

RICE UNIVERSITY

Efficient Radiometric Signature Methods for Cognitive Radio Devices

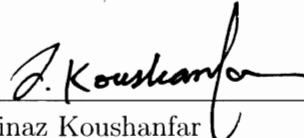
by

Övünç Kocabaş

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Master of Science

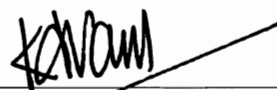
APPROVED, THESIS COMMITTEE:



Farinaz Koushanfar
Assistant Professor of Electrical and
Computer Engineering and Computer
Science



Richard Baraniuk
Victor E. Cameron Professor of Electrical
and Computer Engineering



Kartik Mohanram
Assistant Professor of Electrical and
Computer Engineering and Computer
Science

Houston, Texas

August, 2010

ABSTRACT

Efficient Radiometric Signature Methods for Cognitive Radio Devices

by

Övünç Kocabaş

This thesis presents the first comprehensive study and new methods for radiometric fingerprinting of the Cognitive Radio (CR) devices. The scope of the currently available radio identification techniques is limited to a single radio adjustment. Yet, the variable nature of the CR with multiple levels of parameters and adjustments renders the radiometric fingerprinting much more complex. We introduce a new method for radiometric fingerprinting that detects the unique variations in the hardware of the reconfigurable radio by passively monitoring the radio packets. Several individual identifiers are used for extracting the unique physical characteristics of the radio, including the frequency offset, modulated phase offset, in-phase/quadrature-phase offset from the origin, and magnitude. Our method provides stable and robust identification by developing individual identifiers (classifiers) that may each be weak (i.e., incurring a high prediction error) but their committee can provide a strong classification technique. Weighted voting method is used for combining the classifiers. Our hardware implementation and experimental evaluations over multiple radios demonstrate that our weighted voting approach can identify the radios with an average of 97.7% detection probability and an average of 2.3% probability of false alarm after testing only 5 frames. The probability of detection and probability of false alarms both rapidly improve by increasing the number of test frames.

Acknowledgments

First and foremost I wish to express my gratitude to my thesis supervisor Prof. Fari-naz Koushanfar for her valuable advise and guidance throughout my research. This thesis would not be complete without her complementary knowledge and inspiration.

I would like to thank also my thesis committee members Prof. Richard Baraniuk and Prof. Kartik Mohanram for their valuable reviews and comments on my thesis.

Furthermore I would like thank Texas Instruments for their financial support during my graduate study so that I can concentrate my research and finish my thesis.

Last but not the least, I would like to thank my beloved family and Gizem, for always being there and supporting me unconditionally. Without their support, this work could have never finished.

Contents

Abstract	ii
Acknowledgments	v
List of Illustrations	vii
List of Tables	ix
1 Introduction	1
2 Related Work	3
2.1 Location Based Identification	4
2.2 Radio-frequency Fingerprinting	6
3 Preliminaries	9
3.1 Cognitive Radios	9
3.2 Wireless Open-Access Research Platform (WARP)	11
3.2.1 WARPLab	12
4 Classifying Variables	16
4.1 Classifiers	16
4.2 Experiment Setup	20
4.3 Classifier Analysis	22
4.3.1 Phase Error	23
4.3.2 Magnitude Error	25
4.3.3 Error Vector Magnitude	25
4.3.4 I/Q Offset	25

	vi
4.3.5 Frequency Offset	38
5 Classification	42
5.1 Fingerprinting Mechanism	42
5.1.1 Signature Extraction	43
5.1.2 Combining Classifiers	46
5.2 Performance of Classifiers	50
5.2.1 Results with Office Environment	52
6 Conclusion	55
Bibliography	58

Illustrations

3.1	Spectrum Utilization [1]	10
3.2	Spectrum hole [2]	11
3.3	WARP Kit	13
3.4	WARP Hardware Platform	13
3.5	WARPLab Design Flow [3]	14
3.6	WARPLab Architecture [3]	15
4.1	Received signal from different boards	17
4.2	Received signal for different power levels (BPSK)	18
4.3	Received signal for different power levels (QPSK)	18
4.4	Error vector magnitude (EVM)	19
4.5	I/Q Offset	20
4.6	Experiment Setup	21
4.7	Probability Density Function estimation with histograms	24
4.8	Boxplots for Phase Error over different configurations	26
4.9	Frequency response for Phase Error	27
4.10	Power response for Phase Error	28
4.11	Boxplots over the different configurations of Magnitude Error	29
4.12	Frequency response for Magnitude Error	30
4.13	Power response for Magnitude Error	31
4.14	Boxplots over the different configurations of EVM	32
4.15	Frequency response for EVM	33

4.16	Power response for EVM	34
4.17	Boxplots over the different configurations of I/Q Offset	35
4.18	Frequency response for I/Q Offset	36
4.19	Power response for I/Q Offset	37
4.20	Boxplots over the different configurations of Frequency Offset	39
4.21	Frequency response for Frequency Offset	40
4.22	Power response for Frequency Offset	41
5.1	The flow of signature extraction and signature matching approach. . .	43
5.2	Probability density function of frequency offset classifier	45

Tables

5.1	KL distance of boards	45
5.2	P_D and P_{FA} for BPSK identifiers	46
5.3	P_D and P_{FA} for QPSK identifiers	46
5.4	α values for different configurations (BPSK)	49
5.5	α values for different configurations (QPSK)	49
5.6	Voting Example for BPSK	50
5.7	Combining classifiers: Voting and ML (BPSK)	51
5.8	Combining classifiers: Voting and ML (QPSK)	51
5.9	P_D and P_{FA} for BPSK identifiers	52
5.10	P_D and P_{FA} for QPSK identifiers	52
5.11	Combining classifiers: Voting and ML (BPSK)	53
5.12	Combining classifiers: Voting and ML (QPSK)	54

Chapter 1

Introduction

Rapid technological advances in wireless communication has enabled the transition from low content rate voice and telephony usage to high data rate seamless multimedia and interactive Internet applications. Proliferation of embedded tether-less appliances and the growing number of users demand new methods for identification, coexistence, and management of radio devices. Wireless radio identification is typically performed using the digital identifiers or keys. For example, cryptographic methods actively control accessing the device using the key-exchange protocols, or IP addresses are used for passively tracking the user access. While such methods provide the required level protection for many applications, they may be vulnerable for a number of applications, in particular when users have physical access to the devices. Vulnerabilities include extraction of digital keys by side-channel attacks, or replaying the IPs. To ensure identification certainty, a suit of radiometric fingerprinting methods that rely on unclonable minute variations of the physical radio were proposed in [4–12].

