^*\cdot\exp[-\lambda^*|\tau|]\right]\right], \end{aligned} \tag{4.5}$$

and with

$$
\begin{aligned}
&p \cdot \exp[-\lambda|\tau|] + p^* \cdot \exp[-\lambda^*|\tau|] \\
&= (\Re\{p\} + j\Im\{p\}) \cdot \exp\{-\lambda|\tau|\} + (\Re\{p\} - j\Im\{p\}) \cdot \exp\{-\lambda^*|\tau|\} \\
&= \Re\{p\} \cdot \exp\{-\Re\{\lambda\}|\tau|\} \cdot 2\cos(\Im\{\lambda\}|\tau|) \\
&\quad + j\Im\{p\} \cdot \exp\{-\Re\{\lambda\}|\tau|\} \cdot -2j\sin(\Im\{\lambda\}|\tau|) \\
&= 2 \cdot \exp\{-\Re\{\lambda\}|\tau|\} \cdot (\Re\{p\} \cdot \cos(\Im\{\lambda\}|\tau|) + \Im\{p\} \cdot \sin(\Im\{\lambda\}|\tau|))
\end{aligned}
\tag{4.6}
$$

we get

$$
\begin{aligned}
&R_y(\tau) = \exp(-j\omega_0\tau) \cdot \exp\left[-0.5\omega_0^2 c_{ref}|\tau|\right] \\
&\cdot \exp\left[-2\omega_0^2\left[\Re\{p\} - \exp\left[-\Re\{\lambda\}|\tau|\right] \cdot (\Re\{p\}\cos(\Im\{\lambda\}|\tau|) + \Im\{p\}\sin(\Im\{\lambda\}|\tau|))\right]\right].
\end{aligned}
\tag{4.7}
$$

Equation (4.7) shows multiple factors that determine the dependence of the envelope of the autocorrelation function on the PLL parameters. The exponential decay factor $\exp\left[-0.5\omega_0^2 c_{ref}|\tau|\right]$ depends on the quality of the crystal oscillator. The dependence of the envelope of $R_y(\tau)$ on other PLL components, which we want to use for user identification, is most pronounced at small values of $|\tau|$, for which the $\exp\left[-0.5\omega_0^2 c_{ref}|\tau|\right]$ factor is close to 1, and hence can be neglected in the signature extraction process. The envelope $\mathcal{E}_{R_y}(\tau)$ of the autocorrelation function for small values of $|\tau|$ can thus be expressed as

$$
\begin{aligned}
&\mathcal{E}_{R_y}(\tau) = \exp\left[-2\omega_0^2\left[\Re\{p\} - \exp\left[-\Re\{\lambda\}|\tau|\right]\right.\right. \\
&\left.\cdot (\Re\{p\}\cos(\Im\{\lambda\}|\tau|) + \Im\{p\}\sin(\Im\{\lambda\}|\tau|))\right]] \\
&= \exp\left[-2\omega_0^2\left[\Re\{p\} - \exp\left[-\Re\{\lambda\}|\tau|\right]\right.\right. \\
&\cdot\sqrt{\Re\{p\}^2 + \Im\{p\}^2} \cdot \cos\left(\Im\{\lambda\}|\tau| + sgn\left(\Im\{\lambda\}\right) \cdot \arccos\left(\frac{\Re\{p\}}{\sqrt{\Re\{p\}^2 + \Im\{p\}^2}}\right)\right)\Bigg]\Bigg],
\end{aligned}
\tag{4.8}
$$

which, with $\mathcal{E}_{R_y}(0) = 1$, becomes

$$\mathcal{E}_{R_y}(\tau) = \exp\left[-2\omega_0^2 \Re\{p\}\left(1 - \exp\{-\Re\{\lambda\}|\tau|\} \cdot \cos\left(\Im\{\lambda\}|\tau|\right)\right)\right]. \tag{4.9}$$

The characteristics of the envelope of the autocorrelation function at small $|\tau|$ can be used as a unique feature identifying a given oscillator, and the parameter vector

$$F = [\Re\{p\} \quad \Re\{\lambda\} \quad \Im\{\lambda\}] \tag{4.10}$$

can be used as a unique fingerprint that directly depends on the values of the components comprising the PLL circuit. For the range of $|\tau|$ considered for the signature extraction, the undersampling rate should be chosen such that enough samples are available for an accurate signature estimate.

4.3 Identification Method

4.3.1 Distribution of the Envelope of the Sample Estimate of the Autocorrelation Function of the PLL Output

Consider an output $y(t)$ of the PLL, sampled with the frequency f_s. A sample estimate of the autocorrelation of the random process $y(t)$ calculated based on a record $y[n]$, $n = 1, \ldots, N$ can be obtained as

$$\underline{\hat{R}}_y[m] = \frac{1}{(N-m)} \sum_{n=1}^{N-k} y[n]y[n+m], \qquad m = 0, 1, \ldots, N-1. \tag{4.11}$$

Denote the autocorrelation coefficient $\rho_m = cov(y[n], y[n+m]/var(y[n]))$. If N goes to infinity, then the joint distribution of any finite set of elements of $\left(\underline{\hat{R}}_y[m]/\underline{\hat{R}}_y[0] - \rho_m\right)$

becomes jointly normally distributed with the covariance matrix W, the elements of which are defined with [82, Eq.(1.4)]:

$$w_{i,j} = \sum_{\nu=-\infty}^{\infty} \left(\rho_\nu \rho_{\nu+i-j} + \rho_\nu \rho_{\nu+i+j} + 2\rho_i \rho_j \rho_\nu^2 - 2\rho_i \rho_\nu \rho_{\nu+j} - 2\rho_j \rho_\nu \rho_{\nu+i}\right), \tag{4.12}$$

assuming that the stochastic process $y[n]$ is stationary with mean μ_y and can be represented as

$$y[n] - \mu_y = \sum_{i=-\infty}^{\infty} h[i]\epsilon[n-i], \tag{4.13}$$

where $\epsilon[n-i]$ are independently and identically distributed zero-mean, random variables with finite variances, and $\sum_{i=-\infty}^{\infty} |h[i]| < \infty$. In other words, the considered discrete random process $y[n]$ must have a representation that is equivalent to a linear, time invariant filtering of a discrete white noise process, which is the case for the modeled phase noise process as will be shown in Section 4.4.1.

Assume access to the noise-corrupted PLL output records $y[n] = p[n] + \eta[n]$ undersampled with sampling rate f_s. $\underline{\hat{R}}_y[m]$, calculated from these undersampled records, oscillates with frequency $f = \min_N |(f_0 - N f_s)|$, which because of the variations of f_0 and $\Delta(t)$ from (4.1), varies among the devices and over time. Wireless devices could potentially be identified by comparing vectors of envelopes of sample estimates of the autocorrelation functions $\underline{\mathcal{E}}_{\underline{\hat{R}}_y}[m]$ at small values of $|m|$, normally distributed with covariance matrix elements given in (1.4) [82].

For the autocorrelation function of the PLL output (4.7) $R_y(\infty) = \mu_y = 0$, however, because (4.9) was derived for small values of the time shift, $\underline{\mathcal{E}}_{\underline{R}_y}(\infty) = \exp\left[-2\omega_0^2 \Re\{p\}\right]$, where p is device dependent. Since in [82] the covariance matrix of the sample autocorrelation estimate was obtained for zero-mean stochastic processes, to calculate the elements of the covariance matrix with the infinite sums from (1.4) [82], we subtract the offset $\underline{\mathcal{E}}_{\underline{R}_y}(\infty)$, and, for the white noise corrupted PLL output records, obtain

$$\rho_m = cov(y[n], y(n + m \cdot T_s))/var(y[n])) = \frac{\underline{\mathcal{E}}_{\underline{R}_y}(|m \cdot T_s|) - \underline{\mathcal{E}}_{\underline{R}_y}(\infty) + \sigma_\eta^2 \cdot \delta[|m \cdot T_s|]}{1 - \underline{\mathcal{E}}_{\underline{R}_y}(\infty) + \sigma_\eta^2}, \tag{4.14}$$

with a unit impulse $\delta[n]$, $\underline{\mathcal{E}}_{\underline{R}_y}$ from (4.9), and $\sigma_\eta^2 = 10^{NPNR/10}$, where $NPNR$ is the ratio of the power of the white noise to that of the phase noise

$$NPNR = 10 \log_{10}(P_\eta / P_p). \tag{4.15}$$

4.3.2 Optimal Hypothesis Test

We consider first a two-device identification scenario. After the PLL output record is captured from the device on the crime scene, the two hypotheses of the identification test are $\mathcal{H}_1$: device 1 is the transmitting device; $\mathcal{H}_2$: device 2 is the transmitting device. The likelihood ratio test is

$$\Lambda = \frac{p_{\underline{\mathcal{E}}_{\hat{\underline{R}}_y}|\mathcal{H}_1}(\underline{\mathcal{E}}_{\hat{\underline{R}}_y}|\mathcal{H}_1)}{p_{\underline{\mathcal{E}}_{\hat{\underline{R}}_y}|\mathcal{H}_2}(\underline{\mathcal{E}}_{\hat{\underline{R}}_y}|\mathcal{H}_2)} \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \varsigma. \tag{4.16}$$

For equally probable hypotheses the threshold $\varsigma = 1$ minimizes the risk of the test (4.16) (pg. 26, Section 2.2 [76]). With jointly Gaussian distributed vectors $\underline{\mathcal{E}}_{\hat{\underline{R}}_y}$,

$$p(\underline{\mathcal{E}}_{\hat{\underline{R}}_y} \,|\, \mathcal{H}_k) = \frac{1}{(2\pi)^{M/2} \det\{W_{\mathcal{H}_k}\}^{1/2}} \exp\left\{ -\frac{1}{2} \left(\underline{\mathcal{E}}_{\hat{\underline{R}}_y} - \underline{\mathcal{E}}_{\hat{\underline{R}}_{y,\mathcal{H}_k}} \right)^H W_{\mathcal{H}_k}^{-1} \left(\underline{\mathcal{E}}_{\hat{\underline{R}}_y} - \underline{\mathcal{E}}_{\hat{\underline{R}}_{y,\mathcal{H}_k}} \right) \right\}, \tag{4.17}$$

where $\underline{\mathcal{E}}_{\hat{\underline{R}}_{y,\mathcal{H}_k}}$ are envelopes of accurate estimates of the autocorrelation functions obtained from the devices from the pool of suspects. The binary decision rule becomes

$$\ln\left(\det\{W_{\mathcal{H}_1}\}\right) + \left(\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_1}}\right)^H W_{\mathcal{H}_1}^{-1} \left(\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_1}}\right)$$

$$\begin{matrix} \mathcal{H}_2 \\ \gtrless \\ \mathcal{H}_1 \end{matrix} \ln\left(\det\{W_{\mathcal{H}_2}\}\right) + \left(\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_2}}\right)^H W_{\mathcal{H}_2}^{-1} \left(\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_2}}\right). \tag{4.18}$$

The two-device scenario can easily be generalized to a K-device scenario, for which the identified device k is the device for which the likelihood function takes its maximal value

$$\begin{aligned} k_{opt} = \max_{k=1,\ldots,K} p(\underline{R}_y \,|\, \mathcal{H}_k) = \\ \min_{k=1,\ldots,K} \left\{ \ln\left(\det\{W_{\mathcal{H}_k}\}\right) + \left(\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_k}}\right)^H W_{\mathcal{H}_k}^{-1} \left(\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_k}}\right) \right\}. \end{aligned} \tag{4.19}$$

The power levels of the phase noise are much below the carrier power even for inexpensive commercially used PLLs (e.g., -81dBc/Hz at 1kHz offset from the carrier for ADF4360-1 [4]). Thus the measurement noise dominates the phase noise at common SNR values. For the discrete additive white Gaussian noise (AWGN) random process, ρ_m from (4.14) is dominated by the unit impulse and the covariance matrix W becomes an identity matrix. As shown in Section 4.4, for practical SNR levels, the approximation of W from (4.18) and (4.19) with the identity matrix does not cause a noticeable degradation in identification performance. This allows for significant simplification of the decision rules (4.18) and (4.19). Respectively, for the binary scenario,

$$||\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_1}}||_2 \begin{matrix} \mathcal{H}_2 \\ \gtrless \\ \mathcal{H}_1 \end{matrix} ||\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_2}}||_2, \tag{4.20}$$

and, for the K-ary scenario,

$$k_{opt} = \min_{k=1,\ldots,K} ||\underline{\mathcal{E}}_{\underline{\hat{R}}_y} - \underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_k}}||_2. \tag{4.21}$$

4.3.3 Practical Identification Algorithm

One possible way to obtain access to the undersampled PLL output in practice is to utilize (at least one) carrier phase recovery pilot tone, which for accurate extraction of the phase noise needs to be sufficiently separated from the data tones [83]. Although not present in current standards, security is becoming a critical issue in mobile radio applications, and it is reasonable to understand the potential benefit if future communication standards provide additional tones for security level enhancements. In fact, the relative expense required decreases with the increase of bandwidth utilized by individual users and hence the cost of adding such a tone is already trending rapidly towards a negligible amount.

The autocorrelation function estimates (4.11) are calculated based on individual signal records captured from the devices over time. Out of all of the samples of the estimate of the autocorrelation function estimated based on a given signal record, only for a subset of samples do we have $\underline{\mathcal{E}}_{\underline{\hat{R}}_y} \approx \underline{\hat{R}}_y$ (samples close to the local extrema of the autocorrelation function). With a fixed sampling rate, because of variations of f_0 among devices, as well as because of the time-varying frequency offset $\Delta(t)$ from (4.1), the subsets of samples for which $\underline{\mathcal{E}}_{\underline{\hat{R}}_y} \approx \underline{\hat{R}}_y$ can be different among the devices and vary over time. Thus, to obtain accurate estimates of the fingerprint F (4.10) for each device from the pool of suspects, the envelopes of the estimates at small values of $|\tau|$ are matched to the model (4.9) through exhaustive search of the values of (4.10) for each record available from a given device and averaged over these records. The same procedure is followed for the records captured from the unknown device from the crime scene. For the device identification $p(\underline{\mathcal{E}}_{\underline{\hat{R}}_y} \,|\, \mathcal{H}_k)$ are compared. $\underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_k}}$'s and $\underline{\mathcal{E}}_{\underline{\hat{R}}_y}$ for each hypothesis and for the device from the crime scene respectively are

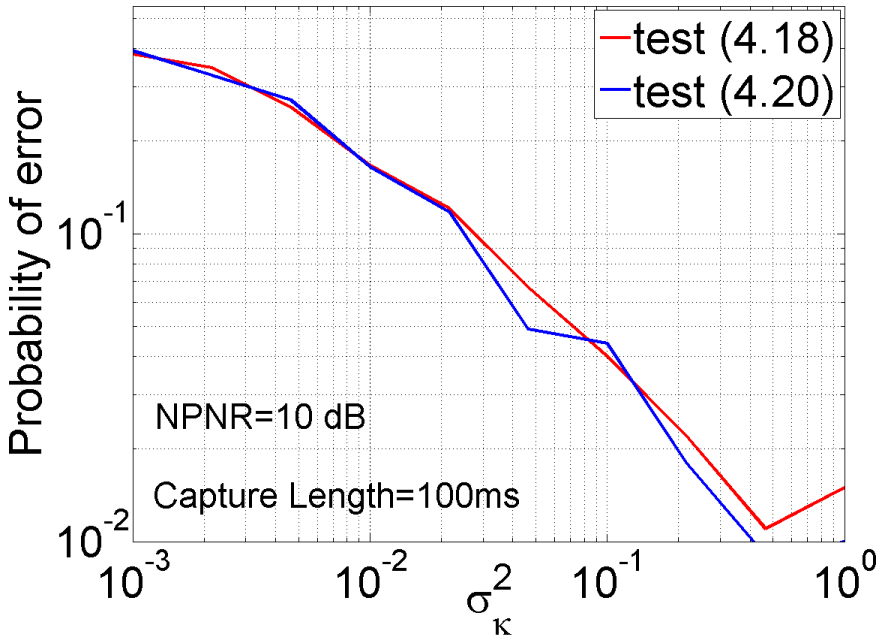

Figure 4.2. Probability of error of the binary hypothesis test (4.18) and (4.20) averaged over 500 trials for $NPNR = 10dB$ and for capture length $l_c = 100ms$ as a function of the standard deviation σ_κ used to artificially generate the oscillator pairs.

reconstructed for the same sets of $|\tau|'s$ with the fingerprint (4.10) and the model (4.9). Algorithm 2 summarizes the proposed device identification scheme.

4.4 Measurements and Numerical Results

The performance of the proposed identification method is considered here with simulations and hardware measurements. In Section 4.4.1, similarly as in [84], sample paths of the PLL phase noise are simulated by numerically solving a discrete-time version of equations set modelling a 1^{st} order charge pump PLL. Most important is Section 4.4.2, where PLL output signals are captured from commercially used PLLs and the performance of the identification method is analyzed at 15dB and 35dB SNR with records of length 200ms.

Algorithm 2 Device Identification Algorithm

For the Device that Needs to be Identified
for number of available captured records **do**
- Capture an output record of the PLL undersampled with f_s
- Calculate estimate of the capture's autocorrelation function $\underline{\hat{R}}_y[m]$ with (4.11)
- Find values of m for which $|\underline{\hat{R}}_y[m]|$ has local maxima
- Based on $(m, \underline{\hat{R}}_y[m])$ pairs and model (4.9) find the estimate of the parameter vector F (4.10) via exhaustive search

end for
- Average the estimates of the parameter vector F over the number of captured records
- Reconstruct the autocorrelation function $\underline{\mathcal{E}}_{\underline{\hat{R}}_y}$ with the model (4.9) and the averaged fingerprint F for a fixed set of $m's$

For the Devices Building the Pool of Suspects
for number of devices building the pool of suspects **do**
  for number of available captured records **do**
    - Capture an output record of the PLL undersampled with f_s
    - Calculate estimate of the capture's autocorrelation function with (4.11)
    - Find values of m for which $|\underline{\hat{R}}_y[m]|$ has local maxima
    - Based on $(m, \underline{\hat{R}}_y[m])$ pairs and model (4.9) find the estimate of the parameter vector F (4.10) via exhaustive search

  end for
  - Average the estimates of the parameter vector F over the number of captured records to obtain a very accurate fingerprint

end for
for number of hypothesis (devices from the pool of suspects) **do**
  - Reconstruct the autocorrelation function $\underline{\mathcal{E}}_{\underline{\hat{R}}_{y,\mathcal{H}_k}}$ with the model (4.9) and the averaged fingerprint F for a fixed set of $m's$

end for
- Make the identification decision based on (4.19) or (4.21)

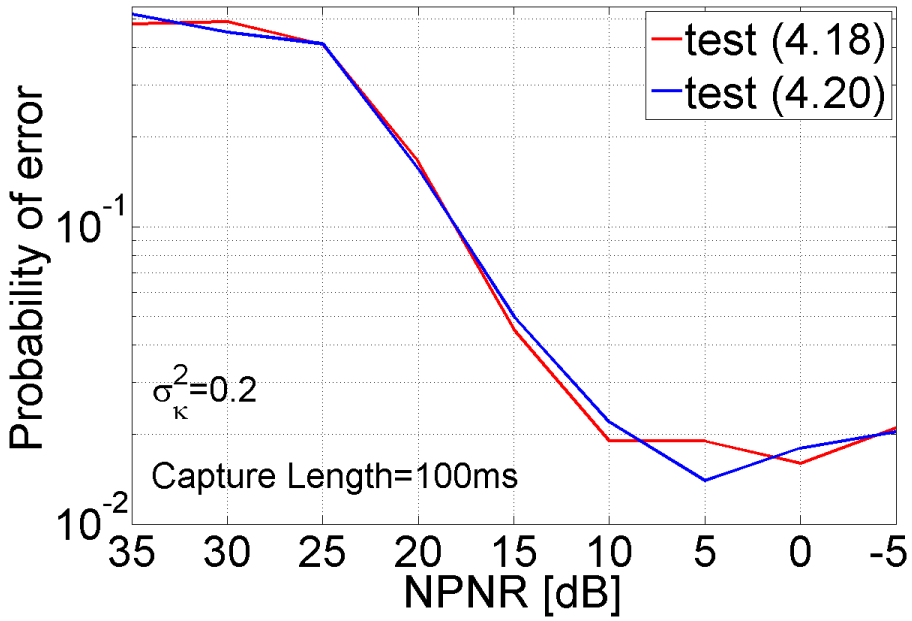

Figure 4.3. Probability of error of the binary hypothesis test (4.18) and (4.20) averaged over 500 trials for the standard deviation $\sigma_\kappa = 0.2$ used to artificially generate the oscillator pairs and for capture length $l_c = 100ms$ as a function of the $NPNR$.