To manage the multiplicity of devices and address the growing application demand for higher bandwidth, innovative and complex radio technologies that can more efficiently use the available radio spectrum are being developed. The emerging cognitive radio (CR) engines sense the bandwidth and other physical layer properties and then use the sensed data to make intelligent situation-aware decisions about their operation. The CR is able to adaptively adjust its physical and link layer parameters thus,

is typically more complex than most of the radios presently in operation and use.

This thesis presents the first comprehensive study and new methods for radiometric fingerprinting of the CR devices. Radiometric fingerprinting based on transient signal analysis has been widely studied for electromagnetic characteristics and antenna-level correlation and properties. Recent work has demonstrated that radiometric identification can be done more accurately by differentiating the characteristics of the individual wireless frames in the modulation domain [11]. It is possible to further ameliorate the radio identification accuracy by employing symbol-based frame analysis and improved statistical classification methods [12].

The scope of the currently available radio identification techniques is limited to a single radio adjustment. The variable nature of the cognitive radio with multiple levels of parameters and adjustments renders the radiometric fingerprinting much more complex. For example, the error magnitude of the transmitted signal is commonly used as a metric for device identification. However, it is not clear if this value remains the same for different transmission power levels and various frequencies. While we initially study the modulation domain properties in the channel emulator environment where the channel impact is masked, our analysis also includes validation of the identifiers in indoor office environment.

Chapter 2

Related Work

In this chapter, a brief survey of related literature is provided which has influenced and inspired this work. Signal detection and identification source of an emitted signal is one the most challenging problems in wireless communication due to the broadcasting nature of the wireless communication which poses critical security threats.

The early research in signal detection and identification of an emitted signal goes back to 1960's where finding the source of a radar signal in military has been of the utmost importance. Several methods such as Special Emitter Identification (SEI) and Special Emitter Verification (SEV) techniques are developed for identifying and verifying the source of a received signal. While SEI technique [13,14] identifies source of a signal by matching the received waveform to unique emitter, SEV technique [13,14] is used for verification of the transmitter by looking at the external features of the signal where there is a priori knowledge of the transmitter. In these methods, identification is performed by measuring the unique features of the received signal, then these unique features are compared and mapped to existing clusters. Similar techniques are employed for also combating against fraud in cellular network [15, 16].

The importance of identifying source of a signal has been amplified with the emergence of wireless communication. During the past decade, with the help of vast advances in Integrated Circuit (IC) technology, the use of wireless devices in daily life has increased dramatically. Pervasive use of wireless devices and broadcasting nature of the wireless communication necessitate a secure medium for communication. In [17]

a risk analysis of threats for wireless communication is provided according to their implementation difficulty and potential impact on wireless network. Among these threats, impersonation attack is listed as one of the most critical attacks due to its ease of implementation by using off-the-shelf equipment. Various forms of this attack exist such as device cloning, address spoofing, unauthorized access and replay attacks. For instance, an attacker could spoof one of the device identities in wireless communication, MAC (*Media Access Control*) address, and use MAC address to access a network. This attack could be prevented by using cryptographic protocols, namely public-key cryptography. Public-key cryptography algorithms are effectively used for providing security and authentication mechanisms for wired networks. Yet implementation of this method has severe disadvantages for wireless networks. Public-key algorithms perform computationally heavy operations which will consume a lot of power and require considerable processing power. But most of the wireless devices are operating on battery power which makes battery life a critical issue for these devices. Also public-key algorithms require a key management infrastructure which incurs an overhead. For these reasons, lightweight and passive security mechanisms are needed for device identification in wireless networks.

A large body of work exists for identifying devices in wireless networks and these methods can be grouped into two classes based on their identification mechanism: location based identification, radio-frequency (RF) fingerprinting.

2.1 Location Based Identification

The wireless link between a transmitter and receiver depends on frequency, time and space therefore it is unique for a communication pair. Location based identification mechanisms [18–22] utilize the properties of this unique connection to identify devices.

Faria et al. [18] employ the received signal strength (RSS) for identification of devices. RSS is the measure for signal strength of received frame and depends on the power of the signal, wireless channel and distance between devices. Therefore, RSS values are unique for each transmitter and they cannot be forged by the attacker. Authors generate signalprints by using RSS values to identify devices and implement several attacks for IEEE 802.11 networks. Finally, they show that proposed signalprints can detect intruders with high probability.

In [20], authors use off-the-shelf air monitors (AM) to detect spoofing attacks. RSS of frames are captured via AMs and modeled as Gaussian Mixture Model (GMM). GMM is proposed to distinguish between two signals from different transmitters which have frequency and spatial variation of multipath channel at the receiver. Authors propose an expectation-maximization algorithm based on RSS profiles. They test their algorithm with 20 AMs and report 3% false positive and up to 98.7% detection rate.

Li et al. [22] analyze multi-path effects of the wireless channel for identification. Their method relies on the fact that the channel between transmitter and receiver will show unique properties. Authors propose authentication and confidentiality schemes on USRP/GNURadio SNR platform and show that they can detect spoofing attacks. In [19], a robust location distinction mechanism is proposed by exploiting physical layer characteristic of the radio channel between a transmitter and a receiver. A temporal link signature, which is the sum of the effects of the multiple paths from the transmitter to the receiver, each with its own time delay and complex amplitude, is generated. Authors show that the signature changes with the device location and can be used to identify transmitters.

Chen et al. [21] propose a method which relies on an attack detector based on

statistical testing of RSS values. An attack is detected by looking at the distances between centroids obtained via K-means cluster of RSS data. Authors evaluate their method both on Wi-Fi and 802.15.4 (Zigbee) networks and show that their method can achieve 95% of detection rate and less than 5% false positive rate.

Although location based identification methods provide a mechanism for identifying different transmitters, the major shortcomings of these methods are that devices are assumed stable at one location and the probability of detection is highly dependent on distance of the devices.

2.2 Radio-frequency Fingerprinting

RF fingerprinting methods are based on physical properties of the devices for identification. Device identification by the unique variations in the physical properties has been subject of research in integrated circuits. The work in [23] uses the delay variation of CMOS logic components to extract a digital secret. The properties of reconfigurable platforms are used in [24] to build a secure and robust authentication system based on the present delay variations. A post-fabrication nondestructive gate-level characterization for IC identification is also presented in [25]. Similarly, RF fingerprinting methods rely on the hardware imperfections inherited during the fabrication to uniquely identify wireless devices. Due to the manufacturing variability, devices have minute imperfections which will cause deviations in the transmitted signals. These deviations are hard to model and forge therefore they can be used for device identification.