4.4.1 Simulated Oscillators

We generated pairs of phase noise paths $\Theta_{PLL_k}[n], k = 1, 2$, by numerically solving a discrete-time version of the set of equations modeling a 1^{st} order, charge pump PLL [84, Eq.(8) and (13)]:

$$\beta[n] = \beta[n-1] + \sqrt{c_{contr}}\Delta t\gamma[n-1] + \sqrt{c_{VCO}} \cdot \eta_{VCO}[n] - \sqrt{c_{ref}} \cdot \eta_{ref}[n], \quad (4.22)$$

$$\gamma[n] = (1-\omega_{G_{PLL}}\Delta t)\gamma[n-1] - k_{pd}\omega_{cp}\Delta t\beta[n-1] - k_{pd}(\sqrt{c_{VCO}}\cdot\eta_{VCO}[n] + \sqrt{c_{ref}}\cdot\eta_{ref}[n]). \quad (4.23)$$

The generated phase noise path of the PLL is then

$$\Theta_{PLL}[n] = \left(\beta[n] + \sqrt{c_{ref}} \cdot \eta_{ref}[n]\right) \cdot 2\pi f_{0_{PLL}}. \quad (4.24)$$

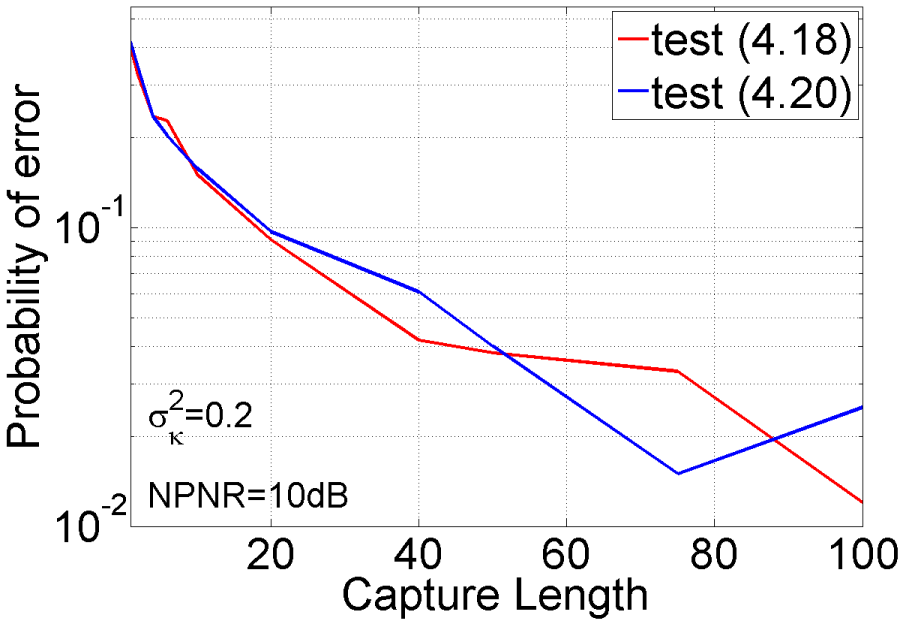

Figure 4.4. Probability of error of the binary hypothesis test (4.18) and (4.20) averaged over 500 trials for the standard deviation $\sigma_\kappa = 0.2$ used to artificially generate the oscillator pairs and for $NPNR = 10dB$ as a function the capture length.

β and γ in (4.22) and (4.23) are, respectively, output of the phase detector and input to the voltage controlled oscillator (Figure 4.1). η_{VCO} and η_{ref} are independent discrete white Gaussian noise processes $\sim \mathcal{N}(0, \Delta t)$. We set Δt to $0.04\mu s$ (sampling rate $f_s = 25Msps$) and $f_{0_{PLL}}$ to 2.4GHz. We generated the parameters used to generate each of the paths of a given pair randomly by multiplying nominal values of the parameters defined in [84]: quality parameters of, respectively, the voltage controlled and crystal oscillators $c_{VCO} = 15 \cdot 10^{-19}$ and $c_{ref} = 10^{-25}$; cut-off frequency of the PLL structure $\omega_{G_{PLL}} = 2\pi \cdot 10^4$; cut-off frequency of the charge pump $\omega_{cp} = 2\pi \cdot 16 \cdot 10^3$; phase detector gain $k_{pd} = 1$ and $\sqrt{c_{contr}} = \omega_{G_{PLL}/k_{pd}}$, with a factor $(1 + |\kappa|)$, where $\kappa \sim \mathcal{N}(0, \sigma_\kappa)$. We then generated the third path (potential capture from the crime scene) using the first set of parameters. We then added white Gaussian noise with elements $\eta[n] \sim \mathcal{N}(0, \sigma_\eta)$

to the three phase noise paths and estimated the autocorrelation functions from the white noise corrupted phase noise paths $y_k[n] = \Theta_k[n] + \eta_k[n], k = 1, 2, 3$.

An oscillation of a frequency $f_{0_{PLL}}$ under-sampled with a sampling rate f_s oscillates with a frequency $f = \min_N |(f_{0_{PLL}} - Nf_s)|, N \in \mathbb{Z}$. Since for the artificial phase noise path generation, as described above, the choice of $f_{0_{PLL}}$ and f_s is arbitrary, $f_{0_{PLL}}$ and f_s, as well as relative time shift between the sampling sequence and the PLL output were chosen such that the undersampled output of the PLL and the phase noise path were equivalent.

To show that the PLL output $y[n]$ can be represented in the form (4.13), and hence to justify application of the tests from Section 4.3.2, we write the z-transforms of the coupled Equations (4.22) and (4.23)

$$\mathcal{B}(z) \cdot (1 - z^{-1}) = c_1 \Gamma(z) \cdot z^{-1} + H_1(z) \cdot E(z), \tag{4.25}$$

$$\Gamma(z) \cdot (1 - c_2 z^{-1}) = c_3 \mathcal{B}(z) \cdot z^{-1} + H_2(z) \cdot E(z), \tag{4.26}$$

where c_1, c_2 and c_3 are constants $c_1 = \sqrt{c_{control}\Delta t}$, $c_2 = (1 - \omega_{G_{PLL}}\Delta t)$, $c_3 = -k_{pd}\omega_{cp}\Delta t$ and $E(z)$ is a z-transform of the white noise sequence $\epsilon[n]$ (4.13). With (4.25) and (4.26) one can write

$$\Gamma[z] = H(z) \cdot E(z), \tag{4.27}$$

where

$$H(z) = \frac{z\left(zH_2(z) + c_3 H_1(z) - H_2(z)\right)}{z^2 - (1 + c_2)z + (c_2 - c_1 c_3)}. \tag{4.28}$$

For the parameter values used in the simulations both of the complex poles of the transfer function (4.28) had amplitudes smaller than one, which is equivalent with absolute summability of the corresponding impulse response and implies that the simulated phase noise process has the representation (4.13).

Figures 4.2, 4.3 and 4.4, respectively, show the probability of error P_{err} of the binary hypothesis tests (4.18) and (4.20) averaged over 500 trials as a function of the

standard deviation σ_κ used to artificially generate the oscillator pairs (Figure 4.2); as a function of the additive white noise power to the phase noise power ratio $NPNR$ (4.15) (Figure 4.3); and as a function the capture length (Figure 4.4). The region of the autocorrelation function employed was $\tau \in (0.01,\ 0.15)ms$. Covariance matrices $W_{\mathcal{H}_k}$ from (4.18) were calculated with (4.9), (4.12) and (4.14) with the assumption of known oscillators' parameters. Plots from Figures 4.2, 4.3 and 4.4 show a potential for effective device identification based of oscillator non-idealities, even if only a single capture from the devices building the pool of suspects and from the device that needs to be identified is available; however, the variation of component values was generated quite artificially; hence, hardware measurements are critical. These are provided in the next section.

4.4.2 Measured Oscillators

After the qualitative performance analysis from Section 4.4.1, we analyzed the effectiveness of the proposed technique for the case of commercially employed PLLs. We considered the most challenging identification scenario, when the PLL's that need to be told apart are of the same model and from the same manufacturer. We measured eight Analog Devices ADF4360-1 [4] oscillators, oscillating at $f_0 = 2.4$GHz, on a Tektronix $DPO71254B$ oscilloscope. We captured 50 output records of length 200ms sampled with $f_s = 62.5$Msps for each of the PLLs.

Table 4.1 shows the probability of identification error averaged over 250 trials for all possible pairs from the group of 8 measured oscillators for the test (4.20) at $SNR = 15$dB (lower left, below the diagonal) and at $SNR = 35$ dB (upper right, above the diagonal), when all 50 captured records were used to extract the fingerprints (4.10), and a single record, randomly chosen from the group of all 50 captured records, was used as a capture from the crime scene. The region of the autocorrelation function employed was $\tau \in (0,\ 0.075)ms$. Increase of the SNR

PLL #	1	2	3	4	5	6	7	8
1	-	0.000	0.000	0.000	0.000	0.000	0.016	0.000
2	0.000	-	0.000	0.164	0.000	0.000	0.000	0.000
3	0.000	0.000	-	0.000	0.216	0.000	0.000	0.000
4	0.000	0.228	0.000	-	0.000	0.000	0.000	0.000
5	0.000	0.000	0.228	0.000	-	0.000	0.000	0.020
6	0.000	0.000	0.000	0.000	0.000	-	0.000	0.008
7	0.008	0.000	0.000	0.000	0.000	0.000	-	0.000
8	0.000	0.000	0.000	0.000	0.028	0.036	0.000	-

Table 4.1. Probability of device identification error for test (4.20), averaged over 250 trials, for all possible pairs from the group of 8 measured oscillators at $SNR =$ 15dB (lower left, below the diagonal) and at $SNR = 35$dB (upper right, above the diagonal), when all 50 captured records of length $12.5 \cdot 10^6$ samples were used to extract the fingerprints of the devices from the pool of suspects, and a single record from the crime scene, randomly chosen from the group of all 50 captured records, was used for the identification.

did not bring significant performance improvement, as error floors emerged for some pairs of oscillators. Low error probabilities from Table 4.1 justify application of the proposed identification method for establishing probable cause and make it attractive for cyber-crime investigations.

4.4.3 Influence of Aging of the PLLs on Identification Performance

It is known that aging can lead to changes in the PLL performance over long time periods. Changes in the nominal frequency shift and character of the phase noise can be expected as the elements of the PLL circuit age. To verify the influence of aging on the character of the envelope of the autocorrelation function, and thus on the character of the tags that we want to use for identification, we re-measured the PLL outputs of the eight considered PLLs three months after the original measurements used to obtain the results reported in Section 4.4.2 were taken. In a potential crime investigation described in Chapter 1, the difference between the time of the capture from the crime scene and the time of investigation would not exceed three months in

most cases. Figure 4.5 shows the envelopes of the autocorrelation function calculated based on the original measurements and on the measurements from three months later. Different colors correspond to different oscillators. As shown in the Figure 4.5, the character of the envelopes is very similar when calculated based on captures from three months apart.

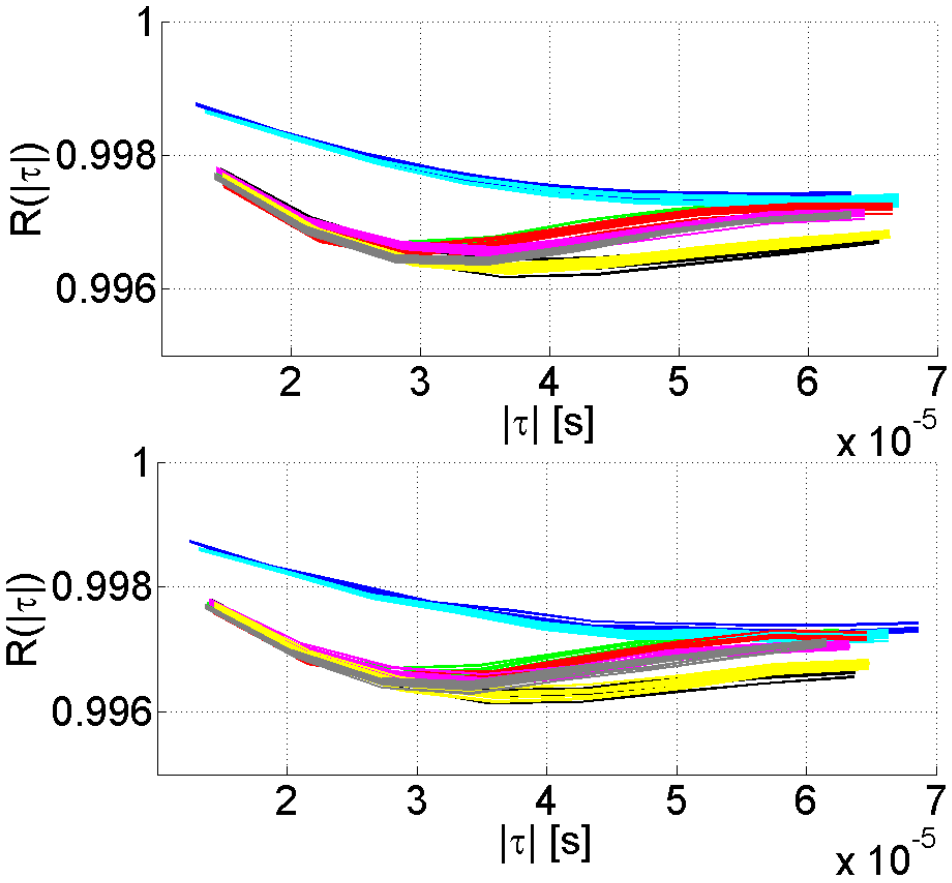

Figure 4.5. Envelopes of sample autocorrelation functions calculated for 8 measured Analog Devices ADF4360-1 oscillators [4] for 10 original captures, used to obtain the results reported in Section 4.4.2 (upper plot) and for 10 captures from 3 months after the original captures were taken (bottom plot). Colors correspond to different oscillators.

Having these two sets of captures, we re-ran our identification algorithm such that captures from the new measurement set were used to extract the fingerprints of the devices from the pool of suspects, and a single record, randomly chosen from the

PLL #	1	2	3	4	5	6	7	8
1	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	-	0.00	0.27	0.00	0.00	0.00	0.00
3	0.00	0.00	-	0.00	0.27	0.00	0.00	0.01
4	0.00	0.32	0.00	-	0.00	0.00	0.00	0.00
5	0.00	0.00	0.24	0.00	-	0.00	0.000	0.01
6	0.00	0.00	0.00	0.00	0.00	-	0.00	0.07
7	0.01	0.00	0.00	0.00	0.00	0.00	-	0.00
8	0.00	0.00	0.00	0.00	0.02	0.04	0.00	-

Table 4.2. Probability of device identification error for test (4.20), averaged over 100 trials, for all possible pairs from the group of 8 measured oscillators at $SNR =$ 15dB (lower left, below the diagonal) and at $SNR = 35$dB (upper right, above the diagonal), when all 10 new captured records of length $12.5 \cdot 10^6$ samples were used to extract the fingerprints of the devices from the pool of suspects, and a single record, randomly chosen from the group of 50 old captured records, was used as a capture from the crime scene.

group of old captured records, was used as a capture from the crime scene. Table 4.2 shows the probability of identification error averaged over 100 trials for all possible pairs from the group of 8 measured oscillators at $SNR = 15$dB (lower left, below the diagonal) and at $SNR = 35$ dB (upper right, above the diagonal). Tables 4.2 and 4.1 show only a very slight identification performance degradation, which presumably was caused by aging.

Some variation in the performance of the oscillators over time can also be expected because of chips' temperature changes. We performed measurements to examine how the frequency of the considered PLLs changes over time after the chips are powered up. Figure 4.6 shows the measured frequencies of two oscillators randomly picked from the group of the eight considered oscillators, measured every 10 minutes over 1.5h after the power-up (upper plot), as well as measured every 15 seconds over the first five minutes after the power-up (bottom plot). We see that for the considered chips, the changes of the carrier frequency over time, which might be caused by temperature changes of the chip after power-up, are in the order of tens of Hz. This

is a very slight variation that is much smaller than the frequency delta between the oscillators. This makes us believe that the influence of the temperature variation on the PLL performance can be ignored in the initial investigation. Refinement of the algorithm by taking into consideration these variations is, however, an interesting topic for future research.

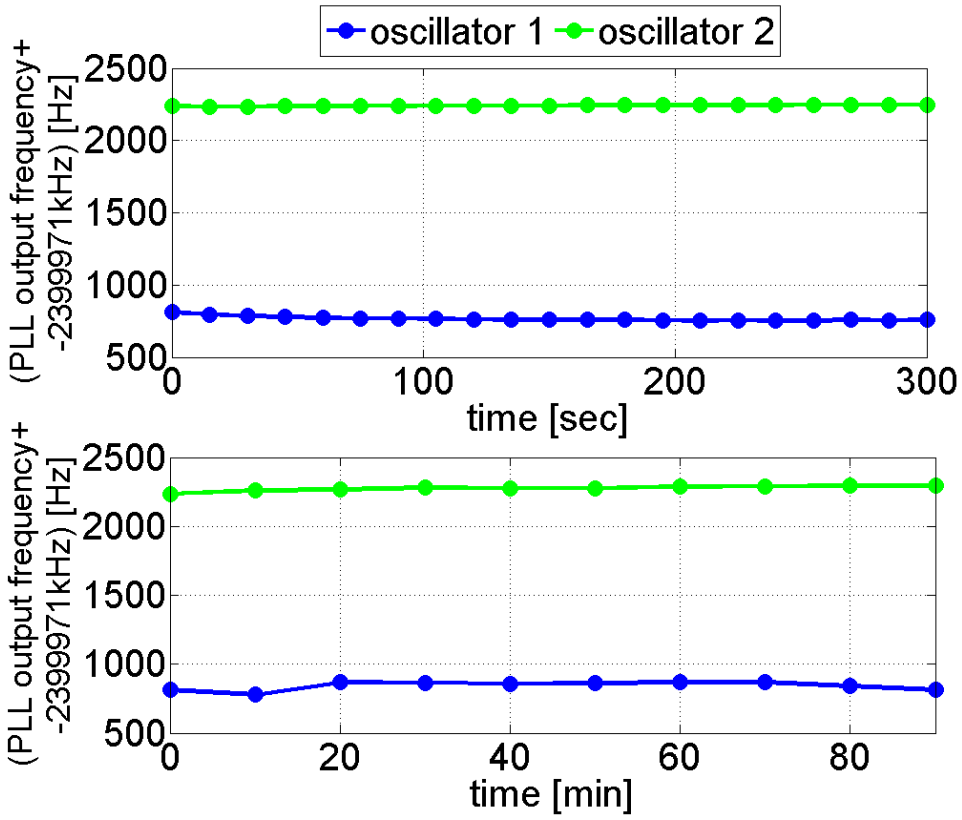

Figure 4.6. Measured frequencies of two oscillators randomly picked from the group of the eight considered oscillators, measured every 10 minutes over 1.5h after the power-up (upper plot), as well measured every 15sec over first 5 minutes after the power-up (bottom plot).

4.5 Conclusions

In this chapter, we analyzed the degree to which a wireless device can be identified from unique characteristics of the phase noise of the transmitter's RF oscillator.

Measurements of commercially used chips indicate that oscillators can be identified at practical $SNRs$ and with short observed sequences to the accuracy required to establish probable cause. The extension to higher-order PLL models that more accurately match characteristic of commercial PLLs could lead to improvement of the identification performance. Among the topics for future research are the consideration of frequency and temperature dependence of the characteristics of the phase noise. While the first is not critical, as the access point can assign devices that need to be identified to arbitrary frequency channels, the latter should be an important consideration for further refinement of the identification methods.

CHAPTER 5

PERFORMANCE BOUNDS FOR GROUPED INCOHERENT MEASUREMENTS IN COMPRESSIVE SENSING

5.1 Problem Statement

As discussed in Section 1.2.2, in this chapter we study the requirements on the number of measurements needed for successful recovery of sparse signals when, instead of individually random measurements, as assumed by conventional compressive sensing [56], the measurements are taken uniformly at random in pre-defined, non-overlapping *groups* of equal size. Such a *grouped* sampling scheme can be of practical interest for multiple application, e.g. application of interference-robust, wide-band interferer, or applications of medical imaging and remote sensing.

Consider a wide-band receiver, receiving a message of interest $m(t)$ together with an in-band, powerful and known interferer $i(t)$ that saturates the receiver's front-end and is uncorrelated with the message of interest. For undistorted recovery of the message of interest, samples can only be taken at times when the interferer's amplitude values are small. Figures 5.1 and 5.3 visualize such interference-dependent sampling schemes. In Figure 5.1 the samples of $m(t)$ are taken in *groups* of size 5 around the zero-crossings of the interferer's waveform. While the zero-crossings of the uncorrelated interferer are random with respect to the message of interest, the remaining elements of the *groups* are chosen in a deterministic way as the adjacent samples. Given a band-limited interference, the samples around the zero-crossing are small, and hence the nonlinear distortion is avoided. Such a *grouped* sampling strategy is similar to a strategy, where the entire sampling space (all possible sampling times

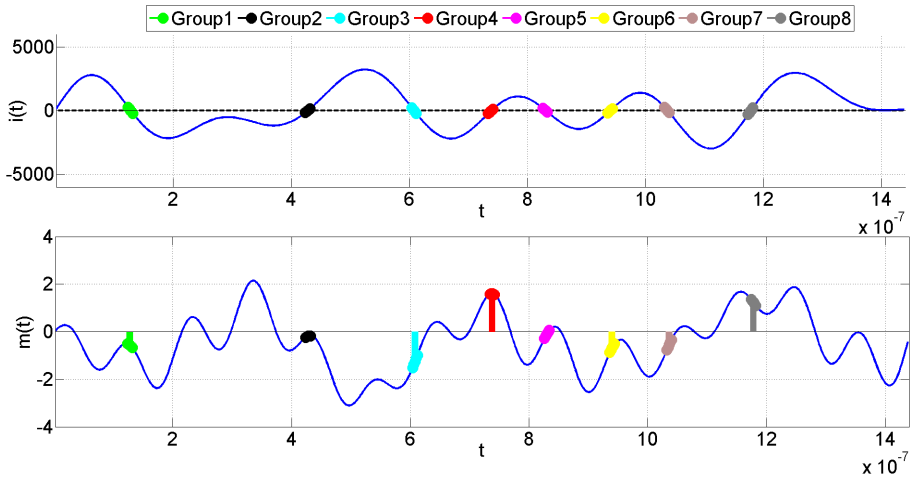

Figure 5.1. A powerful interferer $i(t)$ (upper subplot) and a message of interest $m(t)$ (lower subplot) sampled around the zero-crossings of the interferer.