In one of the early works [4], the electromagnetic (EM) signal transmitted by different WLAN (*wireless local area network*) cards is analyzed and shown that they have distinct properties due to manufacturing variability and antenna topology. The

authors investigate EM signals from 6 different card and able to generate a unique signature for each card.

Gerdes et al. [5] use a different approach and utilize matched filter to generate profiles based on signal to noise ratio (SNR) with the Gaussian noise presence. They show that by using a conventional matched filter, transmitters can be uniquely identified.

A large body of work exists in the literature [6–10] which utilizes the transient behavior of the wireless devices. A transient behavior is the anomaly observed when a device changes its state, such as activation or turn-on. This behavior is characteristic to each device and related with the components (i.e. phase-lock-loop (PLL), modulators, amplifiers, antennas) of the device. An RF fingerprint is generated by detecting and extracting the transient signal. First, signal is captured and initial point of the transient is detected. Then a fingerprint is generated by extracting the features of the transient signal. Features of the transient signal can be extracted by investigating instantaneous phase and amplitude of the received signal. The key part of this method relies on detecting the transient which is challenging due to noise present in the received signal. Therefore, transition point from noise to transient should be correctly identified for unique identification of devices. In [6, 7], a Bayesian detector is used for estimating the starting point of the transient. Received signal is assumed to have two different gaussian distribution, one for noise and the other for transient signal. The transition from noise to transient then estimated by looking at the drastic changes in mean and the variance which can be detected by using Bayesian detector. In [10], authors use the variance trajectory of instantaneous amplitude and phase for transient detection. They extract power spectral density fingerprints and use spectral correlation for classification. Finally they report 80% detection accuracy of their

experiments with 3 devices by collecting 802.11a OFDM signals. Tekbas et al. [9] investigate the effects of environmental conditions (i.e. power and temperature) on RF fingerprinting methods. Authors use variance dimensions for detecting transient signal, then probabilistic neural network (PNN) method is used for classification. 10 different VHF radio transmitters are tested by varying the power and temperature. Results show that fingerprints are susceptible to change of environment.

A different and recent method is proposed which uses modulation domain features by Brik et al. [11] instead of using transients for radio-frequency fingerprinting. Authors claim that transmitters can be uniquely identified by looking at deviations of emitted signal from ideal I/Q plane. The main reasons for the deviations from ideal I/Q domain are channel effect, noise at the receiver and hardware imperfections. Hardware imperfections are related with the manufacturing variability of devices and can be used for identifying transmitters uniquely. Authors propose using five radiometric identity metrics based on deviations for identification. The metrics for identification process are as follows: frequency error, SYNC correlation, I/Q offset, magnitude error and phase error. Finally authors use machine learning algorithms for classification of transmitters according to these metrics. K-nearest neighbor (kNN) and supporting vector machines (SVM) are implemented for classification. Authors perform experiments with identical 130 NIC (*network interface card*) cards and show that their method can differentiate transmitters with 99% accuracy.

Chapter 3

Preliminaries

In this chapter, first a background information will be provided about cognitive radios. Then our experiment platform, Wireless Open-Access Research Platform (WARP), will be introduced briefly.

3.1 Cognitive Radios

Ever increasing usage of mobile wireless devices and temporal-spatial inefficiency of using licensed spectrum necessitate a paradigm shift for wireless communication. Currently, wireless spectrum is controlled via fixed spectrum assignment policy, which assigns the portions of the spectrum to licensed users. Yet, Federal Communications Commission (FCC) reports [1] show that fixed assignment policy is inefficient due to temporal and spatial variations which is also depicted in Figure 3.1. A dynamic access of spectrum is required as the spectrum, which is the lifeline for wireless communication, is utilized inefficiently. Initiatives have been taken such as, next generation networks which is also called Dynamic Spectrum Access network for implementing policy based intelligent radios is proposed by DARPA [26, 27]. Most recently, FCC announces [28] opening up TV white spaces for mobile devices on unlicensed basis.

The key enabling technology for the initiatives is cognitive radio (CR), which is capable of sharing the spectrum in an opportunistic manner. CR concept is first proposed by J. Mitola [29] where it is defined as a software-defined radio which can

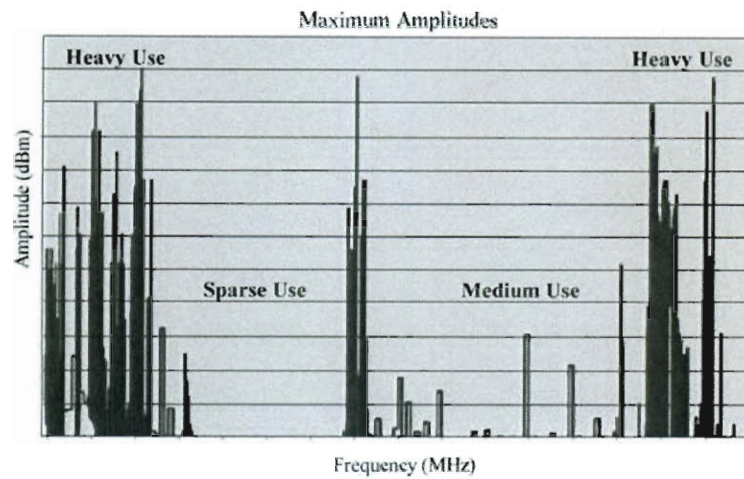


Figure 3.1 : Spectrum Utilization [1]

adjust its parameters depending on the spectrum status. The ultimate goal of CR is to determine best available spectrum without interfering the licensed users. Sharing the spectrum with the primary users poses a challenge which requires CR to be always aware of its environment and to find the temporarily unused portions of the spectrum, which is called *spectrum hole* or *white spaces* [2] as shown in Figure 3.2. Thus, main functionalities of CR can be listed as follows [30]:

- Finding white spaces and sharing it without interfering with primary users (*Spectrum sensing*)
- Determining the best available spectrum to meet communication requirements (*Spectrum management*)
- Continuation of seamless communication during a transition to another white space (*Spectrum mobility*)
- Sharing the available white space with other CR users (*Spectrum sharing*)