Figure 5.2. Exemplary partition of the entire one dimensional sampling space into non-overlapping groups of equal size.

of $m(t)$) is partitioned into a set of pre-defined, non-overlapping *groups* of equal size. The groups of samples, instead of individual samples, are then drawn at random. A *grouping* structure that partitions a 30 samples-long message block into 6 groups of size 5 is visualized in Figure 5.2.

Different ways of partitioning the sampling space that can lead to different performance of the *grouped* sampling are, of course, possible. As an example, consider a *grouping* structure dictated by a powerful interferer $i(t)$ that repeats periodically over the duration of the block of the signal of interest $m(t)$. While the zero-crossing times during the first period of the interferer are random with respect to the message of interest $m(t)$, the zero-crossing times during the consecutive interferer's periods are shifted by the multiples of the length of the period. Figure 5.3 visualizes a sampling

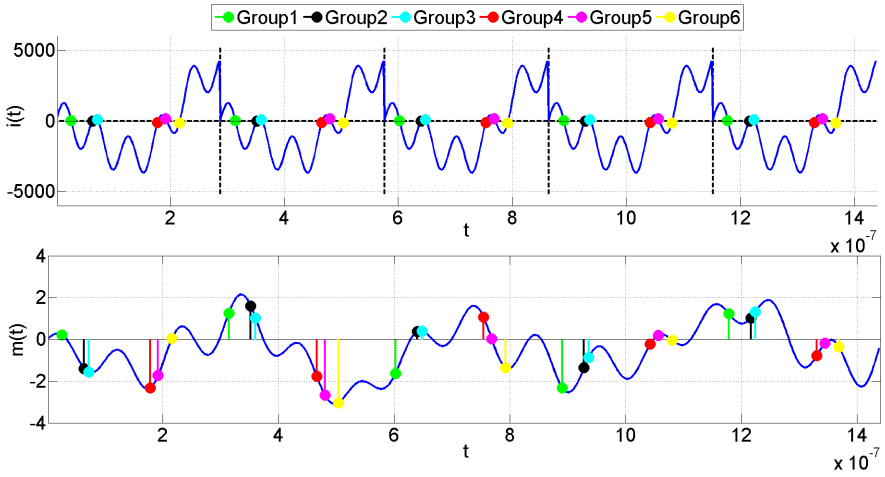

Figure 5.3. Periodically repeating powerful interferer $i(t)$ (upper subplot) and a message of interest $m(t)$ (lower subplot) sampled at times when amplitude values of $i(t)$ are closest to zero.

Figure 5.4. Exemplary partition of the entire one dimensional sampling space into non-overlapping groups of equal size.

scheme dictated by a powerful interferer repeating periodically 5 times over the duration of the message signal block, when only a single sample around the zero-crossings of $i(t)$ are kept for the recovery of $m(t)$. For such a sampling scheme, each *group* of samples consists of 5 samples taken at dependent sampling times and the entire sampling space (all possible sampling times of $m(t)$) is divided into a set of pre-defined, non-overlapping *groups* of samples of equal size, as visualized in Figure 5.4 for a 30 samples-long message block and an interferer with a period 5 times shorter than the message block.

For the applications of medical imaging and remote sensing it is common for the sensors to follow pre-defined trajectories during the acquisition process. Consider Magnetic Resonance Imaging (MRI) [85], where the measurements in the 2-D Fourier

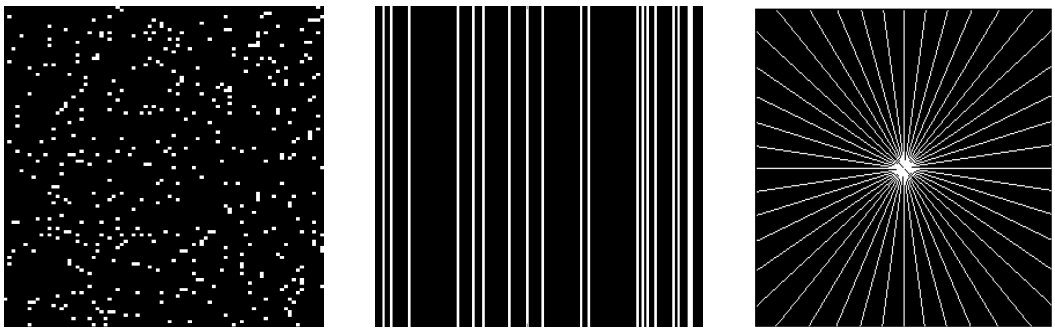

Figure 5.5. Left: Independently random 2-D sampling. Middle: Vertical line trajectories used for MRI. Right: Radial acquisition trajectories used for MRI. The trajectories group measurement selections into slices of the 2-D Fourier domain.

Figure 5.6. Exemplary remote sensing applications, for which sensors follow predefined trajectories.

space cannot be taken at random, as illustrated in the left subplot of Figure 5.5, but need to follow relatively smooth sampling trajectories to satisfy hardware and physiological constraints [86]. Two such trajectories: vertical lines and radial lines are visualized in the Figure 5.5 (middle and right subplots respectively). For compressive sensing MRI, in order to obtain a *group* of samples from the measurement space the tomographic scanner is first *randomly* oriented, and then a number of samples along a smooth trajectory is acquired. Many practical trajectories, such as the vertical lines from Figure 5.5 or trajectories considered in Section 5.3.2, partition the 2D-Fourier sampling space into a set of non-overlapping, pre-defined *groups* of equal size. For such trajectories, in this chapter we introduce a metric that upper bounds

the multiplicative penalty on the number of required measurements, with respect to the conventional compressive sensing, where individual measurements are taken uniformly at random. This factor can then be used to indicate trajectories that lead to signal recovery from smaller subset of measurements, and hence lead to a reduction of acquisition time and cost.

Similarly to the medical imaging applications, in the remote sensing applications structure is introduced to the measurement scheme, as the sensors follow specific trajectories (Figure 5.6). The cost of moving the sensor can be significant for many remote sensing applications, therefore recovery of a sparse signals from *grouped* measurements is of high practical interest. As an example of a practical application consider recovery of a wavelet domain sparse image of the bottom of the ocean from measurements taken from a ship that is moved randomly to different positions. After the sensor (the ship) is randomly positioned, it is cost-effective to take a number of local measurements in addition to the random measurement. Performance bounds for signal recovery from random grouped measurements derived in this chapter can help to make a decision on how to partition a spacial area of interest into groups in order to reduce the total number of measurements required for successful recovery.

5.1.1 Compressive Sensing Background

Consider the acquisition of an $N \times 1$ signal vector $\underline{x}$. Assume that $\underline{x}$ is known to be sparse in some basis; that is, we say the signal $\underline{x}$ is K-sparse for some integer K if $\underline{x}$ has a representation $\underline{c} = U^H \underline{x}$ having only K non-zero entries in some known orthonormal basis U, although the value and location of those non-zero entries may be unknown. In the compressive sensing framework, we acquire the $M \times 1$ output $\underline{y} = \Phi \underline{x}$, for some $M \ll N$, where Φ is the measurement matrix. According to CS theory, given certain constraints on Φ and M, $\underline{x}$ can be reconstructed from $\underline{y}$ with high probability.

5.1.2 Incoherent Measurements

Given an orthonormal measurement basis V, a K-sparse signal $\underline{x} = U\underline{c}$, sparse in some known orthonormal basis U can be reconstructed successfully from a set of M independently drawn random samples $\Omega \subseteq \{1, \ldots, N\}$ of $y = V^H U\underline{c}$ with probability not lower than $1 - \delta$, for any $1 > \delta > 0$, as long as the number of samples is large enough. Define $A = V^H U$ and denote by A_Ω the matrix built from the M rows of A corresponding to the index set Ω. Define a coherence parameter $\mu(A)$ of the matrix A as

$$\mu(A) = \max_{i,j} |A(i,j)|, \tag{5.1}$$

which has range $\mu(A) \in [\frac{1}{\sqrt{N}}, 1]$ [56]. A pair of bases V and U for which the minimal value of $\mu(A)$ is achieved is referred to as a *perfectly incoherent* pair of bases.

When the elements of Ω are drawn independently at random, it can be shown that the number M of measurements required for successful recovery of sparse $\underline{x}$ depends on the coherence parameter $\mu(A)$ (5.1).

Theorem 1. *[56] Let A be an $N \times N$ orthogonal matrix ($A^H A = I$) with coherence parameter $\mu(A)$. Fix an arbitrary subset T of the signal domain. Choose a subset Ω of the measurement domain of size $|\Omega| = M$ and a sign sequence z on T, both uniformly at random over all possible choices. Suppose that*

$$M \geq Const \cdot N\mu^2(A)|T| \log(N/\delta). \tag{5.2}$$

Then with probability exceeding $1 - \delta$, every signal $\underline{c}_0$ supported on T with signs matching z can be recovered from $y = A_\Omega \underline{c}_0$ by solving the linear program

$$\min_{\underline{c}} ||\underline{c}||_1 \qquad s.t. \qquad A_\Omega \underline{c} = A_\Omega \underline{c}_0. \tag{5.3}$$

Theorem 1 shows that the number of measurements required for successful recovery of a sparse signal scales linearly with the signal's sparsity, but only logarithmically with its length, as long as V and U are perfectly incoherent.

5.1.3 Grouped Incoherent Measurements

In certain applications, the assumptions of Theorem 1 are violated as measurements must be taken in groups instead of independently at random. More specifically, divide the set of N rows of A into N/g disjoint groups G_i, $i = 1, \ldots, N/g$, of size g each. Note that it will still be possible to take a set of measurements Ω for a signal, following Theorem 1, by selecting M/g groups out of the N/g groups available, independently at random.[1] We say that such a process provides a *grouped incoherent measurement scheme*. Grouped incoherent measurement schemes can be seen as a generalization of the standard incoherent measurement scheme used in Theorem 1 by setting $g = 1$. As discussed in Chapter 1, signal acquisitions applications for which the samples are drawn at random in predefined groups include interference robust wide-band receiver, medical imaging and remote sensing applications. In the following sections, we introduce the penalty factor on the number of required measurements due to the *grouped* measurement scheme and derive recovery guarantees.

5.2 Performance Analysis for Grouped Incoherent Measurements

5.2.1 Performance Metric

The grouped incoherent measurement scheme introduced in Section 5.1.3 violates the assumptions of Theorem 1 and causes an increase of the number of measurements needed for successful recovery of sparse signals. Such a penalty factor depends on the

[1] We assume that g divides both M and N for simplicity.

structure of the groups $G = \{G_1, \ldots, G_{N/g}\}$, on the product of the measurement and transformation basis $A = V^H U$, and on the set T defining the sparse signal support. We define a penalty factor

$$\gamma(A, T, G) = \max_{i \in 1,\ldots,N/g} \left\| \overline{A_{G_i T}} \right\|_{2 \to 1}, \tag{5.4}$$

where $\|M\|_{p \to q} = \max_f \|Mf\|_q / \|f\|_p$ denotes the $p \to q$ operator norm of the matrix M, $\overline{M}$ denotes the matrix M after row normalization, and $A_{G_i T}$ is the submatrix of A that preserves the g rows corresponding to the group G_i and the $|T|$ columns corresponding to the sparsity set T. Given the set T defining the sparse support, the penalty factor $\gamma(A, T, G)$ is a measure of similarity among the rows of $A_{G_i T}$ for each i. For example, if the rows of $A_{G_i T}$ are equal for some i, we will have $\gamma(A, T, G) = g$; in contrast, if all rows of $A_{G_i T}$ are mutually orthogonal for each i, then we will have $\gamma(A, T, G) = \sqrt{g}$. Figure 5.7 shows the structure of the matrix $A_{\Omega,T}$ obtained by drawing two out of four groups for two example grouping structures G^1 and G^2; here $N = 16$ and $g = 4$. In the case of G^1, groups are build out of samples that are separated by N/g and spread over the entire sample space, whereas in the case of G^2, groups are built out of adjacent samples.

5.2.2 Recovery Guarantees

We now provide requirements on the number of measurements needed for successful recovery of the sparse signal $\underline{x}$ when the subset Ω of the measurement domain is built out of predefined measurement groups.

Theorem 2. *Let A be an $N \times N$ orthogonal matrix ($A^H A = I$) with coherence parameter $\mu(A)$. Fix an arbitrary subset T of the signal domain. Choose a subset Ω of the measurement domain of size $|\Omega| = M$ as the union of M/g groups from*

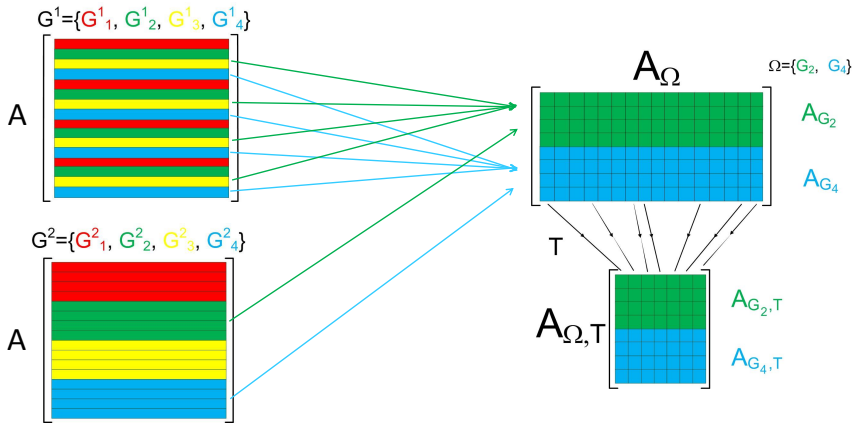

Figure 5.7. Visualization of the structure of the $A_{\Omega,T}$ matrix, for $N = 16$ and $g = 4$, obtained by drawing two out of four groups, for two example grouping structures G^1 and G^2. In the case of G^1, groups are build out of samples that are separated by N/g and spread over the entire sample space, whereas the in the case of G^2, groups are built out of adjacent samples.

$G = \{G_1, \ldots, G_{M/g}\}$ *and a sign sequence* z *on* T*, both uniformly at random over all possible choices. Suppose that*

$$M \geq \gamma(A, T, G) \cdot Const \cdot \mu^3(A) N^{3/2} |T| \log(N/\delta). \tag{5.5}$$

Then with probability exceeding $1 - \delta$*, every signal* $\underline{c}_0$ *supported on* T *with signs matching* z *can be recovered from* $y = A_\Omega \underline{c}_0$ *by solving the linear program (5.3), for any* $1 > \delta > 0$*.*

The theorem shows that for a perfectly incoherent measurement and sparsity bases, $\gamma(A, T, G)$ provides a multiplicative penalty on the number of measurements necessary for successful signal recovery due to the grouped structure of the incoherent measurement selection. Note that for a group size $g = 1$ and for perfectly incoherent pair of bases V and U our result coincides with Theorem 1 as it is equivalent to drawing elements of Ω uniformly at random.

Proof. In the following, we will prove the result of Theorem 2 with a small modification on the distribution of the submatrices: instead of a uniform distribution among all subsets Ω containing M/g out of the N/g available groups, we pose an independent Bernoulli selection for each group submatrix G_i, $i = 1, \ldots, N/g$, belonging in Ω with selection probability $P(\delta_i = 1) = M/N$. This independent model results in the expected number of selected groups being equal to M/g. Furthermore, one can show that since the probability of failure is a non-increasing function of the size M of the set Ω, the probability of failure under the uniform distribution used in Theorem 2 is upper-bounded by a constant times the probability of failure under the independent selection model used in the proof (a property dubbed *poissonization* in [56]). Thus, the effect of the conversion of the subgroup selection model is a constant multiplicative factor in the required number of measurements, which is accounted for by the constants in (5.2) and (5.5).

Following the argument of [56], one can show that the signal $\underline{c}_0$ is the unique solution to (5.3) if and only if there exists a dual vector $\pi \in \mathbb{R}^N$ that has following properties:

- π is in the row space of A_Ω,
- $\pi(t) = \text{sign}\{\underline{c}_0(t)\}$ for $t \in T$,
- $|\pi(t)| < 1$ for $t \in T^c$.

As in [56], we consider the candidate

$$\pi = A_\Omega^H A_{\Omega T} (A_{\Omega T}^H A_{\Omega T})^{-1} z_0, \tag{5.6}$$

where z_0 is a $|T|$-dimensional vector whose entries are the signs of $\underline{c}_0$ on T. To prove Theorem 2 we need to show that under its hypothesis: (i) $A_{\Omega T}^H A_{\Omega T}$ is invertible and (ii) $|\pi(t)| < 1$ for $t \in T^c$. We begin by showing that $A_{\Omega T}^H A_{\Omega T}$ is invertible with

high probability given the requirement (5.5) on M. The following theorem is proven in Appendix A and shows that if M is large enough then, on average, the matrix $A_{\Omega T}^H A_{\Omega T}$ does not deviate much from $\frac{M}{N} I$, where I is the identity matrix.

Theorem 3. *Fix an arbitrary subset T of the signal domain. Define N/g index groups $G = \{G_1, \ldots, G_{N/g}\}$ of the measurement domain, each of size g, and draw each group independently at random with probability M/N into a set Ω. If*

$$M \geq \frac{28}{3} \cdot \gamma(A, T, G) \cdot N \cdot \mu^2(A) \cdot |T| \log\left(\frac{|T|}{\delta}\right), \tag{5.7}$$

with $\gamma(A, T, G)$ introduced in (5.4), then

$$P\left(\left\|\frac{N}{M} A_{\Omega T}^H A_{\Omega T} - I\right\| \geq \frac{1}{2}\right) < \delta, \tag{5.8}$$

where $\|\cdot\|$ denotes the spectral norm

$$\|Y\| = \sup_{\|f_1\|_2 = \|f_2\|_2 = 1} |\langle f_1, Y_i f_2 \rangle|.$$

Theorem 3 shows that if M is large enough, then $A_{\Omega T}^H A_{\Omega T}$ is invertible with high probability. We continue by proving that $|\pi(t)| < 1$ for $t \in T^c$. Following the techniques in [56], we use the following three lemmas, proven in the Appendices B, C and D.

Lemma 1. *Denote by v^0 a row of the matrix $A_\Omega^H A_{\Omega T}$ indexed by $t_0 \in T^C$. Then*

$$E||v^0||^2 < \frac{M}{\sqrt{N}} \mu^3(A) |T| \gamma. \tag{5.9}$$

Lemma 2. *Define*

$$\bar{\sigma}^2 := \gamma\mu^2(A)\frac{M}{N}\cdot\max\left\{1, \sqrt{\gamma}\mu^{3/2}(A)N^{3/4}|T|/\sqrt{M}\right\}. \tag{5.10}$$

For $0 < a \leq \frac{\sqrt{M}}{\mu(A)\sqrt{N\gamma|T|}}$ *if* $\frac{\sqrt{\gamma}\mu^{3/2}N^{3/4}(A)|T|}{\sqrt{M}} < 1$

and $0 < a \leq \left(\frac{M}{\gamma\mu(A)\sqrt{N}}\right)^{1/4}$ *if* $\frac{\sqrt{\gamma}\mu^{3/2}N^{3/4}(A)|T|}{\sqrt{M}} \geq 1,$

we have

$$P\left(||v^0|| > \mu^{3/2}(A)N^{-1/4}\sqrt{\gamma M|T|} + a\bar{\sigma}\right) < 3e^{-\kappa a^2} \tag{5.11}$$

for some positive constant κ.

Lemma 3. *Let* $w^0 = (A^H_{\Omega T}A_{\Omega T})^{-1}v^0$. *With the notations and assumptions of Lemma 2 we have:*

$$P\left(\sup_{t_0\in T^c}||w^0|| \geq 2N^{3/4}\mu^{3/2}\sqrt{\frac{\gamma|T|}{M}} + \frac{2Na\bar{\sigma}}{M}\right) \leq 3e^{-\kappa a^2} + P\left(||A^H_{\Omega T}A_{\Omega T}|| \leq \frac{M}{2N}\right). \tag{5.12}$$

Finally we will use [56, Lemma 3.4], reproduced below.