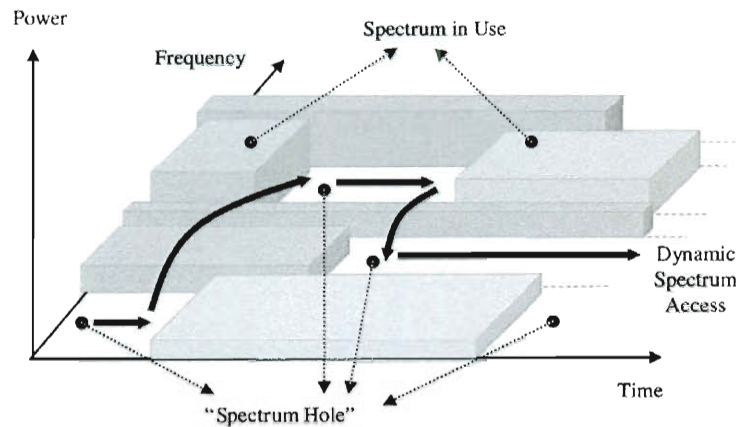


Figure 3.2 : Spectrum hole [2]

As of today, a number of different CR hardware platforms have been proposed. Most of the case, architectures rely on an versatility of an FPGA device to enable reconfigurability and flexibility of the platform. A multi-FPGA testbed for physical and network layers of CR is proposed in [31, 32]. A multiprocessor system-on-chip (MPSoC) design is introduced in [33], in which a system level design methodology is adopted to map a CR on a platform.

3.2 Wireless Open-Access Research Platform (WARP)

Wireless Open-Access Research Platform (WARP) [3] is a scalable and extensible programmable platform designed for prototyping and implementing wireless networks. WARP is an open-access research platform which enables sharing and exchanging wireless network architectures for developing next-generation wireless networks. As of today, WARP has been adopted in more than 50 research groups and is one of the most widely used platform in wireless network research [3].

The WARP Platform is composed of four parts: platform support packages, open-

access repository, research applications and custom hardware [34]. Platform support packages contain design tools for hardware/software design and open-access repository is a collection of the source codes and hardware design files. Algorithm implementations via WARP Platform are shared via research applications.

The custom hardware consists of 3 main components: an FPGA Board, Radio Boards and Clock Board. A fully equipped WARP hardware kit is presented in Figure 3.3. Hardware platform is centered around Virtex-II Pro FPGA Board which is depicted on Figure 3.4. The FPGA board has both configurable logic blocks (CLB) and PowerPC cores. While real-time Digital Signal Processing (DSP) applications which require high-speed communication are implemented by CLB's, PowerPC cores are used for executing network layer protocols developed in C and providing flexible interface between physical (PHY) and Media Access Control (MAC) layer. Radio Board are integrated to FPGA Board via daughtercard slots. The Radio Board supports 2.4 and 5 GHz ISM/UNII bands and capable of performing wideband applications such as OFDM. Clock Board provides clock signal to all boards and contains 2 parts. First part generates signal for radio boards, while second part supplies the signal for FPGA logic and analog converters.

3.2.1 WARPLab

Our experiments are performed with WARP FPGA Boards and WARPLab framework. WARPLab is a non-real time communication framework designed for rapid physical layer prototyping [3]. The framework uses MATLAB and WARP FPGA Board interactively. WARP boards are controlled via MATLAB workspace where user can generate signals, then these signals are transmitted by WARP boards. While wireless communication is done in real time by transmitting signal on the air, all

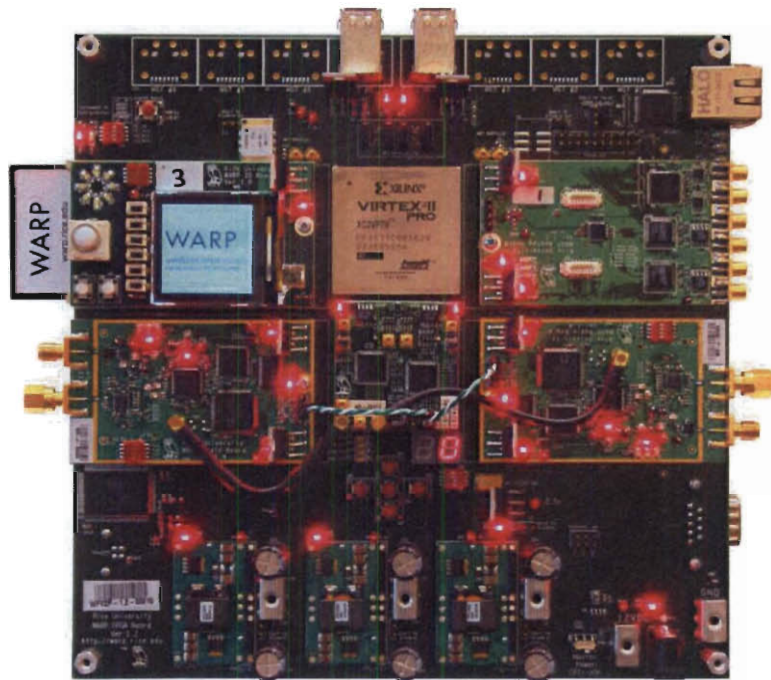


Figure 3.3 : WARP Kit

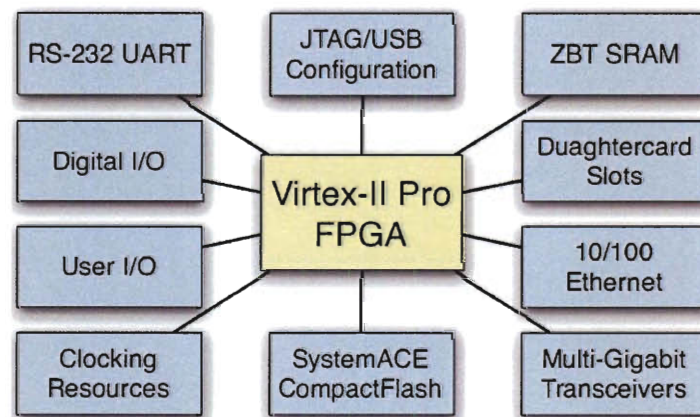


Figure 3.4 : WARP Hardware Platform

the data processing is performed off-line with MATLAB. The design flow for the WARPLab is shown in Figure 3.5. First, signal is generated in the MATLAB by user,

then generated signal is sent to WARP board via Ethernet. WARP board downloads the signal and stores in its buffer. Once trigger signal is sent, WARP board sends the signal in the air and receiver board captures the signal in real time and stores in the buffer. Finally, captured signal is sent to MATLAB for processing via Ethernet.

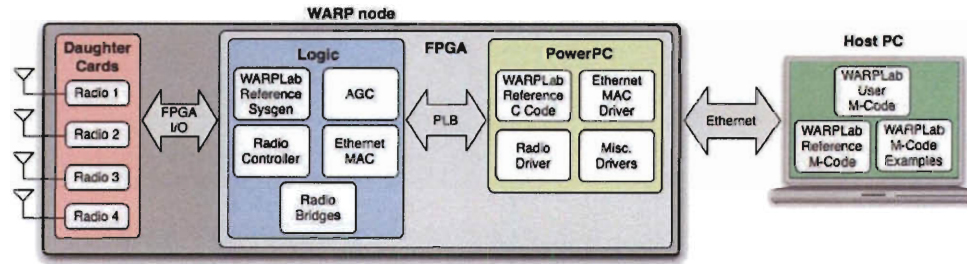


Figure 3.5 : WARPLab Design Flow [3]

The hardware architecture of the WARPLab is composed of two parts, transmitter and receiver, which is presented in Figure 3.6. In the transmitter part, signal is received via Ethernet and stored in Tx I/Q buffers. Then signal is converted to analog by Digital to Analog I/Q Converters (DAC), and analog signal is amplified with Transmitter Base-Band Amplifiers. Baseband signal is upconverted to RF signal through Phase-Locked Loop (PLL). Finally, upconverted signal is amplified with Transmitter RF amplifiers. All the parameters for Base-Band Amplifier, PLL and RF amplifiers are adjustable and can be set by user within MATLAB. For the receiver part, first signal is captured on the air, then it goes through Receiver RF amplifier. The RF signal is downconverted to baseband signal via PLL, then the signal strength is adjusted with Receiver Base-Band Amplifiers. Finally, analog signal is converted to digital by Analog to Digital I/Q (ADC) converters and stored in the Receiver I/Q Buffers. The values then send to MATLAB via Ethernet for further processing.

Like transmitter part, all the parameters for Receiver Base-Band Amplifier, PLL and Receiver RF amplifier can be controlled by user. Receiver part also has a Received Signal Strength Indicator (RSSI), which measures the signal strength of the signal.

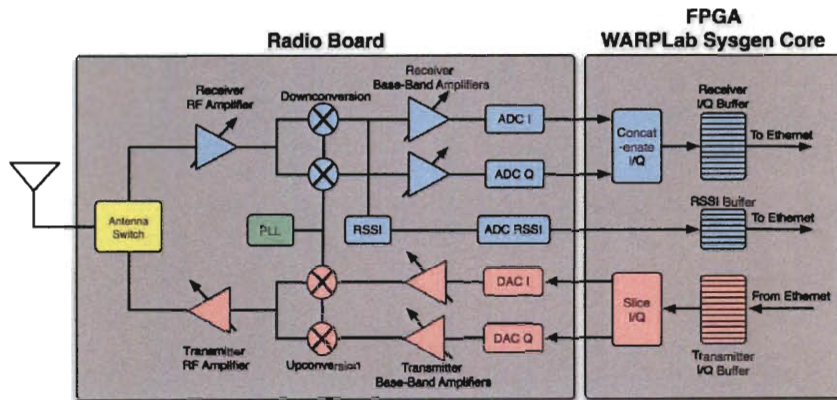


Figure 3.6 : WARPLab Architecture [3]

Chapter 4

Classifying Variables

In this chapter, we introduce our classifiers first which will be integral part of our fingerprinting mechanism. These classifiers will be used later in Chapter 5 to generate signatures for cognitive radio devices. Experiment setup which is used during classifier extraction will be explained next. Finally the response of classifiers for different cognitive radio configurations (i.e. modulation, power and frequency) will be analyzed.

4.1 Classifiers

To authenticate a wireless device in a network, a signature that can uniquely identify each device is needed. Various signature extraction methods are proposed and used for network security but we are interested in a signature scheme based on the unique RF signal characteristics of a device. A signature can be generated for each device by extracting its specific information from the transmitted signal via processing the received signal in the modulation domain. The extracted information will be defined as our classifiers and for the rest of this work, the terms classifiers and identifiers will be used interchangeably.

Our classifiers will be based on the deviation of the signals from the ideal signal on modulation domain. Due to manufacturing variability of the hardware and communication channel, received signals in the receiver end will be different than ideal

signal. Figure 4.1 demonstrates the variation of received signal from different boards. Signals from different radios behave differently and form a distinctively different clusters. Even the very same board acts differently with the changing power which is presented in Figure 4.2 and 4.3.

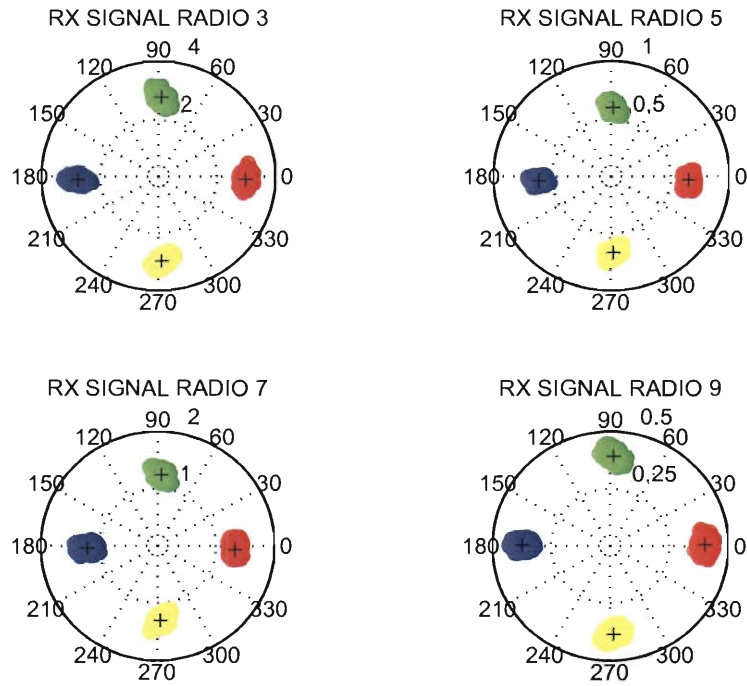


Figure 4.1 : Received signal from different boards

Thus, to extract the hidden information related to hardware variability, we will define our classifiers based on Error Vector Measurements (EVM), I/Q offset and frequency offset. EVM is widely used method for testing the quality of the communication systems [35]. The method analyzes deviations in the received signal to identify source of these distortions such as communication medium, noise and hardware imperfections of devices for troubleshooting.