Lemma 4. *Assume that* $z(t)$, $t \in T$ *is an i.i.d. sequence of symmetric Bernoulli random variables. For each* $\lambda > 0$, *we have*

$$P\left(\sup_{t\in T^c}|\pi(t)| > 1\right) \leq 2Ne^{-1/2\lambda^2} + P\left(\sup_{t\in T^c}||w^0|| > \lambda\right). \tag{5.13}$$

Now that all lemmas are in place, we are ready to prove Theorem 2. If we pick $\lambda = 2N^{3/4}\mu^{3/2}\sqrt{\gamma|T|/M} + 2Na\bar{\sigma}/M$ in (5.13), from (5.12) and (5.13) we get

$$P\left(\sup_{t\in T^c}|\pi(t)| > 1\right) \quad \leq \quad 2Ne^{-1/2\lambda^2} + Ne^{-\kappa a^2} + P\left(||A^H_{\Omega T}A_{\Omega T}|| \leq M/2N)\right).$$

For the right hand side of (5.14) to be smaller than 3δ we need all three summands to be smaller than δ. We now derive conditions on δ that provide this guarantee. We start with the second summand: for it to be no bigger than δ we can set a^2 to be

$$a^2 = \kappa^{-1} \log(N/\delta). \tag{5.14}$$

For the first summand to be no bigger than δ, we need

$$\frac{1}{\lambda^2} \geq 2 \log(2N/\delta). \tag{5.15}$$

If $\frac{\sqrt{\gamma}\mu^{3/2}(A)N^{3/4}|T|}{\sqrt{M}} > 1$, Lemma 2 requires

$$0 < a \leq \left(\frac{M}{\gamma \mu(A) \sqrt{N}} \right)^{1/4}. \tag{5.16}$$

Then with $\bar{\sigma}^2$ from (5.10) we get

$$Na\bar{\sigma}/M \leq \mu^{3/2} N^{3/4} \sqrt{\gamma |T|/M}, \tag{5.17}$$

and so

$$\lambda \leq 4\mu^{3/2}(A) N^{3/4} \sqrt{\gamma |T|/M}. \tag{5.18}$$

Reorganizing terms, we obtain

$$\frac{1}{\lambda^2} \geq \frac{M}{16\mu^3(A) N^{3/2} \gamma |T|}. \tag{5.19}$$

From (5.14) and (5.16) we get the following bound on M:

$$M \geq \gamma \mu(A) \sqrt{N} \kappa^{-2} \log^2(N/\delta). \tag{5.20}$$

Suppose now that $\frac{\sqrt{\gamma}\mu^{3/2}(A)N^{3/4}|T|}{\sqrt{M}} < 1$. Then, with (5.10), $\bar{\sigma}^2 = \gamma\mu^2(A)\frac{M}{N}$. If $\mu(A)\sqrt{N}|T| \geq a^2$, then

$$Na\bar{\sigma}/M \leq \mu^{3/2}N^{-1/4}\sqrt{\gamma|T|/M}, \tag{5.21}$$

and

$$\lambda \leq 4\mu^{3/2}(A)N^{3/4}\sqrt{\gamma|T|/M}, \tag{5.22}$$

and thus

$$\frac{1}{\lambda^2} \geq \frac{M}{16\mu^3(A)N^{3/2}\gamma|T|}, \tag{5.23}$$

which matches the previous condition (5.19). On the other hand, if $\mu(A)\sqrt{N}|T| \leq a^2$ then

$$Na\bar{\sigma}/M \geq \mu^{3/2}N^{-1/4}\sqrt{\gamma|T|/M}, \tag{5.24}$$

and

$$\lambda \leq 4Na\bar{\sigma}/M, \tag{5.25}$$

and thus, with $\bar{\sigma}^2 = \gamma\mu^2(A)\frac{M}{N}$,

$$\frac{1}{\lambda^2} \geq \frac{M^2}{16N^2a^2\bar{\sigma}^2} = \frac{M}{16a^2\gamma\mu^2(A)N}. \tag{5.26}$$

And so with (5.23) and (5.26) we can write

$$\frac{M}{16\gamma\mu^2(A)N}\min\left(\frac{1}{\mu(A)N^{1/2}|T|}, \frac{1}{a^2}\right) \geq 2\log(2N/\delta), \tag{5.27}$$

$$M \geq 16\gamma\mu^2(A)N\max\left(\mu(A)N^{1/2}|T|, a^2\right)2\log(2N/\delta), \tag{5.28}$$

which with (5.14) gives

$$M \geq \text{Const}\cdot\gamma\mu^2(A)N\max\left(\mu(A)N^{1/2}|T|, \log\left(\frac{N}{\delta}\right)\right)\log\left(\frac{N}{\delta}\right). \tag{5.29}$$

Due to Theorem 3, for the third summand to be smaller than δ, we need

$$M \geq \frac{28}{3} \cdot \gamma \cdot N \cdot \mu^2(A) \cdot |T| \log\left(\frac{|T|}{\delta}\right). \tag{5.30}$$

Thus from (5.20), (5.29) and (5.30) we see that the overall requirement on M is:

$$M \geq \text{Const} \cdot \gamma(A, T, G) \cdot \mu^3(A) N^{3/2} |T| \log(N/\delta), \tag{5.31}$$

which finishes the proof of the Theorem 2. □

5.2.3 Calculation of the Performance Metric

For a fixed sparsity set T, we can obtain lower and upper bounds on the value of $\gamma(A, T, G)$ by leveraging the Pietsch Factorization theorem [87].

Theorem 4. *Each matrix B can be factored as $B = FD$ where D is a nonnegative, diagonal matrix with trace$(D^2) = 1$ and $\|B\|_{\infty\to 2} \leq \|F\|_2 \leq K_p \|B\|_{\infty\to 2}$, where K_p is a constant equal to $\sqrt{\frac{\pi}{2}} \approx 1.25$ for the real field and $\sqrt{\frac{4}{\pi}} \approx 1.13$ for the complex field.*

Since $\|M\|_{2\to 1} = \|M^H\|_{\infty\to 2}$, thanks to the duality of the operator norms, we can find bounds on γ by performing Pietsch factorization of the matrices $(\overline{A_{G_i T}})^H = F_i D_i$, for $i = 1, \ldots, N/g$, where D_i is a nonnegative diagonal matrix with trace$(D_i^2) = 1$. The value of $\gamma(A, T, G)$ can then be bounded by

$$\frac{1}{K_p} \max_i ||F_i||_2 \leq \gamma(A, T, G) \leq \max_i ||F_i||_2, \tag{5.32}$$

The Pietsch factorization of matrix B can be performed by solving a semidefinite program [87].

5.3 Simulations

In this section, we present simulation results that justify the utility of the penalty factor γ (5.4) as an indicator of the recovery performance of different group structures for the grouped incoherent measurement scheme. First, one-dimensional Fourier sparse signals are considered. Next, we present the dependency of the recovery performance on the penalty factor for multiple different grouping structures for images.

5.3.1 Fourier-Domain Sparse 1-D Signals

We generate discrete signals s of length $N = 1100$ and sparsity $|T| = 5\% \cdot N$, sparse in the frequency domain, generated as a product of an orthonormal Fourier basis of size $N \times N$ and a sparse coefficient vector $\underline{c}$ with values of non-zero entries distributed uniformly: $\sim \mathcal{U}(-1, 1)$. We evaluate two different configurations for the grouped incoherent measurements:

- G^1: 100 groups of size 11 were constructed such that the first sample of each of the groups was chosen out of the first 100 samples of s: $\{s[1], \ldots, s[n]\}$, and the remaining 10 samples for each group were shifted with respect to the first sample by multiples of 100. More specifically, $G_i^1 = \{i, i + 100, i + 200, \ldots, i + 1000\}$. This configuration appears in the interference-robust compressive wide-band receiver application. The first sample corresponds to a random zero-crossing of a modulated interferer. Additional samples correspond to subsequent zero-crossings of the interferer's carrier.

- G^2: 100 groups of size 11 were constructed such that each group contained 11 consecutive, adjacent samples. More specifically, $G_i^2 = \{s[i+(i-1)\cdot 11] : s[i\cdot 11]\}$. Such configuration assumes that the samples are taken in sequential bursts.

Figure 5.8 shows the relation between the penalty factor $\gamma(A, T, G)$ from (5.4) and the ratio between the number M of samples required for successful recovery for the two described group structures and the number of samples M_0 required for successful

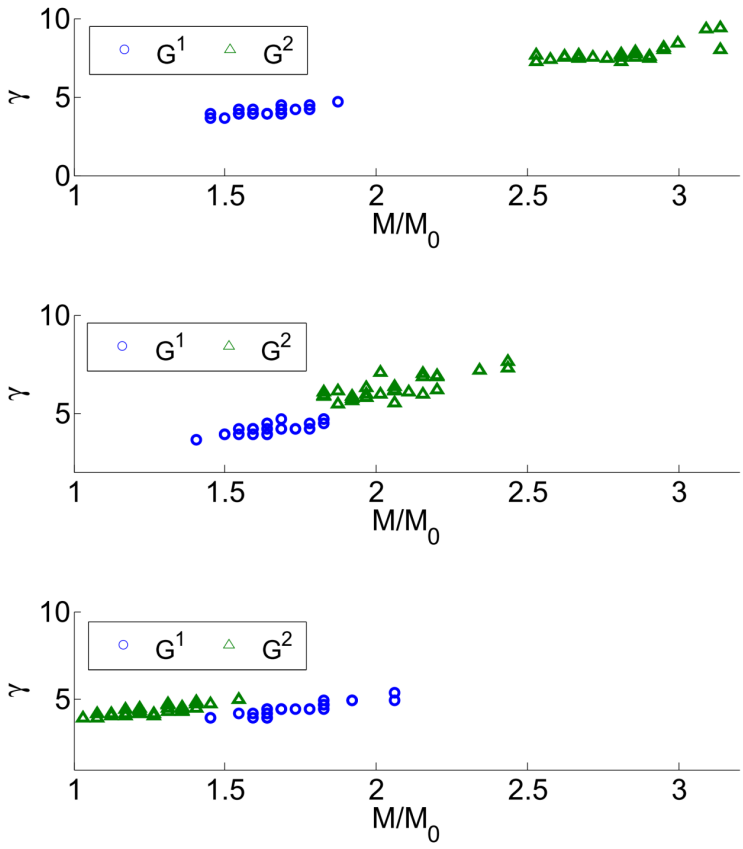

Figure 5.8. γ versus M/M_0 for group structures G^1 and G^2 for different concentrations of the nonzero Fourier coefficients of a 5% sparse signal s. Top: a sub-band built out of two 5%-wide channels; middle: a sub-band built out of four 5%-wide channels; bottom: the entire band.

recovery for random sampling. The values shown are the minimal number of measurements needed to obtain normalized recovery error $NRE = \|s - \hat{s}\|/\|s\| < 0.001$ for 99 out of 100 draws of the measurement groups (uniformly at random) and the values of the Fourier coefficients (from $\mathcal{U}[-1, 1]$).[2] Each point of the scatter plots corresponds to a fixed signal support. We consider three different classes of signal supports: for the first two classes, the positions of the non-zero Fourier coefficients

[2]Throughout this section, the SPGL1 solver [88, 89] was used for recovery, while the CVX optimization package [90] was used to solve a semidefinite program [87] for Pietsch factorization of the matrices $(\overline{A_{G_i T}})^H$ and subsequent calculation of the penalty factors $\gamma(A, T, G)$.

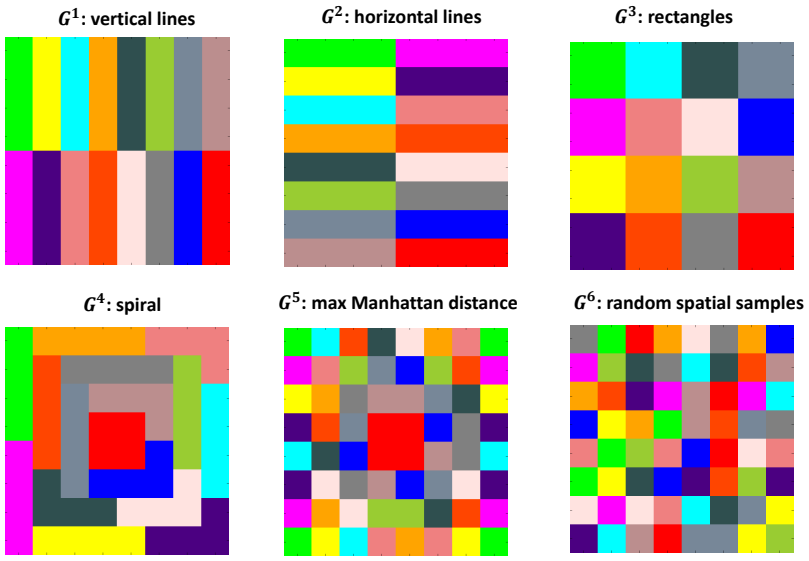

Figure 5.9. Illustration of tested group structures for 8×8-pixel images and for a group size $g = 4$, where elements of the same group are marked with the same color.

are chosen uniformly at random within a sub-band built out of two and four 5%-wide channels, respectively, positioned uniformly at random within the entire frequency band; we then compare their performance against the baseline of signals with unrestricted sparse supports. Figure 5.8 shows that for the first two classes γ was a good performance indicator; in contrast, for the last class the values of γ misleadingly suggest that both group structures perform equally well.

5.3.2 Wavelet domain sparse 2-D signals

Next, we consider the recovery of images from grouped measurements. For different measurement trajectories (group structures), we use the penalty factor to assess the suitability of different group measurement structures to obtain successful recovery with the least number of measurements. We consider six different 2-D group structures:

- G^1: vertical lines;

- G^2: horizontal lines;
- G^3: $g/2 \times 2$ rectangles;
- G^4: spiral;
- G^5: maximal Manhattan distance; and
- G^6: groups build out of random spacial samples.

Figure 5.9 shows the structures for 8×8-pixel images and for a group size $g = 4$, where elements of the same group are marked with the same color. The group structure G^5 was constructed as follows: the upper left pixel was chosen as the first element of the first group, and successive elements of the group were chosen from the remaining pixels to maximize the total Manhattan distance between the new element and the existing elements of the group. After all elements of the group were chosen, a new group was constructed starting with the pixel closest to the top left corner among those remaining, following the same procedure as the first group afterwards; this procedure was repeated for all other groups.

The suitability of the penalty factor γ as an indicator of the performance of different 2-D group measurement structures was evaluated with two sets of experiments. The first experiment evaluates grouped spatial measurements. The second experiment evaluates grouped frequency-domain measurements that emulate MRI acquisition.

5.3.2.1 Recovery of Satellite Terrain Images

The images used in the first experiment were taken from a satellite terrain image of areas around the town of Amherst, MA that was obtained from Google Maps. 25 low-resolution (32×32 pixels) tiles were gray-scaled and compressed using wavelet transform coding to 51 coefficients. We study the recovery of these images from grouped pixel measurements under configurations G^1-G^6 with groups of size $g = 8$. Figure 5.10 shows the relationship between the penalty factor $\gamma(A, T, G)$ and

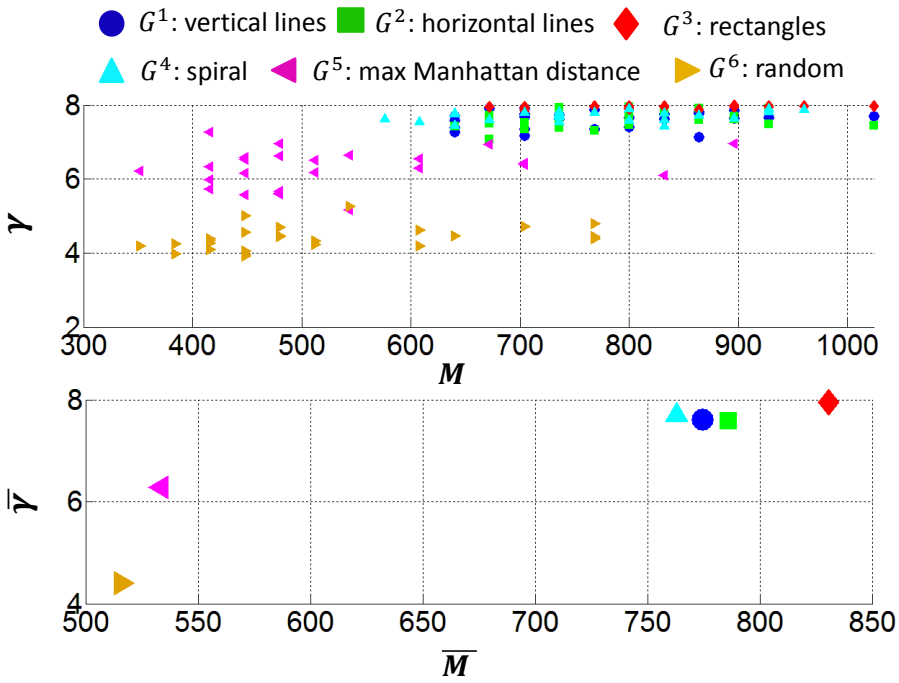

Figure 5.10. Top: relationship of M versus γ for the six considered group structures, for 25 low-resolution (32×32 pixels) compressed images from a satellite terrain images of areas around the town of Amherst; bottom: average value of γ and M, averaged over the 25 segments.

the number M of samples required for successful recovery for each of the six group structures from Figure 5.9. Each point of the top scatter plot corresponds to a single 32×32-pixel tile, while each point of the bottom scatter plot shows the average values of γ and M, over all of the tiles, for each of the grouped measurement configuration. In these experiments, recovery success is defined by a normalized recovery error $NRE = \|s - \hat{s}\|/\|s\| < 0.1$ for 49 out of 50 draws of the measurement groups, uniformly at random. The values of M tested are multiples of $4 \cdot g = 32$.

Figure 5.10 shows how the value of $\gamma(A, T, G)$ increases as a function of the number of measurements M required for successful recovery until it reaches its maximal value $\gamma = g = 8$ for the group structure G^3. The Figure shows that the metric γ can be a useful indicator of the performance for group structures of practical interest. The

Figure 5.11. 160×160-pixel chest MRI image used in the experiment.

metric indicates a superior performance of the randomized sampling structure G^6, as well as the Manhattan distance-based group structure G^5, both of which bear out in practice. Out of the four group structures G^1, G^2, G^3 and G^4, characterized with continuous measurement trajectories, G^3 exhibited the worst performance, and the highest value of the penalty $\gamma(A, T, G)$. The recovery performance, as well as the value of $\gamma(A, T, G)$, was very similar for group structures G^1, G^2 and G^4. Despite similar performances for group structures G^5 and G^6 a certain level of variation of the γ factor was observable. This is indicative of the potential looseness of the bound provided by Theorem 2. We believe that such looseness is characteristic of guarantees that rely on worst-case metrics, such as the coherence parameter $\mu(A)$ from (5.1) and our metric $\gamma(A, T, G)$, and is compounded by the looseness in the estimate of $\gamma(A, T, G)$ obtained via Theorem 4 (of up to $1/\sqrt{\pi/2} \approx 21\%$).

5.3.2.2 Recovery of MRI Images

In the second experiment, we study the recovery of MRI images from grouped measurements taken in the Fourier domain. 25 small-scale (32×32 pixels) images

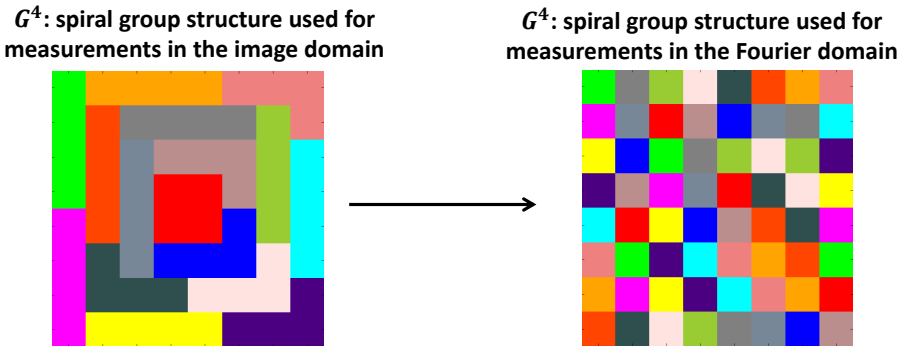

Figure 5.12. Grouped measurement structure G^4 used in the MRI experiments.

were obtained as segments of an 160×160 pixels chest MRI image from Figure 5.11 and compressed using wavelet transform coding to 51 coefficients. The group size was again set to $g = 8$. For the MRI experiments, the spiral group structure G^4 shown in Figure 5.9, where adjacent measurements form a spiral trajectory, was replaced with a structure where adjacent measurements in the same spiral trajectory are assigned to the different groups lexicographically and cyclically. For such a grouping structure, the measurements contributing to a given group were spread across the spectrum of the considered 2-D signal – including both low and high-frequency measurements in each group. Figure 5.12 visualizes the new grouping structure G^4 for the Fourier measurement domain of size 8×8 and for a group size $g = 4$.

Figure 5.13 shows the relationship between the penalty factor $\gamma(A, T, G)$ and the number M of samples required for successful recovery for each of the six aforementioned group structures. Each point of the top scatter plot corresponds to a single 32×32-pixel tile, while each point of the bottom scatter plot shows the average values of γ and M, over all of the tiles, for each of the grouped measurement configuration. In these experiments, recovery success is defined by a normalized recovery error $NRE = \|s - \hat{s}\|/\|s\| < 0.1$ for 19 out of 20 draws of the measurement groups, uniformly at random. The values of M tested are once again multiples of $4 \cdot g = 32$. The figure shows that while the group structures G^1, G^2, G^3 and G^5 demonstrate sim-

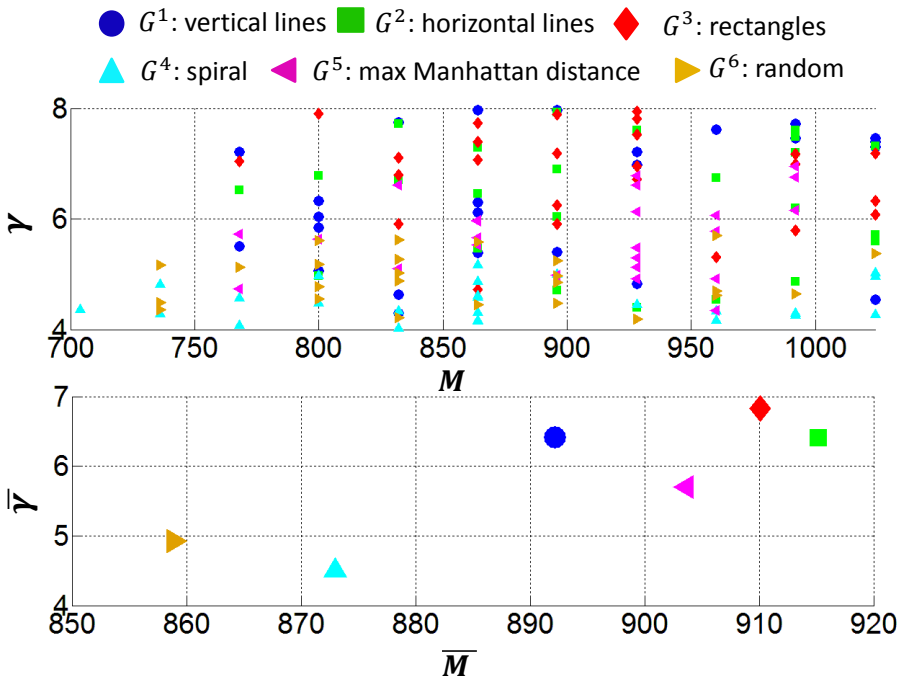

Figure 5.13. Top: relationship of M versus γ for the six considered group structures, for 25 small-scale (32×32 pixels) compressed images from a 160×160-pixel chest MRI image (cf. Figure 5.11); bottom: average value of γ and M, averaged over the 25 segments.

ilar performance and values of γ, the group structure G^4 and the randomized group structure G^5 exhibit smaller values of γ and lead to lower requirements on the number of measurements, which suggest the utility of γ as a performance indicator for the Fourier domain grouped sampling schemes.

5.4 Conclusions

In this chapter, we presented an analytically derived multiplicative penalty on the number of measurements needed for compressive sensing recovery when the measurements exhibit grouped structure instead of the usual independently drawn measurement assumption taken by most existing CS literature. Such grouped sampling is of large practical interest as full randomization of measurements is difficult to achieve

in many compressive sensing acquisition systems. We showed the utility of the introduced penalty factor as an indicator of the performance for acquisition scenarios of practical interest. A notable limitation of the introduced penalty factor γ is that it is dependent on the signal support.

CHAPTER 6

RECOVERY OF SPARSE SIGNALS FROM AMPLITUDE-LIMITED SAMPLE SETS

6.1 Problem Statement

In Chapter 5 we studied the recovery of sparse signals from *grouped* measurements, where the measurement scheme was independent from the signal that needed to be recovered (it was dictated by physical constraints of the acquisition system). In this chapter, we study compressive sensing recovery of sparse signals from irregular *signal dependent* samples. In particular, we study recovery from samples taken only when the amplitude of the signal that needs to be recovered is small; that is, we attempt to recover the signal using samples with values within a range $[-\tau, \tau]$. The small-amplitude signal sampling approach introduced in this chapter falls in the field of signal dependent non-uniform sampling. Early and significant work on signal-dependent sampling was done by Logan [91], who established sufficient conditions for the zero-crossings of a signal to uniquely determine it. Existing practical recovery algorithms from the zero-crossing information are, however, known to be unstable. Boufounos and Baraniuk [92] introduced an additional signal sparsity assumption to gain robustness in signal recovery from zero-crossings. Recovery of frequency-sparse signals from non-zero level crossings as well as from multiple level crossings has been addressed recently by Sharma and Sreenivas [93]. Our work is significantly different from [92, 93]: instead of sampling non-uniformly at the times when the signal crosses predefined levels, we consider sampling the signal uniformly at high sampling rates and then selecting only the samples whose amplitudes are below a given threshold

τ, while discarding potentially nonlinearly distorted samples with values that exceed the threshold.

Perhaps the prior contribution most closely related to the work presented is the recent independent work of [94], which also considers the recovery of frequency-sparse signals from a reduced set of samples. The sample subselection in [94] is driven by signal clipping; the resulting algorithms that account for clipping are similar to those we discuss here. However, in contrast to [94], we study the performance of CS with the proposed algorithms as a function of the amplitude of the threshold that samples must meet to be deemed suitable for undistorted signal recovery. Such threshold τ controls the fraction of samples that are used in recovery, cf. Figure 6.4. Our results show the impact that different options to leverage sample selection information during recovery have on CS performance.

Consider a frequency sparse signal s, with an unknown support, captured at the receiver. s needs to be recovered from a set of its small-amplitude samples. For that purpose, we solve a linear program

$$\hat{S} = \arg\min_{\bar{S}} ||\bar{S}||_1 \quad \text{s.t.} \quad A\bar{S} = AS, \tag{6.1}$$

where $S = \mathcal{F}s$ is the Fourier representation of s, $A = M\mathcal{F}^H$ is a transformation matrix, M is a measurement matrix, and $\mathcal{F}^H$ is the Hermitian conjugate of the Fourier matrix $\mathcal{F}$. The probability of recovery error P_{err} is defined as the probability of the normalized recovery error

$$NRE = \|s - \hat{s}\|_2 / \|s\|_2 \tag{6.2}$$

being above a target value ρ.

The characteristics of the measurement matrix depend on the measurement scheme. In order to reduce the sampling rate of analog-to-digital converters (ADCs), individ-

ual measurements can be built as a linear combination of multiple time samples [95]. We assume that the measurement matrix M is built out of rows of an identity matrix that correspond to the indices of small-amplitude samples.

It is well known [56] that if the time samples are taken uniformly at random, then the recovery guarantees of compressive sensing are independent from the support of the frequency-sparse signal that needs to be recovered. When the uniform randomness of the sampling scheme is violated, recovery performance can become support-dependent (see Chapter 5). The signal-dependent sampling approach considered in this work clearly violates the randomness assumptions of compressive sensing. Thus, it is expected that the recovery method (6.1) will have varying performance for signals with different supports of the same size, even if the size of the set of small-amplitude samples used for the recovery is the same. The next section presents approaches for enhancing the recoverability of the sparse signals from the amplitude limited sample sets for the cases when the standard recovery (6.1) leads to erroneous results.

6.2 Approaches

We will demonstrate in Section 6.3 that an ℓ_1-norm minimization (6.1) fitting only the values of the samples with amplitude less than τ will encounter ambiguities for some signals (i.e., signal-dependent performance). In this section, we consider three approaches for improving performance of the recovery. Methods described in Subsections 6.2.1 and 6.2.2 have been considered in independent work of [94], where signal clipping was driving small sample selection.

6.2.1 ℓ_1-norm minimization with inequality constraints

The recoverability of a frequency-sparse signal s from a small-amplitude sample set can be enhanced by taking into account additional information about s that becomes available while discarding large-amplitude samples. In particular, the indices

of samples whose amplitudes exceed the predefined threshold τ are known, and this information can be exploited via inequality constraints in the linear program (6.1). The resulting optimization problem becomes

$$\hat{S} = \arg\min_{\bar{S}} ||\bar{S}||_1 \quad \text{s.t.} \quad A\bar{S} = AS, \quad |s(\underline{\Gamma})| > \tau \tag{6.3}$$

where $\underline{\Gamma}$ is a vector of time stamps of samples that have been discarded. The incorporation of these inequality constraints into standard CS recovery was suggested in [60], where unbounded measurement quantization errors caused by the saturation of ADCs were considered. Because of the relatively easy implementation of threshold comparators at the receiver, the extension of the constraints from (6.3) to multiple thresholds $\tau_n > \tau_{n-1} > \cdots > \tau$ is worth considering for our application of interest. The advantage of adding a second threshold and additional constraints to (6.3) of the form

$$A\bar{S} = AS, \quad \tau_2 \geq |s(\underline{\Gamma})| > \tau, \quad |s(\underline{\Gamma}_2)| > \tau_2, \tag{6.4}$$

will be studied in Section 6.3.

6.2.2 Iterative ℓ_1-norm minimization

A second approach for performance enhancement of (6.1) when only small-amplitude samples are available is iterative ℓ_1-norm minimization. The minimization problem from (6.1) is a relaxation of a computationally intractable combinatorial problem of ℓ_0-"norm" minimization

$$\hat{S} = \arg\min_{\bar{S}} ||\bar{S}||_0 \quad \text{s.t.} \quad A\bar{S} = AS. \tag{6.5}$$

The problem (6.1) can be solved efficiently and a body of existing work has shown there exist conditions under which the combinatorial problem (6.5) and its relaxation

(6.1) are equivalent [96]. However, with the signal-dependent sampling scheme considered in this work, the conventional assumptions of CS are violated, which often leads to non-equivalence of (6.1) and (6.5). In [97], the authors introduced an iterative recovery algorithm consisting of a sequence of weighted ℓ_1-norm minimizations that promotes the sparsity of the result of the computationally tractable ℓ_1-norm minimization for the cases when (6.1) and (6.5) are not equivalent:

$$\hat{S}_i = \arg\min_{\bar{S}} \|C_i \bar{S}\|_1 \quad \text{s.t.} \quad A\bar{S} = AS. \tag{6.6}$$

The diagonal matrix C_i in (6.6) contains positive weights that are updated in every iteration i to be inversely proportional to the values of the solution of the previous iteration:

$$C_{i+1}(k,k) = \frac{1}{|\hat{S}_i(k)| + \epsilon}, \tag{6.7}$$

with ϵ being a positive constant used for stability; one can set C_1 to be the identity. The algorithm is robust with respect to the choice of ϵ, which, as found empirically [97], should be set to a value smaller than the expected amplitudes of coefficients of the solution. The weights (6.7) promote sparsity of the solution, as the coefficients with small amplitude values contribute strongly to the weighted ℓ_1-norm $\|C_i \bar{S}\|_1$ in consecutive iterations. Thus, the final solution tends to consist of a small number of coefficients of highest significance.

As shown in Section 6.3, the iterative recovery algorithm (6.6) can lead to successful recovery of signals from small-amplitude samples when (6.1) and (6.5) are not equivalent due to a violation of the assumptions in CS, which leads to an incorrect solution during the first iteration of (6.6).

6.2.3 Injection of artificial interferers

As a third approach to enhance the performance of (6.1) when only small-amplitude sample sets are available, we consider injection of a known interferer to the signal s.

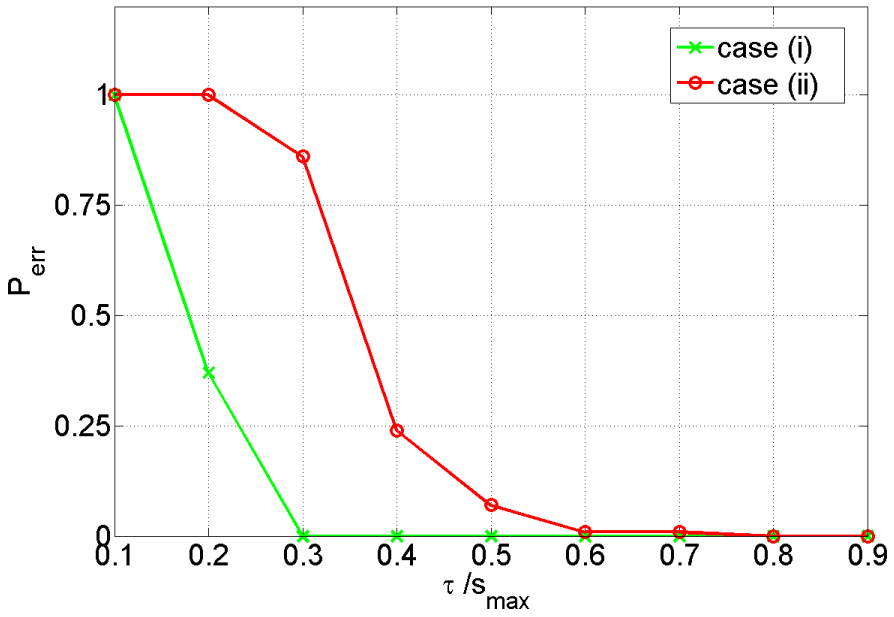

Figure 6.1. Probability of recovery error for the cases (i) and (ii) for ℓ_1-norm minimization (6.1) as a function of the threshold τ.

This corresponds to the addition of a known interferer to the received signal before the LNA in a wide-band receiver. After injection of the interferer, the samples of the signal $s' = s + i_{add}$ for which the amplitude exceeds τ are discarded. Since the interferer is known, the values of i_{add} for the samples retained are subtracted from the respective samples of s' and the resulting signal is used for recovery. If the injected interferer is uncorrelated with the signal s and the power of s and i_{add} are similar, then the sampling times get decorrelated from the frequency content of the signal s. The level of the randomization is higher as the injected interferer becomes more unstructured. Since in practice the threshold τ is a fixed value specified by the nonlinearity of the LNA, the injection of the interferer implies a reduction of the number of samples retained, due to the increased power in s' with respect to s. However, as will be shown in Section 6.3, the injection of a known interferer can significantly enhance

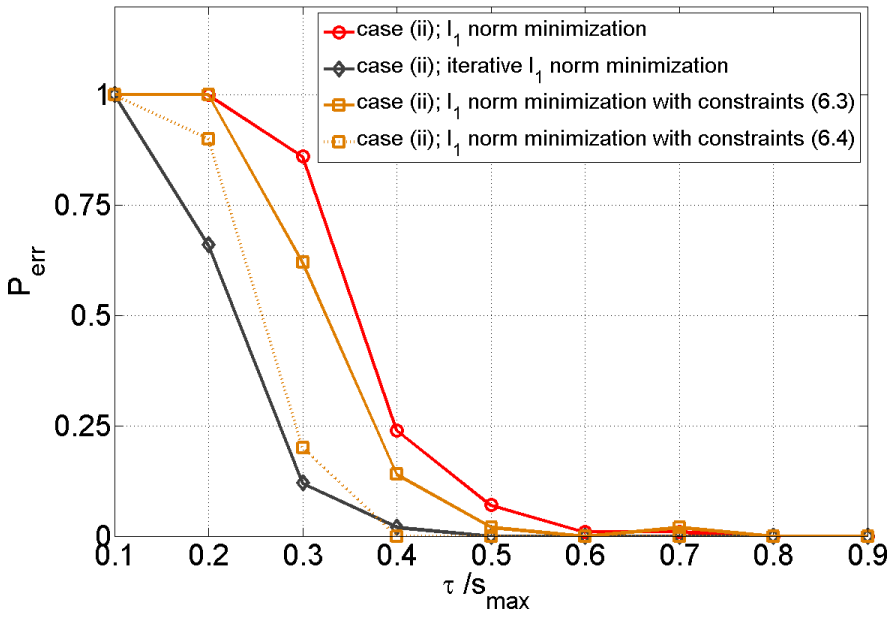

Figure 6.2. Probability of recovery error for the case (ii) for ℓ_1-norm minimization (6.1), for iterative ℓ_1-norm minimization (6.6) and for constrained ℓ_1-norm minimization (6.3) and (6.4) as a function of the threshold τ.

recoverability, despite the penalty (decrease) on the number of samples caused by the increase of the power of the sampled signal s'.

6.3 Simulations

In this section, we present simulation results for CS recovery from sets of small-amplitude samples of frequency-sparse signals. Consider a discrete signal s of length $N = 751$ that consists of 10 tones. We consider two cases: (i) the tones are randomly located on the frequency axis; and (ii) the tones are positioned adjacently to build a single frequency band, located randomly on the frequency axis. For both considered cases (i) and (ii), the amplitudes and phases of the tones are chosen uniformly at random from respective ranges: $[0, 1]$ and $[0, 2\pi]$. We let the threshold τ vary over the range $[0, s_{max}]$, where s_{max} is the maximal amplitude of the signal s. We then discard

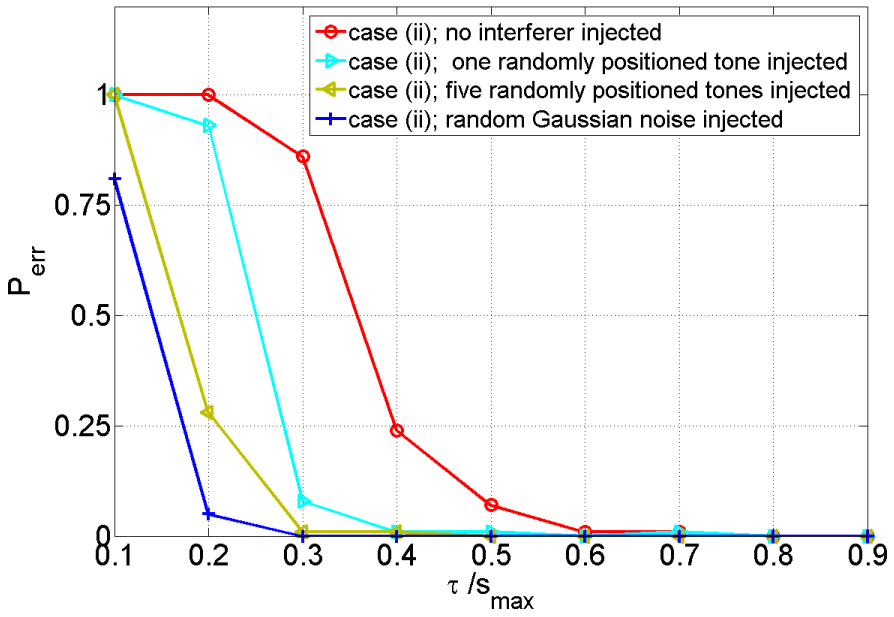

Figure 6.3. Probability of recovery error for the case (ii) for ℓ_1-norm minimization (6.1) as a function of the threshold τ, for different types of known injected interferers.

all samples whose amplitudes are above the threshold τ and preserve the remaining samples as measurements. These measurements are used to solve the minimization problem (6.1) and to find the estimate of the message signal $\hat{s}(t)$ as described in Section 6.1. Figure 6.1 shows the probability of recovery error P_{err} of (6.1), defined as the probability that NRE from (6.2) is above $\rho = 3\%$, calculated over 100 trials for both considered cases (i) and (ii) as a function of the threshold τ. The figure shows that signal recovery is possible from fewer low-amplitude samples of s for the case (i) as compared to the case (ii).

For the case (ii), ℓ_1-norm minimization with inequality constraints and iterative ℓ_1-norm minimization (cf. Sections 6.2.1 and 6.2.2) were applied to improve recovery performance from small-amplitude samples. Figure 6.2 shows the probability of error P_{err} calculated over 50 trials as a function of the threshold τ for the case (ii) when (6.1), (6.3) and (6.6) were used. Five iterations were used for method (6.6); increasing

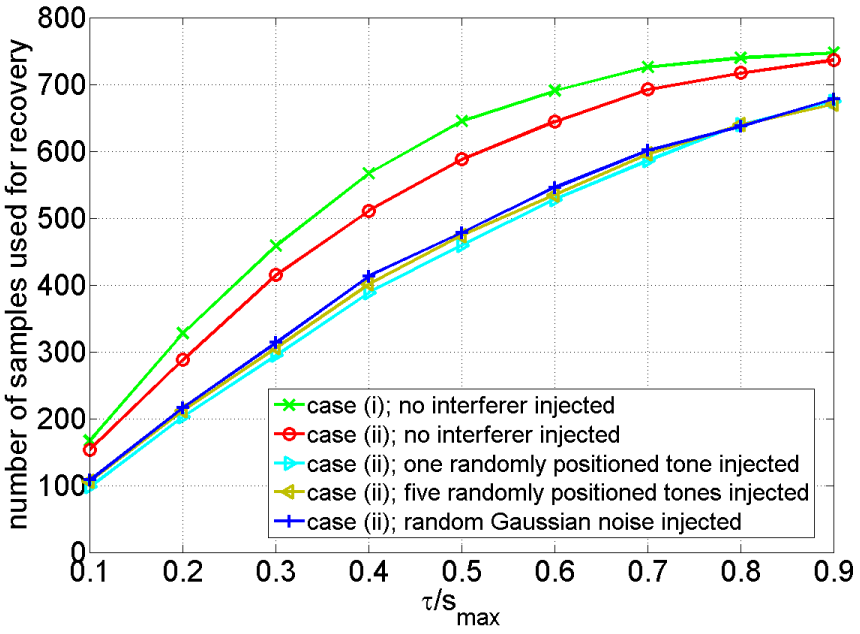

Figure 6.4. Number of small-amplitude samples used for recovery as a function of the threshold τ for the case (i) and for the case (ii) for different types of known interferers injected.

the number of iterations above five did not lead to meaningful performance improvements. Figure 6.2 also shows P_{err} for the case (ii) when (6.3) was used with the additional threshold constraint (6.4). The second threshold τ_2 was used only when $\tau < 0.7 \cdot s_{max}$ and was set to $\tau_2 = 0.75 \cdot s_{max}$.

Finally, we study the recovery performance improvement achieved via injection of known interferers. Figure 6.3 shows P_{err} of (6.1), calculated over 100 trials as a function of the threshold τ for the case (ii), when three different types of known interferers were injected: 1 and 5 randomly positioned tones and a random Gaussian noise. The average power of the injected interferer was set to be equal to the power of the signal s. For the considered frequency-sparse signal s, even a highly structured injected interferer (i.e., the sum of 5 tones) leads to significant decorrelation of the sampling times from the signal structure and thus significant recovery performance

enhancement. Figure 6.3 shows that, even for a fixed threshold, which in practice is dictated by the characteristics of the nonlinearity of the LNA, adding an interferer i_{add} enhances recovery performance despite a reduction of the number of samples due to the average power increase of $s' = s + i_{add}$ with respect to s. Figure 6.4 shows the mapping between the threshold τ and the number of small-amplitude samples used for recovery, calculated over 100 trials, for different types of known injected interferers. It shows how a fraction of samples is lost due to the injected interferer, and how the choice of the interferer is causing only a small difference in the number of samples preserved.

6.4 Conclusions

In this chapter we studied compressive sensing recovery of frequency-sparse signals from irregular, small-amplitude samples. We have shown that the standard ℓ_1-norm minimization recovery performance becomes signal-dependent due to the correlation between the signal structure and the location of small-amplitude samples, thus motivating the exploration of enhanced CS recovery schemes. Three such schemes that have been presented in this chapter show significant improvement over the standard ℓ_1-norm minimization recovery from small-amplitude signal samples.

CHAPTER 7

MITIGATION OF SPECTRAL LEAKAGE FOR SINGLE CARRIER, BLOCK-PROCESSING COGNITIVE RADIO RECEIVERS

7.1 Problem Statement

As discussed in Chapter 1, in addition to the receiver's nonlinearities, time-truncation of the processed signal records causes degradation of the dynamic range of wide-band receivers. While the dynamic range enhancement via selected sampling that leads to linear operation of the receiver's front-end was studied in Chapters 5 and 6, in this chapter we consider the latter cause of degradation and study a method for adaptation of the processing window size for dynamic range improvement for the receivers of single-carrier, block transmission.