EVM measurements are performed in the modulation domain. Received and ideal signal are defined as a phasor in I/Q domain. Error vector is then defined as the

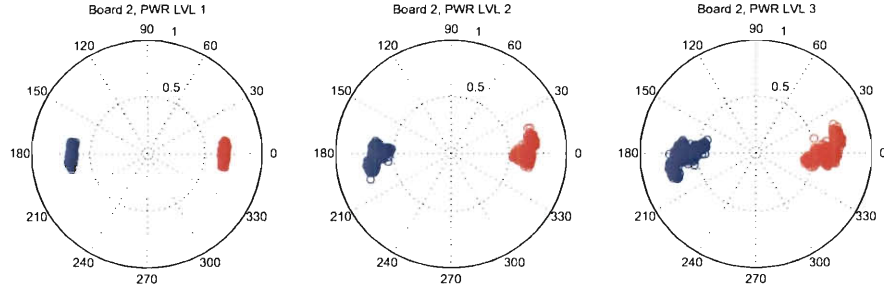


Figure 4.2 : Received signal for different power levels (BPSK)

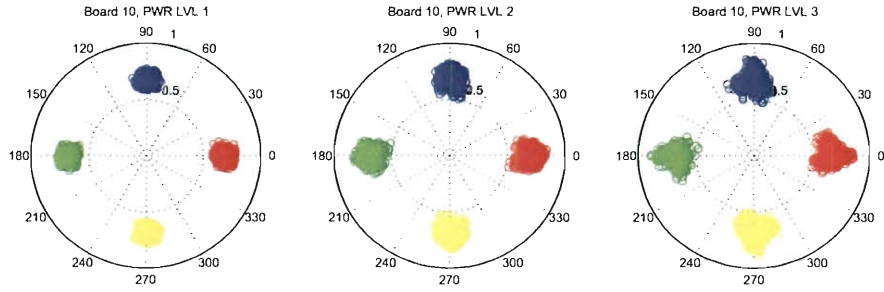


Figure 4.3 : Received signal for different power levels (QPSK)

magnitude of the distance vector between received signal and ideal signal vector which is also shown in Figure 4.4. Other metrics based on error vector can be defined as follows:

- **Magnitude Error:** Magnitude difference between received signal and ideal signal phasor.
- **Phase Error:** Angular difference between received signal and ideal signal phasor.

- **Error Vector Magnitude:** Scalar distance between received and ideal signal phasor.

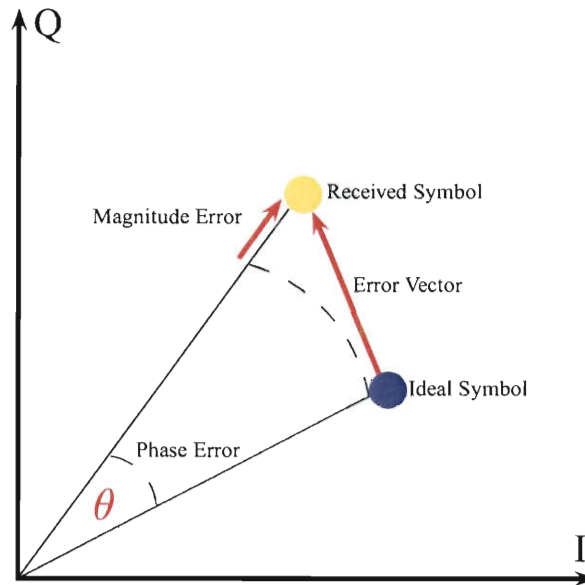


Figure 4.4 : Error vector magnitude (EVM)

These metrics will be used as our identifiers along with I/Q offset and frequency offset to generate fingerprints for WARP boards.

- **I/Q Offset:** The distance between the origin of the I/Q domain (O) and origin of the received signal (O') which is shown in Figure 4.5.
- **Frequency Offset:** Frequency difference between transmitted carrier signal and ideal carrier signal.

We choose two different modulation types in our experiments, differential BPSK and QPSK, to analyze the effects of modulation. Since EVM related classifiers are defined per symbol, two modulations will have different number of classifiers.

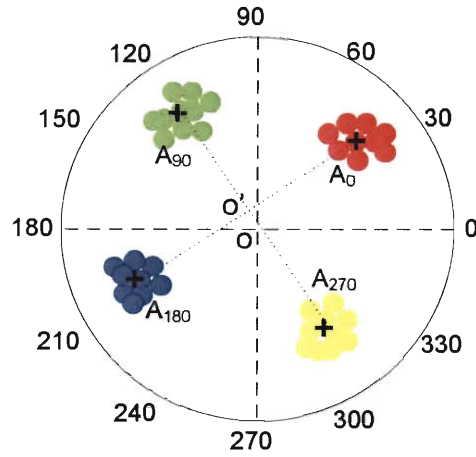


Figure 4.5 : I/Q Offset

- **BPSK**: Total of 8 classifiers: 2 Phase Error, 2 Magnitude Error, 2 Error Vector Magnitude, 1 I/Q Offset, 1 Frequency Offset
- **QPSK**: Total of 14 classifiers: 4 Phase Error, 4 Magnitude Error, 4 Error Vector Magnitude, 1 I/Q Offset, 1 Frequency Offset

4.2 Experiment Setup

Our signature scheme is based on the deviations of the transmitted signal due to hardware imperfections of transmitters. Therefore, channel effects and environment noise in the transmitted signal should be eliminated first to observe only hardware effect. For this purpose, Spirent SR5500 Wireless Channel Emulator is employed in the experiments and wireless link between transmitter and receiver is set to static channel to eliminate possible channel effects. WARPLab reference design [3], which is used for prototyping physical layer algorithms, handles the communication. While WARP boards are employed for real time communication, MATLAB is used for off-

line signal processing.

Our experiment setup is presented in Figure 4.6. Transmitter and receiver nodes are connected to each other via channel emulator's input and output connections. The Ethernet switch is used to connect WARP boards and PC. We use 12 WARP boards to simulate different CR transmitter nodes trying to communicate with a common receiver node.