Block transmissions, for which groups of data symbols are processed as a unit, allow for the implementation of frequency domain channel equalization (FDE), which for broadband transmissions in rich multi-path environments can bring significant complexity relaxations when compared to time-domain equalization [98]. To allow for FDE, block transmissions employ a cyclic prefix (CP). The CP is a copy of the end of the signal block attached in front of the block. To avoid inter-block interference (IBI), the length of the CP is chosen larger than the maximum delay spread of the channel. Because of the dynamic character of wireless channels, value of the delay spread changes over time and space [99]. For the high reliability demanded of modern communication systems, the tail of the distribution of the delay spread dictates a conservative choice of the length of the cyclic prefix, which is fixed during the design stage.

Since truncation in the time domain causes spreading in the frequency domain, weak signals in a block transmission can be corrupted by strong in-band interferers (see Section 2.1). In this work we propose enhancement of the dynamic range of a block processing CR receiver by an adaptive choice of the processing block size to minimize the spectral leakage into the frequency sub-bands occupied by the signal of interest. In particular we will show that with a non-complete removal of the CP at the receiver, which can be employed in wireless environments with channel delay spreads even slightly shorter than the length of the CP, significant receiver dynamic range improvements can be achieved for block transmissions.

The discrete Fourier transform (DFT) is an invertible signal processing operation that projects a time-limited signal onto a set of complex frequencies, and gives a discrete representation of the signal's spectrum. The values of the frequencies that the signal is projected onto build a discrete grid, equidistantly dividing the entire sampling bandwidth. A complex sinusoid oscillating with a frequency off the discrete frequency grid cannot be represented with a single element of the grid, and its energy *leaks* between multiple elements, which leads to misinterpretation of the spectral content. This is visualized in Figure 7.1, where the observation time of a complex oscillation is shown for two different cases: (a) the frequency of the sinusoid overlaps with that of one of the grid points (left subplots); (b) the frequency of the sinusoid lies off the grid which leads to spectral leakage (right subplots). For wide-band, block-processing receivers, the finite block length can cause misinterpretation of the Fourier coefficients of the in-band interference, possibly orders of magnitude stronger than the message of interest, which can lead to significant contamination of the message.

A cyclic prefix is employed in the single-carrier frequency division multiplexing (SC-FDM) scheme [100], which is often considered for broadband transmissions over wireless channels. For example SC-FDM has been selected as an uplink communication scheme for the Long Term Evolution (LTE) standard for wireless, high-speed

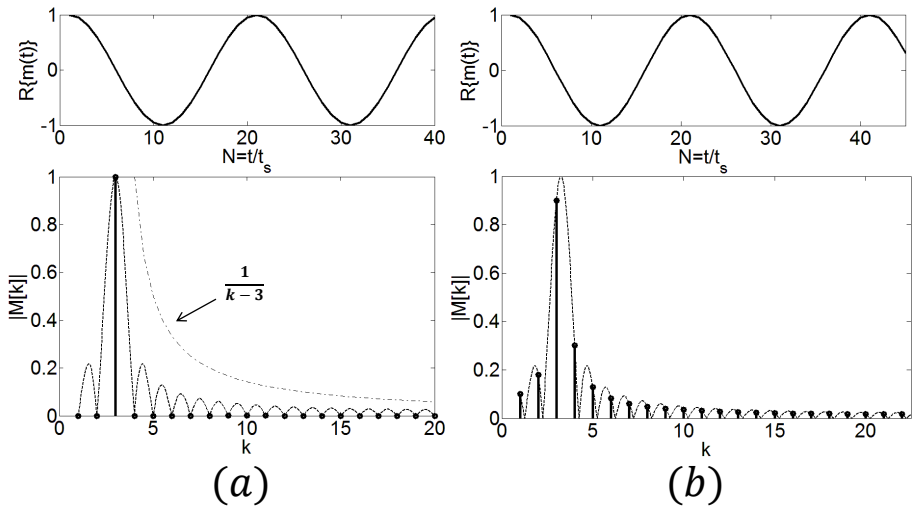

Figure 7.1. Spectral representation of a complex sinusoid with the DFT. Observation time is such that: (a) the frequency of the oscillation overlaps with one of the DFT grid points; (b) the frequency lies off the grid which leads to the spectral leakage.

data communication for mobile phones and data terminals [101]. Similar to orthogonal frequency division multiplexing (OFDM), SC-FDM processes data in blocks; however, unlike OFDM, it utilizes a single carrier modulation at the transmitter, which allows for reduction of the peak-to-average power ratio (PAPR) [100] and thus improves the efficiency of RF power amplifiers. The inherent DFT operation at the receiver (Figure 7.2) makes SC-FDM suitable for spectral sensing applications of cognitive radio. If the length of the CP is longer than the longest delay path of the channel, then the received blocks arriving at the receiver over different channel paths appear as circularity shifted versions of the transmitted block. A circular shift of a discrete-time record corresponds to a multiplication of its DFT with a linear phase. Therefore FDE can be implemented as a simple, frequency domain multiplication with an inverse of the estimate of the channel transfer function, in contrast to time domain equalization, which can involve time-domain adaptive filters with tens of taps and hundreds of multiplication operations required per data symbol.

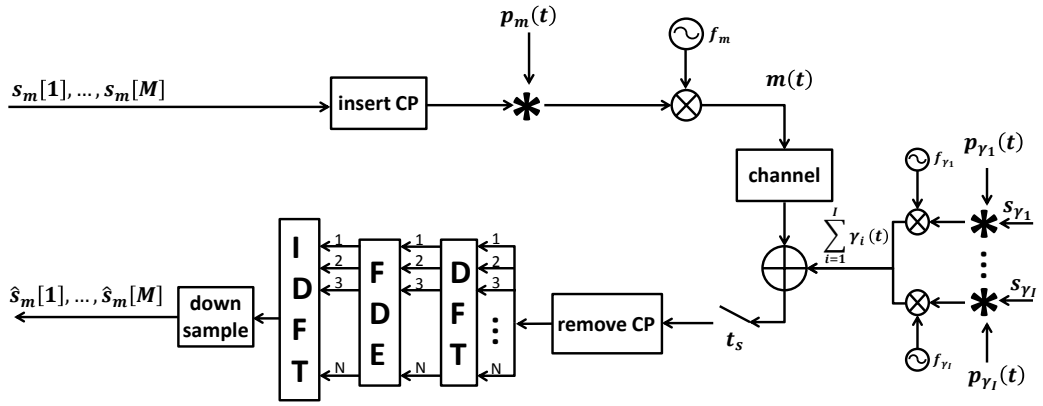

Figure 7.2. Block diagram for a single carrier frequency division multiplexing cognitive radio transmission subject to interference.

Consider a single-carrier, block transmission and a wide-band, highly over-sampling, cognitive radio receiver visualized in Figure 7.2. Message data symbols s_m after cyclic prefix insertion are pulse-shaped with pulse $p_m(t)$, up-converted to frequency f_m, and sent over the channel. The received signal, including an additive interference $\sum_{i=1}^{I} \gamma_i(t)$, possibly with power orders of magnitude higher than that of the message, is over-sampled with a sampling rate $f_s = 1/t_s$. A block of high-rate samples is captured, the cyclic prefix is discarded, and the DFT of the time capture of length N is calculated for frequency domain channel equalization. The transmitted data is then recovered from the time representation of the equalized and digitally filtered parts of the spectrum.

Consider an off-grid complex interferer sinusoid with unit amplitude and a frequency f_γ such that: $f_\gamma \cdot t_s \cdot N = l + a; l \in \mathbb{N}; -0.5 < a \leq 0.5$. The DFT of the sinusoid is ((5.103) [102, pg. 262])

$$\mathbb{O}(f_\gamma, k) = \frac{\sin\left(\pi(f_\gamma \cdot t_s \cdot N - k)\right)}{\sin\left(\frac{\pi}{N}(f_\gamma \cdot t_s \cdot N - k)\right)} \cdot e^{j\pi(f_\gamma \cdot t_s \cdot N - k)\frac{N-1}{N}},$$

$$k = 1, \ldots, N. \tag{7.1}$$

For an interferer $\gamma(t)$ with a continuous bandwidth $(f_\gamma - W_\gamma/2, f_\gamma + W_\gamma/2)$, spectral power leakage into a set K_m of the discrete frequencies occupied by the message of interest (either a message to be sensed or to be received) is therefore

$$L_N[K_m] = \sum_{k \in K_m} \left| \int_{f_\gamma - W_\gamma/2}^{f_\gamma + W_\gamma/2} \Gamma_\gamma(f) \mathbb{O}(f,k) df \right|^2, \tag{7.2}$$

where $\Gamma_\gamma(f)$ are the values of the continuous Fourier transform of the interferer. If the power of the interferer is orders of magnitude higher than the power of the message of interest, the leakage can cause significant degradation of the message's quality.

7.2 Proposed Method for Receiver Dynamic Range Enhancement for Block Transmissions

For conventional block receivers the cyclic prefix is removed completely before the block is processed to retrieve the transmitted message of interest. The complete CP removal allows for the cancellation of IBI caused by the channel delay spread τ_{CH} not exceeding the length L_{CP} of the cyclic prefix and leads to a fixed processing block size N_B. As we saw in the Section 2.1 (Figure 7.1 and eq. (7.1)) the character of the DFT leakage depends on the size N of the processing block. If the delay spread of a considered channel is smaller than the length of the CP, then without compromising the ability to cancel the IBI completely, a potential leakage reduction and therefore dynamic range enhancement can be achieved with an *adaptive*, partial removal of the cyclic prefix. In particular an adaptive choice of the length N of the processing block can be made from a search set $\mathbf{N_{search}} = \left\{ N_B, N_B + 1, \ldots, N_B + \frac{L_{CP} - \tau_{CH}}{t_s} \right\}$, where t_s is the sampling period. The current value of τ_{CH} that determines the size of N_{search} can be estimated at the receiver using conventional channel order estimation methods as studied for example in [103].

To gain an intuition on how the spectral leakage can be controlled with the size N of the observation window, consider an interferer γ consisting of a set of I complex

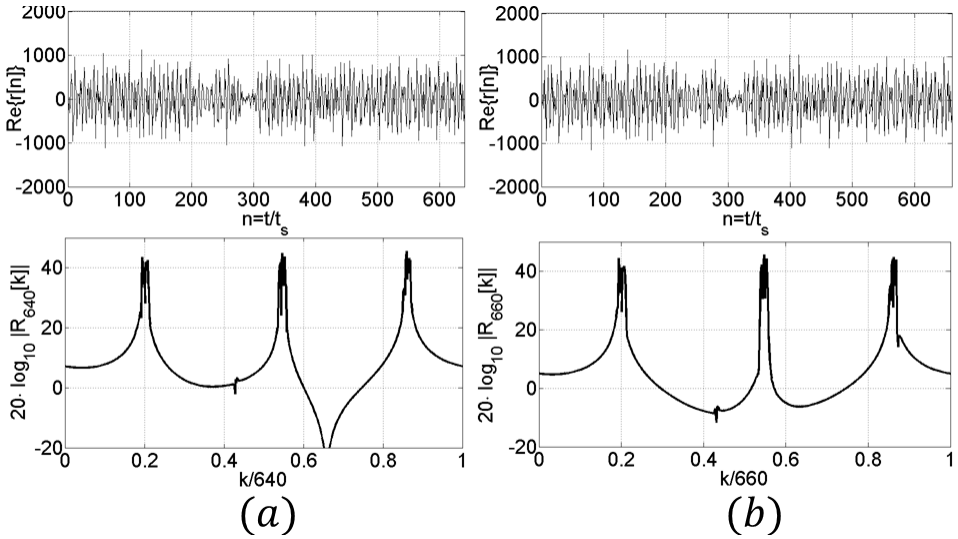

Figure 7.3. Time domain (top) and DFT (bottom) capture ($r[n]$ and $R_N[k]$) of a QPSK block transmission of 5 symbols of interest equipped with 1 cyclic prefix symbol, transmitted at a symbol rate $W_m = 3.84$MHz and power P_m for two different choices of the processing block size: (a) N = 640 after a complete CP removal (b) N = 660 after a partial CP removal. The signal of interest was contaminated with three interferers occupying 7.68MHz sub-bands, with total power P_γ. The signal-to-interference ratio and sampling frequency were set to, respectively: $SIR = 10\log_{10}\frac{P_m}{P_\gamma} = -60dB$ and $f_s = 491.52$MHz. The message signal carrier frequency was located at $f_m/f_s \approx 0.4297$.

oscillations with amplitudes Γ_{γ_i} and frequencies f_{γ_i}, $i = 1, \ldots, I$, and a message m consisting of a set of M complex oscillations with frequencies f_{m_μ}, $\mu = 1, \ldots, M$. Depending on the size N of the observation window, f_{m_μ}, $\mu = 1, \ldots, M$, can either lie on or off the discrete frequency grid. Denote the rounded value of x to the nearest integer as $[\![x]\!]$. The discrete frequency set K_m of the message is thus $K_m = \left\{ \left[\!\left[\frac{f_{m_1}}{f_s} N \right]\!\right], \ldots, \left[\!\left[\frac{f_{m_M}}{f_s} N \right]\!\right] \right\}$. Ignoring the rounding operation when building the set K_m, defining $\Delta_{i,\mu} = \frac{(f_{\gamma_i} - f_{m_\mu})}{f_s}$, with (7.1) and (7.2), the total interference experienced by the message due to the leakage caused by the limited observation window size N is

$$
\begin{aligned}
L_N[K_m] &= \sum_{\mu=1}^{M} \left| \sum_{i=1}^{I} \Gamma_{\gamma_i} \frac{\sin\left(\pi \Delta_{i,\mu} \cdot N\right)}{\sin\left(\pi \Delta_{i,\mu}\right)} e^{j\pi\Delta_{i,\mu}\cdot N} \right|^2 \\
&= \sum_{\mu=1}^{M} \left| \sum_{i=1}^{I} \Gamma_{\gamma_i} \frac{1}{2j \cdot \sin(\pi \cdot \Delta_{i,\mu})} \cdot \left(e^{2j\pi\Delta_{i,\mu}\cdot N} - 1 \right) \right|^2 .
\end{aligned}
\tag{7.3}
$$

Denote $\tilde{\Gamma}_{i,\mu} = \Gamma_{\gamma_i} \cdot \frac{1}{2j \cdot \sin(\pi \cdot \Delta_{i,\mu})}$. The minimization of the interference leakage experienced by the message over the choice of the window size N simplifies to

$$
L_{min}[K_m] = \min_{N \in \mathbf{N_{search}}} \sum_{\mu=1}^{M} \left| \sum_{i=1}^{I} \tilde{\Gamma}_{i,\mu} \cdot \left(e^{2j\pi\Delta_{i,\mu}\cdot N} - 1 \right) \right|^2 , \tag{7.4}
$$

which is a function of $\tilde{\Gamma}_{i,\mu}$'s that depend on the signal transmitted by the interferers, weighted with factors that depend only on the frequency spacing between the message and the interferers and that can be controlled with the choice of the value of the length N of the processing block.

Figure 7.3 shows a visualization of the possible reduction of the leakage of interferers' power into the frequency band occupied by the message of interest that can be achieved with the partial removal of the CP and hence an adjustment of the processing window size from 640 to 660. While increasing the spectral leakage to some parts of the spectrum, the adjustment leads to a roughly 10 dB reduction of the leakage into frequency bins around message carrier frequency.

In practice the receiver does not have access to $L_N[K_m]$ (7.2), (7.3); thus, some other measurable quantity must be used to determine the optimal N. We will employ the total received power in the frequency band occupied by the message of interest. To establish its utility we show that the power of the message of interest and the interference decorrelate quickly with the size N of the processing block. Consider a received signal $r(t)$ consisting of the message of interest $m(t)$ with power P_m and an interferer $\gamma(t)$ with power P_γ. The baseband sampled received signal can be written as $r[n] = m[n] + \gamma[n]$. Its power is

$$P_r = \frac{1}{N}\sum_{n=1}^{N} r[n]^2 = P_m + P_\gamma + 2C; \quad C = \frac{1}{N}\sum_{n=1}^{N} m[n]\cdot\gamma[n]. \tag{7.5}$$

Let $m[n]$ and $\gamma[n]$ be respective digital streams of symbols s_m and s_γ, with periods T_m and T_γ, pulse-shaped with analog pulses $p_m(t)$ and $p_\gamma(t)$, sampled with frequency $f_s = 1/T_s$:

$$m[n] = \sum_{l=-\infty}^{\infty} s_m[l]\cdot p_m(nT_s - lT_m), \quad \gamma[n] = \sum_{l=-\infty}^{\infty} s_\gamma[l]\cdot p_\gamma(nT_s - lT_\gamma). \tag{7.6}$$

Assume $s_m[l]$ and $s_\gamma[l]$ to be uncorrelated, zero-mean random variables, which is the case for commonly used digital modulation schemes. Then $\mathbf{E}\{C\} = 0$ and

$$Var\{C\} = \frac{Var\{s_m\}\cdot Var\{s_\gamma\}}{N^2}\sum_{n=1}^{N}\sum_{l_1=-\infty}^{\infty}\sum_{l_2=-\infty}^{\infty} p_m^2(nt_s - l_1T_m)\cdot p_\gamma^2(nt_s - l_2T_\gamma). \tag{7.7}$$

For analog pulses $p_m(t)$ and $p_\gamma(t)$ and values of N considered in Section 7.3, $Var\{C\}$ is close to zero, and therefore the power of the block of the received signal $r[n]$ concentrated in the frequency range occupied by the message of interest, which is easily measured, can be used to make the decisions on the size of the processing block. Denote the DFT of length N of the received signal $r[n], n = 1, \ldots, N$, carrying a message of interest, occupying a continuous bandwidth $(f_m - W_m/2, f_m + W_m/2)$ as $\mathbf{R}_N$. The optimal size of the processing block can then be chosen as

$$N_{opt} = \underset{N\in\mathbf{N_{search}}}{\arg\min} \sum_{k_m=[\![\frac{f_m-W_m/2}{f_s}N]\!]}^{[\![\frac{f_m+W_m/2}{f_s}N]\!]} |\mathbf{R}_N[k_m]|^2. \tag{7.8}$$

7.3 Numerical Results

In this section we numerically study possible dynamic range improvements for the block CR receivers, employing adaptive, partial removal of the CP, and hence

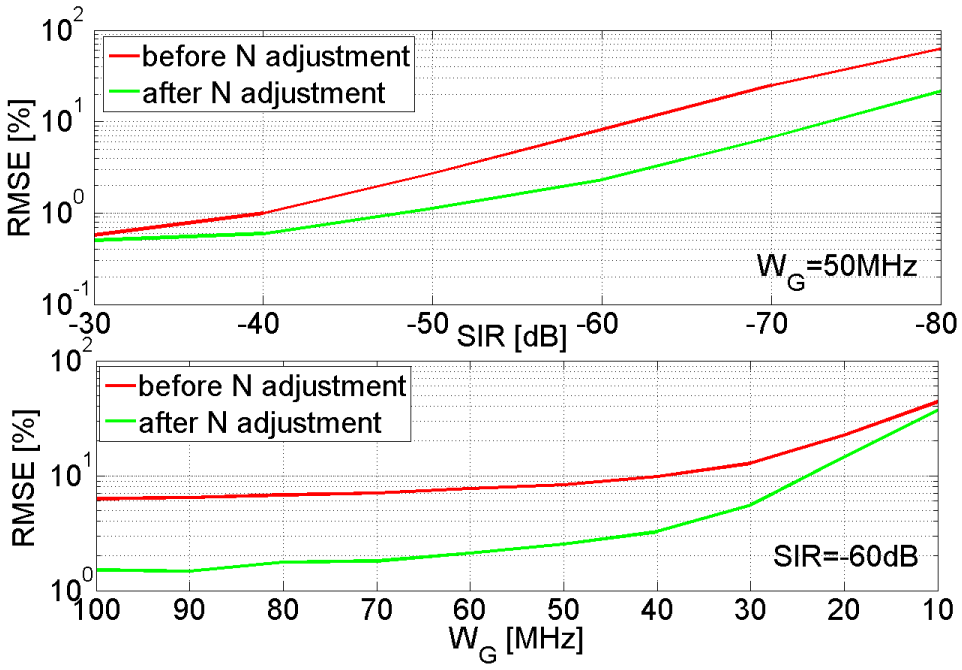

Figure 7.4. RMSE of the recovered QPSK message symbols before and after the adjustment of the processing block size N with (7.8) as a function of the SIR for a fixed guard bandwidth from (7.9) W_G=50MHz (top subplot) and as a function of the guard bandwidth W_G for a fixed SIR=-60 dB (bottom subplot).

with the adaptive, block-to-block choice of the processing block size. Consider a CR receiver working with a sampling rate $f_s = 491.52$MHz receiving a transmission of interest together with strong interference. The transmission of interest $m(t)$, at a carrier frequency f_m, is a QPSK transmission with a symbol rate of $W_m = 3.84$MHz, pulse-shaped with a raised-cosine filter $p_m(t)$ with a roll-off factor $\beta = 0.5$. The interferer signal consists of three QPSK transmissions at carrier frequencies f_{γ_i}, with symbol rates W_{γ_i}, also pulse-shaped with $\beta = 0.5$ raised-cosine filters $p_{\gamma_i}(t)$, $i = 1, 2, 3$; however, the method's performance did not change for various settings of the interferers' excess bandwidth β. The processing block was built out of 32 QPSK message symbols and was equipped with a cyclic prefix of length $6.25\% \cdot 32 = 2$ symbols, which allows the avoidance of IBI caused by a channel with a maximal delay path difference $\Delta d_{CH_1} = c \cdot \tau_{CH_1} = c \cdot \frac{2}{3.84MHz} = 156.14$m, where c is the speed

of light. The number of samples building the processing block at the receiver after the complete CP removal was then $N_B = [\![\frac{32}{3.84MHz} \cdot f_s]\!] = 4096$.

We assume that the transmission occurs in a wireless environment with the maximal channel delay path difference Δd_{CH_2} slightly shorter than Δd_{CH_1}. In such an environment, a complete IBI cancelation is achieved even if a fraction of the CP (up to $\tau_{CH_1} - \tau_{CH_2}$ seconds) is not discarded and kept for further processing. This allows for a processing block size search set: $\mathbf{N_{search}} = \{N_B, \ldots, N_{max}\}$, with $N_{max} = 4096 + (\tau_{CH_1} - \tau_{CH_2}) \cdot f_s$.

For the considered transmission, for $N_{max} = 4134$ corresponding to a search space of size 15% of the CP length, the upper subplot of Figure 7.4 shows the root mean square error (RMSE) of the recovered QPSK message symbols as a function of the SIR before and after the adjustment of N with (7.8). Figure 7.4, averaged over 500 random choices of message and interferers' symbols s_m and s_{γ_i}, carrier frequencies f_m and f_{γ_i} and interferers' symbol rates $W_{\gamma_i} \in \{1.92, 3.84, 7.68, 15.36, 30.72, 61.44\}$ MHz, shows that for a desired value of RMSE a significant reduction of required SIR value, and hence dynamic range improvement of over 10dB, can be achieved with this simple adjustment of the block processing length N. The random choice of f_m, f_{γ_i} and W_{γ_i} was subject to a constraint

$$|f_m - f_{\gamma_i}| < W_G + W_{\gamma_i}/2, \quad i = 1, 2, 3, \tag{7.9}$$

where W_G is a guard bandwidth set to 50MHz. The lower subplot of Figure 7.4 for $N_{max} = 4134$ and $SIR = -60dB$ shows the RMSE of the recovered QPSK message symbols as a function of the guard bandwidth W_G. The effect of the spectral leakage decreases with increasing SIR, and therefore the room for impact for the proposed leakage reduction technique (or any other) is limited at high SIRs. For fixed SIR the performance of the proposed method decreases with decreasing W_G.

Small W_G can lead to small values of $\Delta_{i,\mu}$ from (7.4), which limits the rotation speed of $\left(e^{2j\pi\Delta_{i,\mu}\cdot N}-1\right)$.

7.4 Conclusions

In this chapter an approach for the mitigation of spectral leakage via an adaptive choice of processing window size was proposed for block processing, single-carrier CR receivers. The method is based on an adaptive partial removal of the CP and can be applied in environments with maximal channel delay paths shorter than the length of the CP. The proposed method does not require any structural changes to the receiver and allows for significant dynamic range improvements (over 10 dB) when compared to a studied fixed window size approach from [104].

CHAPTER 8

CONCLUSIONS

Wireless security and interference rejection in wide-band wireless communications, which are often considered as two main challenges in wireless networks' design and research, were studied in this dissertation. We conducted research that illuminates the effects of non-ideal components in wireless communication systems and presented novel signal processing methods that address the two aforementioned challenges. In particular, we proposed novel, model-based approach that exploits transmitters' inherent nonlinearities for the purpose of device identification. We also studied novel signal recovery methods that allow for a higher tolerance of inherent nonlinearities of wireless wide-band receivers.

The proposed model-based identification methods were shown effective via simulations and measurements of components of commercial wireless devices. Nonlinearities of power amplifiers (PAs) and RF oscillators are most suitable for the purpose of device identification. While the character of the PA nonlinearity is power level dependent and can change over time in systems with power control employed, operation of the RF oscillators is power mode independent, which should dictate the choice of the identification methods for different system deployments.

The model-based approach proposed in this dissertation is advantageous over the empirical methods reported in the literature, as it is more resistant against potential countermeasure attacks that can be implemented by sophisticated, masquerading users. Moreover, in contrast to the empirical methods, the model-based approach allows for analysis of the identification performance as a function of important com-

munications parameters such as SNR. Low probabilities of identification errors were reported in this dissertation at practical SNR levels for short capture lengths. It is important to stress here that an increase of available processing power, and therefore an increase of the length of the captures that can be processed for identification, can lead to a proportional improvement in the performance versus SNR.

In addition to its application for post-crime device identification, we see the device fingerprinting work presented here as informative to the recently emerged world-wide debate about the level of personal privacy, as it gives insight into the degree to which such may be compromised due to device hardware imperfections.

It is a common concern that hardware based fingerprints can suffer time-instability because of temperature variations of the device's components. This concern applies to all hardware based fingerprinting approaches, which we summarized in Section 1.1.1. During measurements that we performed to obtain results reported in this dissertation, we observed a very quick temperature stabilization of the measured components, and therefore we initially ignored the temperature variation in our investigations. These variations, mostly due to the ambient temperature variations, should however be considered in the future for refinement of the proposed identification methods.

While the identification methods presented in the first part of this dissertation exploited nonlinearities of transmitter's components, in the second part of the dissertation we presented signal recovery methods for mitigation of nonlinear effects at the receiver, for more efficient spectrum utilization. The contributions that we reported in this area are threefold. First we presented an analytically derived multiplicative penalty on the number of measurements needed for successful recovery of sparse signals, when the measurements, instead of being taken at random, exhibit a grouped structure. Second we introduced three methods that improve recoverability of sparse signals from small-amplitude samples. Third we proposed an adaptive choice of pro-

cessing block size for block-processing receivers that leads to significant reduction of the effect of spectral leakage.

APPENDIX A

PROOF OF THEOREM 3

Denote

$$Y := \frac{N}{M} A_{\Omega T}^H A_{\Omega T} - I = \frac{N}{M} \sum_{i=1}^{N/g} \delta_i A_{G_i T}^H A_{G_i T} - I, \tag{A.1}$$

where δ_i is a Bernoulli random variable with $P(\delta_i = 1) = \frac{M}{N}$. Because $A^H A = I$, we have

$$\sum_{i=1}^{N/g} A_{G_i T}^H A_{G_i T} = I, \tag{A.2}$$

and so we can write

$$Y = \frac{N}{M} \sum_{i=1}^{N/g} \left(\delta_i - \frac{M}{N} \right) A_{G_i T}^H A_{G_i T} =: \sum_{i=1}^{N/g} Y_i. \tag{A.3}$$

We will now use [105, Theorem 1.4], which we include below for completeness.

Theorem 5. *Consider a finite sequence $\{Y_i\}$ of independent self-adjoint random matrices with dimension d. Assume that each matrix Y_i satisfies $\mathbb{E}\{Y_i\} = 0$ and $\|Y_i\| \leq B$ almost surely. Then, for all $t \geq 0$,*

$$P\left\{ \left\| \sum_i Y_i \right\| > t \right\} \leq d \cdot \exp\left(\frac{-t^2/2}{\sigma^2 + Bt/3} \right),$$

$$\textit{where} \quad \sigma^2 = \left\| \sum_i \mathbb{E} Y_i^2 \right\|.$$

For our case,

$$Y_i = \left(\delta_i - \frac{M}{N} \right) A_{G_i T}^H A_{G_i T} \frac{N}{M} \tag{A.4}$$

and $E(Y_i) = 0$. We find a bound B on $\|Y_i\|$:

$$\|Y_i\| = \sup_{f_1,f_2} |\langle f_1, Y_i f_2\rangle| \leq \sup_{f_1,f_2} \left| \left\langle f_1, \frac{N}{M} \sum_{l\in G_i} a^l \otimes a^l f_2 \right\rangle \right| \leq \frac{N}{M} \sup_{f_1,f_2} \sum_{l\in G_i} |\langle f_1, a^l \otimes a^l f_2\rangle|$$
$$= \frac{N}{M} \sup_{f_1,f_2} \sum_{l\in G_i} |\langle f_1, a^l\rangle\langle a^l, f_2\rangle| \leq \frac{N}{M} \left\|a^l\right\|^2 \sup_{f_2} \sum_{l\in G_i} \frac{|\langle a^l, f_2\rangle|}{||a^l||} \leq \frac{N}{M}\mu^2(A)|T|\gamma =: B, \tag{A.5}$$

where the supremum is over unit-norm vectors f_1 and f_2, a^l is the l^{th} row of the matrix A_T, and γ is defined in (5.4).

Next, we calculate σ^2 from Theorem 5 as

$$\sigma^2 = \left\| \sum_{i=1}^{N/g} \mathbb{E}(Y_i^2) \right\| = \text{var}\{\delta_i\} \frac{N^2}{M^2} \left\| \sum_{i=1}^{N/g} (A_{G_iT}^H A_{G_iT})^2 \right\|. \tag{A.6}$$

Since $A_{G_iT}^H A_{G_iT}$ is a Hermitian matrix, its eigendecomposition is $A_{G_iT}^H A_{G_iT} = \Omega_i \Lambda_i \Omega_i^H$, where Ω_i is a matrix whose columns are the orthonormal eigenvectors ω_{ij}, $j = 1, \ldots, |T|$, of the matrix $A_{G_iT}^H A_{G_iT}$, and Λ_i is a diagonal matrix containing the eigenvalues $\{\lambda_{ij}\}_{j=1}^{|T|}$ of the matrix $A_{G_iT} A_{G_iT}^H$. Thus, we can write

$$A_{G_iT}^H A_{G_iT} A_{G_iT}^H A_{G_iT} = \Omega_i \Lambda_i \Omega_i^H \Omega_i \Lambda_i \Omega_i^H = \Omega_i \Lambda_i^2 \Omega_i^H = \sum_{j=1,\ldots,|T|} \lambda_{ij}^2 \omega_{ij} \omega_{ij}^H,$$

and so

$$\left\| \sum_{i=1}^{N/g} A_{G_iT}^H A_{G_iT} A_{G_iT}^H A_{G_iT} \right\| = \left\| \sum_{i=1}^{N/g} \sum_{j=1}^{|T|} \lambda_{ij}^2 \omega_{ij} \omega_{ij}^H \right\|. \tag{A.7}$$

The right side of (A.7) is a weighted double sum of positive semidefinite matrices. The spectral norm of such a sum increases monotonically with the value of each of

the weighting coefficients. Therefore, we can upper-bound (A.7) by replacing λ_{ij} with $\max_{i=1,\ldots,\frac{N}{g}} \max_{j=1,\ldots,|T|} \lambda_{ij}$ and taking it out of the operator norm:

$$\begin{aligned}\left\| \sum_{i=1}^{N/g} A_{G_iT}^H A_{G_iT} A_{G_iT}^H A_{G_iT} \right\| &\leq \max_{i=1,\ldots,\frac{N}{g}} \max_{j=1,\ldots,|T|} \lambda_{ij} \left\| \sum_{i=1}^{N/g} \sum_{j=1}^{|T|} \lambda_{ij} \omega_{ij} \omega_{ij}^H \right\| \\ &= \max_{i=1,\ldots,\frac{N}{g}} \left\| A_{G_iT}^H A_{G_iT} \right\| \cdot \left\| \sum_{i=1}^{N/g} A_{G_iT}^H A_{G_iT} \right\|. \end{aligned} \tag{A.8}$$

With (A.8) we can bound (A.6) by

$$\begin{aligned}\sigma^2 &\leq \mathrm{var}\{\delta_i\} \frac{N^2}{M^2} \max_{i=1,\ldots,\frac{N}{g}} ||A_{G_iT}^H A_{G_iT}|| \cdot \left\| \sum_{i=1}^{N/g} A_{G_iT}^H A_{G_iT} \right\| \\ &= \mathrm{var}\{\delta_i\} \frac{N^2}{M^2} \max_{i=1,\ldots,\frac{N}{g}} ||A_{G_iT}^H A_{G_iT}|| \cdot ||A_T^H A_T|| \\ &= \frac{M}{N}\left(1 - \frac{M}{N}\right) \frac{N^2}{M^2} \max_{i=1,\ldots,\frac{N}{g}} \sup_{\|f_1\|=\|f_2\|=1} \left| \left\langle f_1, \sum_{l \in G_i} a^l \otimes a^l f_2 \right\rangle \right| \\ &\leq \frac{N}{M} \max_{i=1,\ldots,\frac{N}{g}} \sup_{\|f_1\|=\|f_2\|=1} \sum_{l \in G_i} |\langle f_1, a^l \otimes a^l f_2 \rangle| \\ &= \frac{N}{M} \max_{i=1,\ldots,\frac{N}{g}} \sup_{\|f_1\|=\|f_2\|=1} \sum_{l \in G_i} |\langle f_1, a^l \rangle| \cdot |\langle a^l, f_2 \rangle| \\ &\leq \frac{N}{M} \left\| a^l \right\|^2 \max_{i=1,\ldots,\frac{N}{g}} \sup_{\|f_1\|=1} \sum_{l \in G_i} \frac{\langle f_1, a^l \rangle}{||a^l||} \\ &\leq \frac{N}{M} \mu^2(A) |T| \max_{i=1,\ldots,\frac{N}{g}} \sup_{\|f_1\|=1} \sum_{l \in G_i} \frac{\langle f_1, a^l \rangle}{||a^l||} \\ &= \frac{N}{M} \mu^2(A) |T| \gamma = B. \end{aligned} \tag{A.9}$$

We put together (A.5), (A.9) and Theorem 5 to write

$$P(||Y|| \geq 1/2) \leq |T| \cdot \exp\left\{ \frac{-1/8}{7/6 \cdot \frac{N}{M} \mu^2(A) |T| \gamma} \right\}, \tag{A.10}$$

which proves Theorem 3.

APPENDIX B

PROOF OF LEMMA 1

One can express v^0 as

$$v^0 = \sum_{i=1}^{N/g} \delta_i \sum_{l \in G_i} A(l, t_0) a^l. \tag{B.1}$$

Now due to the orthogonality of the columns of the matrix A,

$$\sum_{i=1}^{N/g} \sum_{l \in G_i} A(l, t_0) A(l, t) = 0, \tag{B.2}$$

and we can write

$$v^0 = \sum_{i=1}^{N/g} (\delta_i - E(\delta_i)) \sum_{l \in G_i} A(l, t_0) a^l = \sum_{i=1}^{N/g} Y_i, \tag{B.3}$$

with

$$Y_i := \left(\delta_i - \frac{M}{N} \right) \sum_{l \in G_i} A(l, t_0) a^l. \tag{B.4}$$

We see that $E(Y_i) = 0$ and we can write

$$E||v^0||^2 = E \left\| \sum_{i=1}^{N/g} Y_i \right\|^2 = E \left\langle \sum_{i=1}^{N/g} Y_i, \sum_{i'=1}^{N/g} Y_i \right\rangle = E \sum_{i=1}^{N/g} \langle Y_i, Y_i \rangle + E \sum_{i \neq i'}^{N/g} \langle Y_i, Y_{i'} \rangle = \sum_{i=1}^{N/g} E \langle Y_i, Y_i \rangle . \tag{B.5}$$

Each element of the sum above can be bounded by

$$E \langle Y_i, Y_i \rangle = E \left\langle \left(\delta_i - \frac{M}{N} \right) \sum_{l \in G_i} A(l, t_0) a^l, \left(\delta_i - \frac{M}{N} \right) \sum_{l' \in G_i} A(l', t_0) a^{l'} \right\rangle$$

$$
\begin{aligned}
&= \mathrm{var}\{\delta_i\} \left\langle \sum_{l\in G_i} A(l,t_0)a^l, \sum_{l'\in G_i} A(l',t_0)a^{l'} \right\rangle \\
&= \frac{M}{N}\left(1-\frac{M}{N}\right) \sum_{l\in G_i} A(l,t_0) \sum_{l'\in G_i} \overline{A(l',t_0)} \left\langle a^l, a^{l'} \right\rangle \\
&\le \frac{M}{N} \sum_{l\in G_i} |A(l,t_0)| \sum_{l'\in G_i} |A(l',t_0)| \cdot \left| \left\langle a^l, a^{l'} \right\rangle \right| \\
&\le \mu(A)\frac{M}{N} \sum_{l\in G_i} |A(l,t_0)| \sum_{l'\in G_i} \left| \left\langle a^l, a^{l'} \right\rangle \right| \\
&\le \mu(A)\frac{M}{N} \sum_{l\in G_i} |A(l,t_0)| \sum_{l'\in G_i} \frac{\left| \left\langle a^l, \frac{a^{l'}}{\|a^{l'}\|} \right\rangle \right|}{\|a^l\|} \cdot \|a^l\| \cdot \left\|a^{l'}\right\| \\
&\le \mu^3(A)|T|\frac{M}{N} \sum_{l\in G_i} |A(l,t_0)| \sum_{l'\in G_i} \frac{\left| \left\langle a^l, \frac{a^{l'}}{\|a^{l'}\|} \right\rangle \right|}{\|a^l\|} \\
&\le \mu^3(A)|T|\frac{M}{N}\gamma \sum_{l\in G_i} |A(l,t_0)|. \qquad \text{(B.6)}
\end{aligned}
$$

Putting together (B.5) and (B.6), we get

$$
E||v^0||^2 \le \frac{M}{N}\mu^3(A)|T|\gamma \sum_{i=1}^{N/g} \sum_{l'\in G_i} |A(l',t_0)| = \frac{M}{N}\mu^3(A)|T|\gamma \sum_{l=1}^{N} |A(l,t_0)| = \frac{M}{N}\mu^3(A)|T|\gamma \left\|A(:,t_0)\right\|_1 .
$$

Now since $A^H A = I$, we have $\left\|A(:,t_0)\right\|_2 = 1$, and since for any vector h of length N we have $||h||_1 \le \sqrt{N}||h||_2$, it follows that

$$
E||v^0||^2 \le \frac{M}{\sqrt{N}}\mu^3(A)|T|\gamma, \qquad \text{(B.7)}
$$

which proves Lemma 1.

APPENDIX C

PROOF OF LEMMA 2

By definition,

$$||v^0|| = \sup_{||f||=1} \left\langle v^0, f \right\rangle = \sup_{||f||=1} \sum_{i=1}^{N/g} \langle Y_i, f \rangle , \tag{C.1}$$

with v^0 from (B.3) and Y_i from (B.4). For completeness, we reproduce below [56, Theorem 3.2], which we use to prove Lemma 2.

Theorem 6. *Let $Y_1, \ldots, Y_N$ be a sequence of independent random variables taking values in a Banach space and let Z be the supremum $Z = \sup_{f \in \mathcal{F}} \sum_{n=1}^{N} f(Y_i)$, where $\mathcal{F}$ is a countable family of real-valued functions. Assume that $|f(Y)| < B$ for every $f \in \mathcal{F}$ and all Y, and $\mathbb{E}f(Y_i) = 0$ for every $f \in \mathcal{F}$ and $i = 1, \ldots, N$. Then, for all $t \geq 0$,*

$$P\left(|Z - \mathbb{E}Z| > t\right) \leq 3 \exp\left(\frac{-t}{KB} \log\left(1 + \frac{Bt}{\sigma^2 + B\mathbb{E}\bar{Z}}\right)\right),$$

where

$$\sigma^2 = \sup_{f \in \mathcal{F}} \sum_{i=1}^{N} \mathbb{E}f^2(Y_i),$$
$$\bar{Z} = \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^{N} f(Y_i) \right|,$$

and K is a numerical constant.