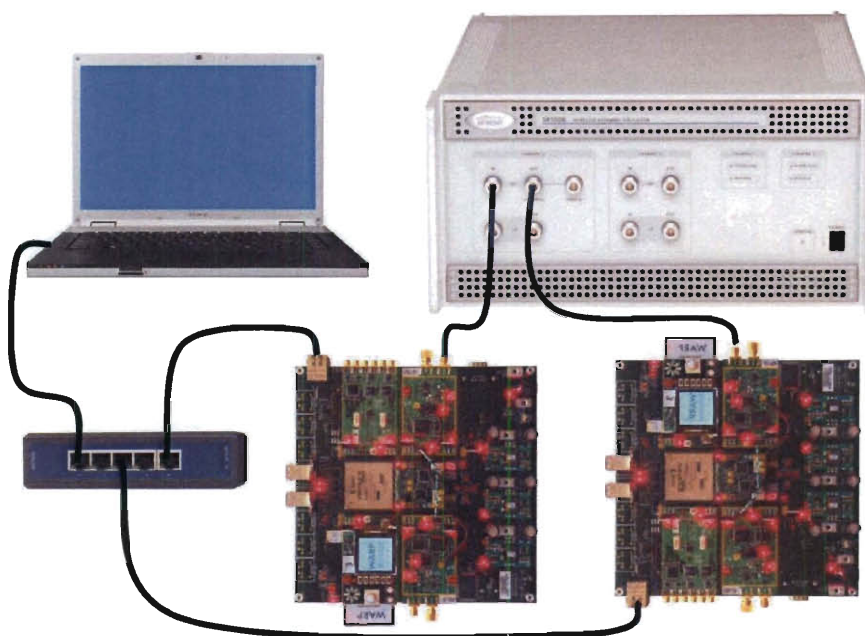


Figure 4.6 : Experiment Setup

To simulate a cognitive radio device, we choose different channel, power and modulation configurations which are listed as follows:

- **Channel:** 4 WLAN channels (1, 2 ,4 and 8)
- **Power:** 3 power settings (low, medium and high)

- **Modulation:** BPSK and QPSK

In each configuration, 200 frames each of which contains 2002 random symbols are generated and processed in MATLAB. Processed data is transferred to transmitter node via the Ethernet and then transmitted through the communication medium via WARP board. Receiver node captures the transmitted signal and sends back captured data to PC via the Ethernet for further processing in MATLAB. In addition, carrier frequencies of the signal is monitored via Agilent ESA series spectrum analyzer to compute frequency offset.

4.3 Classifier Analysis

In the following chapter we will propose a signature scheme for a cognitive radio device operating on over a range of frequencies, power levels, and modulation parameters based on aforementioned classifiers. Therefore, the relationship between parameters and classifiers should be investigated first.

Before going into further discussions, we illustrate a few samples of our visual data analysis. The significance of this phase is that no pattern recognition software has been so far able to match the human pattern recognition ability [36]. The visual trends typically provide a sound guideline on how to organize the experiments and the classifier sensitivity. Boxplots and histograms will be used for identifying visual trends.

A boxplot is a convenient way of graphically depicting groups of numerical data through their five-number summaries (the smallest observation, lower quartile, median, upper quartile, and the largest observation). Boxplot also indicates which observations, if any, might be considered as outliers. It provides a fast method for visual comparison of the density functions and outlier detection.

A histogram on the other hand, approximates the probability density function (pdf), assuming that we normalize the number of values in each bin to the total number of elements (assuming equidistance bins). Figure 4.7 illustrates estimated the probability distribution of different classifiers via histograms. In addition, gamma and normal distribution fittings of histograms are displayed in the figure. We opt to use these fitting functions instead of histograms since all the histogram data could be stored as distribution fitting parameters (i.e. mean and median for normal distribution). We choose gamma distribution to represent pdf of the classifiers since it provides a better fit than normal distribution for all the classifiers.

For the following subsections we will look at trends of each classifier with 3 sets of plots. First set will be boxplots to observe the difference between boards. Second and third sets will be histograms for different power levels and channels to analyze the corresponding effect on each board.

4.3.1 Phase Error

Figure 4.8 presents boxplot diagrams of Phase Error (PE) for different configurations. It can be seen that with the exception of board 12, all boards tend to show similar behavior which makes PE our 'weakest' classifier. The frequency response of the PE is shown in Figure 4.9 via probability distribution functions (pdf) estimated by the gamma probability distribution. For different channels, PE shows little variation. Yet for different power levels, PE acts differently. Figure 4.10 shows the difference of pdfs with different power levels. As the power increases, the pdf tends to spread out which is closely related to the behavior of the boards represented in Figure 4.2 and 4.3. In these figures, it can be seen that with the increase in the power more symbols deviate from the cluster center, which explains the spread in pdfs for different power