Denote the mapping $\langle Y_i, f \rangle$ for a fixed unit vector f as $f(Y_i)$, so that $\bar{Z} = \sup_{||f||=1} \sum_{i=1}^{N/g} f(Y_i) = ||v^0||$. We have $\mathbb{E}\{f(Y_i)\} = 0$, and

$$|f(Y_i)| = \left| \left\langle \left(\delta_i - \frac{M}{N}\right) \sum_{l \in G_i} A(l, t_0) a^l, f \right\rangle \right|$$

$$
\begin{aligned}
&= \left| \left(\delta_i - \frac{M}{N} \right) \sum_{l \in G_i} A(l, t_0) \left\langle a^l, f \right\rangle \right| \\
&< \left| \sum_{l \in G_i} A(l, t_0) \left\langle a^l, f \right\rangle \right| \\
&\leq \max_{l \in G_i} |A(l, t_0)| \cdot \max_{l \in G_i} ||a^l|| \cdot \sum_{l \in G_i} \frac{|\langle a^l, f \rangle|}{||a^l||} \\
&\leq \gamma \cdot \mu^2(A) \cdot \sqrt{|T|} =: B. \qquad \text{(C.2)}
\end{aligned}
$$

Now we find a bound on $\mathbb{E}\{f^2(Y_i)\}$:

$$
\begin{aligned}
\mathbb{E}\{f^2(Y_i)\} &= \mathbb{E}\left\{ \left| \left\langle \left(\delta_i - \frac{M}{N} \right) \sum_{l \in G_i} A(l, t_0) a^l, f \right\rangle \right|^2 \right\} = \operatorname{var}\{\delta_i\} \cdot \left| \sum_{l \in G_i} A(l, t_0) \langle a^l, f \rangle \right|^2 \\
&= \frac{M}{N} \left(1 - \frac{M}{M} \right) \cdot \left| \sum_{l \in G_i} A(l, t_0) \left\langle a^l, f \right\rangle \right|^2 \leq \frac{M}{N} \mu^2(A) \sum_{l \in G_i} \left| \langle a^l, f \rangle \right|^2,
\end{aligned}
$$

and so

$$
\sum_{i=1}^{N/g} \mathbb{E}\{f^2(Y_i)\} \leq \mu^2(A) \frac{M}{N} \sum_{i=1}^{N/g} \sum_{l \in G_i} |\left\langle a^l, f \right\rangle|^2.
$$

We know that $\sum_{i=1}^{N/g} \sum_{l \in G_i} |\langle a^l, f \rangle|^2 = 1$; therefore,

$$
\sum_{i=1}^{N/g} \mathbb{E}\{f^2(Y_i)\} \leq \frac{M}{N} \mu^2(A) =: \sigma^2. \qquad \text{(C.3)}
$$

Plugging (C.2) and (C.3) in Lemma 1, we have

$$
\mathbb{E}\{\bar{Z}\} = \mathbb{E}\{||v^0||\} \leq \mu^{3/2}(A) \sqrt{|T|} \sqrt{\gamma} \frac{\sqrt{M}}{N^{1/4}}. \qquad \text{(C.4)}
$$

Assume that

$$
\frac{\sqrt{\gamma} |T| \mu^{3/2}(A) N^{3/4}}{\sqrt{M}} < 1 \quad ; \quad 0 < a \leq \frac{\sqrt{M}}{\sqrt{\gamma} \mu(A) \sqrt{N} \sqrt{|T|}};
$$

then with (5.10), we have

$$\bar{\sigma}^2 = \gamma\mu^2(A)\frac{M}{N} > B\mathbb{E}\{\bar{Z}\} = B\mathbb{E}\{\left\|v^0\right\|\} > B\mu^{3/2}(A)\sqrt{\gamma}\sqrt{|T|}\frac{\sqrt{M}}{N^{1/4}}, \tag{C.5}$$

and by writing $t = a\bar{\sigma}$ we have

$$Bt = Ba\bar{\sigma} \leq \bar{\sigma}^2. \tag{C.6}$$

For $\frac{\sqrt{\gamma}|T|\mu^{3/2}(A)N^{3/4}}{\sqrt{M}} > 1$ and $0 < a \leq \left(\frac{M}{\gamma\mu(A)\sqrt{N}}\right)^{1/4}$, with (5.10) we have

$$\bar{\sigma}^2 = \gamma^{3/2}\mu^{7/2}(A)|T|\sqrt{M} = B\mathbb{E}\{\bar{Z}\} = B\mathbb{E}\{\left\|v^0\right\|\} = B\mu^{3/2}(A)\sqrt{\gamma}\sqrt{|T|}\frac{\sqrt{M}}{N^{1/4}}$$

and so

$$Bt \leq \bar{\sigma}^2. \tag{C.7}$$

Putting together (C.6), (C.7), and Theorem 6, we can write

$$P\left(||v^0|| > \mu^{3/2}(A)N^{-1/4}\sqrt{\gamma M|T|} + a\bar{\sigma}\right) < 3e^{-\kappa a^2},$$

where κ is a numerical constant $\kappa = \frac{\log(1.5)}{K}$ and K comes from Theorem 6. This completes the proof of Lemma 2.

APPENDIX D

PROOF OF LEMMA 3

Denote the events

$$E_1 : \left\{ \left\| A_{\Omega T}^H A_{\Omega T} \right\| \geq \frac{M}{2N} \right\}$$

and

$$E_2 : \left\{ \sup_{t_0 \in T^c} ||v^0|| \leq \mu^{3/2}(A) N^{-1/4} \sqrt{\gamma M |T|} + a\bar{\sigma} \right\}.$$

We can write

$$P\left(\sup_{t_0 \in T^c} ||w^0|| \geq 2N^{3/4}\mu^{3/2} \sqrt{\frac{\gamma |T|}{M}} + \frac{2Na\bar{\sigma}}{M} \right) \leq P(\overline{E_1 \cap E_2}) = P(\overline{E_1} \cup \overline{E_2}) \leq P(\overline{E_1}) + P(\overline{E_2}),$$

and with Lemma 2 we have

$$P\left(\sup_{t_0 \in T^c} ||w^0|| \geq 2N^{3/4}\mu^{3/2} \sqrt{\frac{\gamma |T|}{M}} + \frac{2Na\bar{\sigma}}{M} \right) \leq P\left(||A_{\Omega T}^H A_{\Omega T}|| \leq \frac{M}{2N} \right) + 3e^{-\kappa a^2}, \tag{D.1}$$

which proves Lemma 3.

BIBLIOGRAPHY

[1] "Data Sheet, SKYWORKS SKY65006-348LF," http://www.skyworksinc.com/uploads/documents/200122H.pdf.

[2] "Internet Crime Complaint Center annual reports," http://www.ic3.gov/media/annualreports.aspx.

[3] "Data Sheet, MAXIM MAX2242, 2.4GHz to 2.5GHz Linear Power Amplifier," http://datasheets.maxim-ic.com/en/ds/MAX2242.pdf.

[4] "Data Sheet, ADF4360-1, integrated synthesizer and VCO," http://www.analog.com/static/imported-files/data_sheets/ADF4360-1.pdf.

[5] R.W. Jackson and M. Shusta, "Interference rejection by time selective sampling," in *43rd European Microwave Conference*, Oct. 2013, pp. 573–576.

[6] "National Center for Missing Exploited and Children. Child Pornography Fact Sheet," http://www.missingkids.com/missingkids/servlet/PageServlet?LanguageCountry=enUS&PageId=2451.

[7] "Business software alliance 2009 year in review," http://www.bsa.org/~/media/Files/General/YIR2009.ashx.

[8] "Federal Trade Commission 2006 Identity Theft Survey Report," http://www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf.

[9] Francois Paget, "Financial fraud and internet banking: Threats and countermeasures," http://www.mcafee.com/us/local_content/reports/6168rpt_fraud_0409.pdf.

[10] N.T. Nguyen, G. Zheng, Z. Han, and R. Zheng, "Device fingerprinting to enhance wireless security using nonparametric Bayesian method," in *INFOCOM, 2011 Proceedings IEEE*, April 2011, pp. 1404–1412.

[11] L.E. Langley, "Specific emitter identification (sei) and classical parameter fusion technology," in *Conference Record, WESCON/'93.*, Sep 1993, pp. 377–381.

[12] K.I. Talbot, P.R. Duley, and M.H. Hyatt, "Specific emitter identification and verification," *Technology Review*, 2003.

[13] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "Fingerprints in ether: Using the physical layer for wireless authentication," in *IEEE Int. Conf. Communications (ICC)*, June 2007, pp. 4646–4651.

[14] L. Xiao, L. Greenstein, N.B. Mandayam, and W. Trappe, "Channel-based detection of sybil attacks in wireless networks," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, pp. 492–503, September 2009.

[15] Z. Li, W. Xu, R. Miller, and W. Trappe, "Securing wireless systems via lower layer enforcements," in *Proceedings of the 5th ACM workshop on Wireless security*, 2006, pp. 33–42.

[16] N. Patwari and S.K. Kasera, "Robust location distinction using temporal link signatures," in *ACM MOBICOM*, 2007, pp. 111–122.

[17] D.B. Faria and D.R. Cheriton, "Radio-layer security: Detecting identity-based attacks in wireless networks using signalprints," in *5th ACM Workshop on Wireless Security (WiSe'06)*, September 2006, pp. 43–52.

[18] Y. Sheng, K. Tan, G. Chen, D. Kotz, and A. Campbell, "Detecting 802.11 mac layer spoofing using received signal strength," in *IEEE INFOCOM 2008. The 27th Conference on Computer Communications.*, April 2008.

[19] O. Ureten and N. Serinken, "Bayesian detection of radio transmitter turn-on transients," in *NSIP99*, 1999, pp. 830–834.

[20] K. Ellis and L. Serinken, "Characteristics of radio transmitter fingerprints," *Journal of Radio Science*, vol. 36, no. 4, pp. 585–597, Dec 2001.

[21] J. Hall, M. Barbeau, and E. Kranakis, "Detection of transient in radio frequency fingerprinting using phase characteristics of signals," in *3rd IASTED International Conference on Wireless and Optical Communications (WOC)*, July, pp. 13–18.

[22] O.H. Tekbas, N. Serinken, and O. Ureten, "An experimental performance evaluation of a novel radio-transmitter identification system under diverse environmental conditions," *Canadian Journal of Electrical and Computer Engineering Canadian Journal of*, vol. 29, no. 3, pp. 203–209, July 2004.

[23] J. Hall, M. Barbeau, and E. Kranakis, "Enhancing intrusion detection in wireless networks using radio frequency fingerprinting," in *3rd IASTED International Conference on Communications, Internet and Information Technology (CIIT)*, November 2004.

[24] J. Hall, "Detection of rogue devices in wireless networks," PhD Dissertation, School of Computer Science, Carleton University, Ottawa, Ontario, 2006.

[25] K.B Rasmussen and S. Capkun, "Implications of radio fingerprinting on the security of sensor networks," in *Third International Conference on Security and Privacy in Communications Networks and the Workshops, 2007. SecureComm 2007.*, Sept 2007, pp. 331–340.

[26] O. Ureten and N. Serinken, "Wireless security through rf fingerprinting," *Canadian Journal of Electrical and computer Engineering*, vol. 32, No. 1, pp. 27–33, Winter 2007.

[27] T. Kohno, A. Broido, and K.C. Claffy, "Remote physical device fingerprinting," in *IEEE Transactions on Dependable and Secure Computing*, 2005, pp. 93–108.

[28] A.A. Tomko, C.J. Rieser, and L.H. Buell, "Physical-layer intrusion detection in wireless networks," in *IEEE Military Communications Conference, 2006.*, Oct 2006, pp. 1–7.

[29] R.M. Gerdes, T.E. Daniels, M. Mina, and S.F. Russel, "Device identification via analog signal fingerprinting: A matched filter approach," in *13th Annual Network and Distributed System Security Symposium (NDSS)*, Feb 2006.

[30] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *14th ACM international conference on Mobile computing and networking*, March 2008, pp. 116–127.

[31] A. Candore, O. Kocabas, and F Koushanfar, "Robust stable radiometric fingerprinting for wireless devices," in *Hardware-Oriented Security and Trust, 2009. HOST '09. IEEE International Workshop on*, July 2009, pp. 43–49.

[32] I.O. Kennedy, P. Scanlon, F.J. Mullany, M.M. Buddhikot, K.E. Nolan, and T.W. Rondeau, "Radio transmitter fingerprinting: A steady state frequency domain approach," in *IEEE 68th Vehicular Technology Conference, 2008. VTC 2008-Fall.*, Sept 2008, pp. 1–5.

[33] B. Danev, H. Luecken, S. Capckun, and K. El Defrawy, "Attacks on physical-layer identification," in *ACM Conference on Wireless Network Security (WiSec10)*, March 2010, pp. 89–98.

[34] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, pp. 201–220, February 2005.

[35] Q. Weng, *Advances in Environmental Remote Sensing: Sensors, Algorithms, and Applications*, CRC Press, 2011.

[36] J.M. Munoz-Ferreras, F. Perez-Martinez, J. Calvo-Gallego, A. Asensio-Lopez, B.P. Dorta-Naranjo, and A. Blanco del Campo, "Traffic surveillance system based on a high-resolution radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, pp. 1624–1633, June 2008.

[37] G.C. Tavik, C.L. Hilterbrick, J.B. Evins, J.J. Alter, Jr. J.G. Crnkovich, J.W. de Graaf, W. Habicht II, G.P. Hrin, S.A. Lessin, D.C. Wu, and S.M. Hagewood, "The advanced multifunction rf concept," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 1009–1020, March 2005.

[38] "802.22 standard," http://www.ieee802.org/22/.

[39] Google Inc., ," https://www.google.com/get/spectrumdatabase/.

[40] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, February 2006.

[41] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1298–1306, April 2006.

[42] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, pp. 118–121, April 2007.

[43] J.A. Tropp, J.N. Laska, M.F. Duarte, J.K. Romberg, and R. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Transactions on Information Theory*, vol. 56, pp. 520–544, January 2010.

[44] S. Kirolos, J.N. Laska, M.B. Wakin, M.F. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk, "Analog to information conversion via random demodulation," in *IEEE Dallas/CAS Workshop on Design, Applications, Integration and Software*, October 2006.

[45] J.N. Laska, S. Kirolos, M.F. Duarte, T. Ragheb, R. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information conversion using random demodulation," in *IEEE Intl. Symp. Circuits and Systems (ISCAS)*, May 2007.

[46] J.N. Laska, S. Kirolos, Y. Massoud, R. Baraniuk, A.C. Gilbert, M. Iwen, and M.J. Strauss, "Random sampling for analog-to-information conversion of wideband signals," in *IEEE Dallas/CAS Workshop on Design, Applications, Integration and Software*, October 2006.

[47] T. Ragheb, S. Kirolos, J. Laska, A. Gilbert, M. Strauss, R. Baraniuk, and Massoud Y., "Implementation models for analog-to-information conversion via random sampling," in *Midwest Symp. Circuits and Systems (MWSCAS)*, August 2007.

[48] M. Davenport, S. Schnelle, J. Slavinsky, R. Baraniuk, M. Wakin, and P. Boufounos, "A wideband compressive radio receiver," in *Military Communications Conference*, November 2010.

[49] M.A. Davenport J.R. Treichler and R. Baraniuk, "Application of compressive sensing to the design of wideband signal acquisition receivers," in *U.S. / Australia Joint Workshop on Defense Applications of Signal Processing (DASP)*, September 2009.

[50] J.P. Slavinsky, J.N. Laska, M.A. Davenport, and R. Baraniuk, "The compressive multiplexer for multi-channel compressive sensing," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011.

[51] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *IEEE Journal of Selected Topics on Signal Processing*, vol. 4, pp. 375–391, April 2010.

[52] M. Mishali, Y.C. Eldar, O. Dounaevsky, , and E. Shoshan, "Xampling: Analog-to-digital at sub-nyquist rates," *IET Circuits, Devices, and Systems*, vol. 5, pp. 8–20, January 2011.

[53] S. Becker, J. Yoo amd M. Loh, A. Emami-Neyestanak, and E. J. Candès, "Practical design of a random demodulation sub-nyquist adc," in *Workshop on Signal Processing with Adaptive Sparse Structured Representations*, June 2011.

[54] S. Hoyos B. M. Sadler X. Chen, Z. Yu and J. Silva-Martinez, "A sub-nyquist rate sampling receiver exploiting compressive sensing," *IEEE Transactions on Circuits an Systems I: Regular Papers*, vol. 58, pp. 507–520, March 2011.

[55] S. Ravindran R. Bland P. Pace G. Fudge, M. Chivers and J. Haupt, "A nyquist folding analog-to-information receiver," in *Asilomar Conf. on Signals, Systems, and Computers*, October 2008.

[56] E. J. Candès and J. K. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, pp. 969–985, 2007.

[57] M. Duarte, M. Davenport, M. Wakin, and R. Baraniuk, "Sparse signal detection from incoherent projections," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2006.

[58] M. Davenport, P. Boufounos, and R. Baraniuk, "Compressive domain interference cancellation," in *Workshop on Signal Processing with Adaptive Sparse Structured Representations*, April 2009.

[59] S. Hwang, S. Jang, D. Kim, and J. Seo, "An mmse-based compressive domain interference cancellator for wideband systems," in *International Conference on Computer and Automation Engineering*, February 2010.

[60] J.N. Laska, P. T. Boufounos, M.A. Davenport, and R. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 31, pp. 429–443, November 2011.

[61] J.N. Laska, M.A. Davenport, and R. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *Asilomar Conference on Signals, Systems and Computers*, November 2009.

[62] J. Bamford, *Body of Secrets: Anatomy of the Ultra-Secret National Security Agency*, Random House, New York, 2001.

[63] "Everything we learned from Edward Snowden in 2013," http://www.nationaljournal.com/defense/everything-we-learned-from-edward-snowden-in-2013-20131231.

[64] G.I. Radulov, M. Heydenreich, R.W. van der Hofstad, J.A. Hegt, and A.H.M. van Roermund, "Brownian-bridge-based statistical analysis of the dac inl caused by current mismatch," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 54, no. 2, pp. 146–150, February 2007.

[65] T.W. Anderson and M.A. Stevens, "The continious and discrete brownian bridges: Representations and applications," *Linear Algebra and its Applications*, vol. 264, pp. 145–171, October 1997.

[66] P. Wambacq and W. Sansen, *The Distortion Analysis of Analog Integrated Circuits*, Kluwer, 1998.

[67] "Data Sheet, Analog Devices AD9776A/AD9778A/AD9779A," http://www.analog.com/static/imported-files/data_sheets/AD9776_9778_9779.pdf.

[68] A. Mehrnia, "Optimum DAC resolution for WMAN, WLAN and WPAN OFDM based standards," in *International Conference on Consumer Electronics*, 2005, pp. 355–356.

[69] "Data Sheet, MAXIM MAX5858 dual, 10-bit, 300msps digital to analog converter," http://datasheets.maxim-ic.com/en/ds/MAX5858.pdf.

[70] J.C. Pedro and S.A. Mass, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 1150 – 1163, April 2005.

[71] M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 348 – 359, January 2006.

[72] D. Schreus, *RF power amplifier behavioral modeling*, Cambridge University Press, 2008.

[73] John A. Gubner, *Probability and Random Processes for Electrical and Computer Engineers*, Cambridge University Press, 2006.

[74] Steven M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, 1988.

[75] H. Ombao, J. Raz, and R. Von Sachs, "A simple generalized crossvalidation method of span selection for periodogram smoothing," *Biometrika*, vol. 88, pp. 1186–1192, 2001.

[76] H. Van Trees, *Detection, Estimation and Modulation Theory*, John Wiley & Sons, Inc., 1968.

[77] R. Hassun, M. Flaherty, R. Matreci, and M. Taylor, "Effective evaluation of link quality using error vector magnitude techniques," in *IEEE Wireless Communications Conference*, August 1997, pp. 89–94.

[78] A.C. Polak, C. Dolatshahi, and D.L. Goeckel, "Identifying wireless users via transmitter imperfections," *IEEE Journal on Selected Areas in Communications*, vol. 29, pp. 1469–1479, August 2011.

[79] A. Demir, A. Mehrotra, and J. Roychowdhury, "Phase noise in oscillators: a unifying theory and numerical methods for characterization," *IEEE Transactions on Circuits and Systems*, vol. 47, pp. 655–674, May 2000.

[80] T.C.W. Schenk, *RF imperfections in high-rate wireless systems: impact and digital compensation*, Springer Publisher, 2008.

[81] A. Mehrotra, "Noise analysis of phase-locked loops," *IEEE Transactions on Circuits and Systems*, vol. 49, pp. 1309–1316, September 2002.

[82] T. W. Anderson and A. M. Walker, "On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process," *The Annals of Mathematical Statistics*, vol. 35, pp. 1296–1303, 1964.

[83] M. El-Tanany, Y. Wu, and L. Hazy, "Analytical modeling and simulation of phase noise interference in OFDM-based digital television terrestrial broadcasting systems," *IEEE Transactions on Broadcasting*, vol. 47, no. 1, pp. 20–31, March 2001.

[84] S. Bittner, S. Krone, and Fettweis G., "Tutorial on discrete time phase noise modeling for phase locked loops," `https://mns.ifn.et.tu-dresden.de/personalSites/stefan.krone/Documents/Bittner_S_PN_Tutorial.pdf`.

[85] G.A. Wright, "Magnetic resonance imaging," *IEEE Signal Processing Magazine*, vol. 14, no. 1, pp. 56–66, Jan 1997.

[86] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 72–82, Mar. 2008.

[87] J. A. Tropp, "Column subset selection, matrix factorization, and eigenvalue optimization," in *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New York, NY, Jan. 2009, pp. 978–986.

[88] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. Sci. Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[89] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, http://www.cs.ubc.ca/labs/scl/spgl1.

[90] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," `http://cvxr.com/cvx`, Apr. 2011.

[91] B.F. Logan Jr., "Information in the zero crossings of bandpass signals," *ATT Technical Journal*, vol. 56, pp. 487–510, Apr. 1977.

[92] P.T. Boufounos and R. Baraniuk, "Reconstructing sparse signals from their zero crossings," *IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP*, p. 3361 3364, Apr. 2008.

[93] N. Sharma and T.V. Sreenivas, "Sparse signal reconstruction based on signal dependent non-uniform samples," *IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP*, pp. 3453 – 3456, March 2012.

[94] A. J. Weinstein and M. B. Wakin, "Recovering a clipped signal in sparseland," *to appear in Sampling Theory in Signal and Image Processing.*, 2013.

[95] J.A. Tropp, J.N. Laska, M.F. Duarte, J.K. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[96] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution," *Commununications on Pure Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[97] E. Candès, M.B. Wakin, and S.P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, pp. 877–905, 2008.

[98] D. Falconer, S. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, pp. 58–66, Apr. 2002.

[99] D.C. Cox and R. Leck, "Distributions of multipath delay spread and average excess delay for 910-mhz urban mobile radio paths," *IEEE Transactions on Antennas and Propagation*, vol. 23, no. 2, pp. 206–213, Mar 1975.

[100] H.G. Myung, J. Lim, and D.J. Goodman, "Single carrier FDMA for uplink wireless transmission," *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 30–38, Sept. 2006.

[101] J. Zyren and W. McCoy, "Overview of the 3GPP Long Term Evolution physical layer," *Freescale Semiconductor White Paper*, July 2007.

[102] U. Kiencke and H Jakel, *Signale und Systeme*, Oldenbourg Verlag, 2008.

[103] A.P. Liavas, P.A. Regalia, and J.-P. Delmas, "Blind channel approximation: effective channel order determination," *IEEE Transactions on Signal Processing*, vol. 47, no. 12, pp. 3336–3344, Dec 1999.

[104] H. Tang, "Some physical layer issues of wide-band cognitive radio systems," in *IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Nov. 2005, pp. 151–159.

[105] J.A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, August 2011.