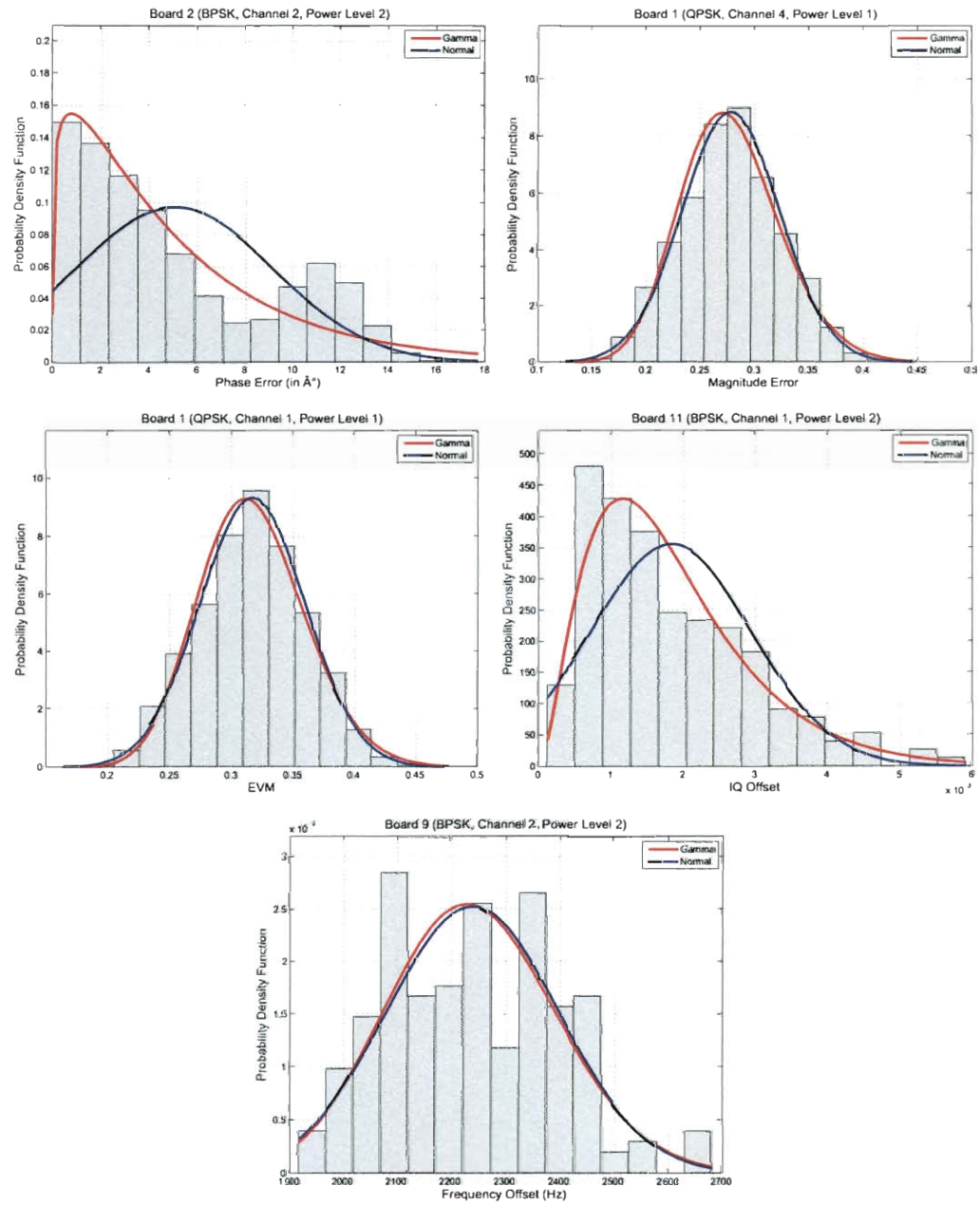


Figure 4.7 : Probability Density Function estimation with histograms

levels.

4.3.2 Magnitude Error

Magnitude Error (ME) classifier performs better than PE which can be observed from boxplots in Figure 4.11. We can see slight difference between the boards yet the patterns are similar. The frequency response of ME differs for each channel but the difference is very small as can be observed in Figure 4.12. With different power levels, ME shows similar behavior with PE which is presented in Figure 4.13. As the power increases more symbols deviate from cluster center thus magnitude error increases which causes the spread out in the pdfs.

4.3.3 Error Vector Magnitude

Error Vector Magnitude (EVM) classifier shows similar characteristics with ME. Boxplots in Figure 4.14 illustrates the same trend as in ME. Frequency response of EVM is quite stable for different channels which is presented in Figure 4.15. With the increasing power, deviation of the symbols from the cluster center increases the EVM value, so same spread out in the EVM pdfs can be observed in Figure 4.16.

4.3.4 I/Q Offset

I/Q Offset values for each board presented in Figure 4.17. Like other classifiers, I/Q Offset pdfs for different frequencies behave similarly as can be seen from Figure 4.18. I/Q Offset values increase with high transmission power which can be seen from Figure 4.19. Increase in the power causes more symbols to spread out, which in turn increases the distance between origins of the received symbols.

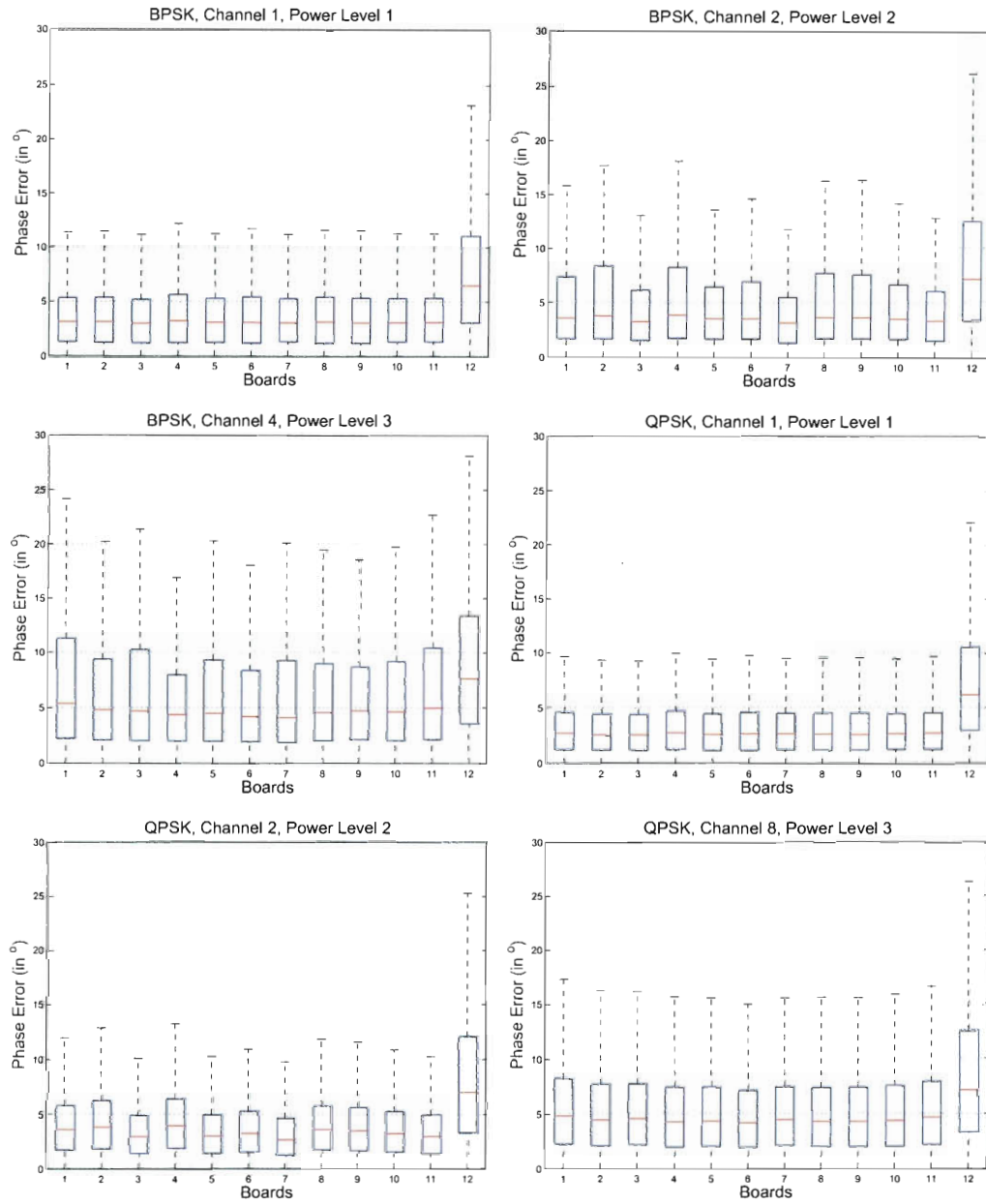


Figure 4.8 : Boxplots for Phase Error over different configurations

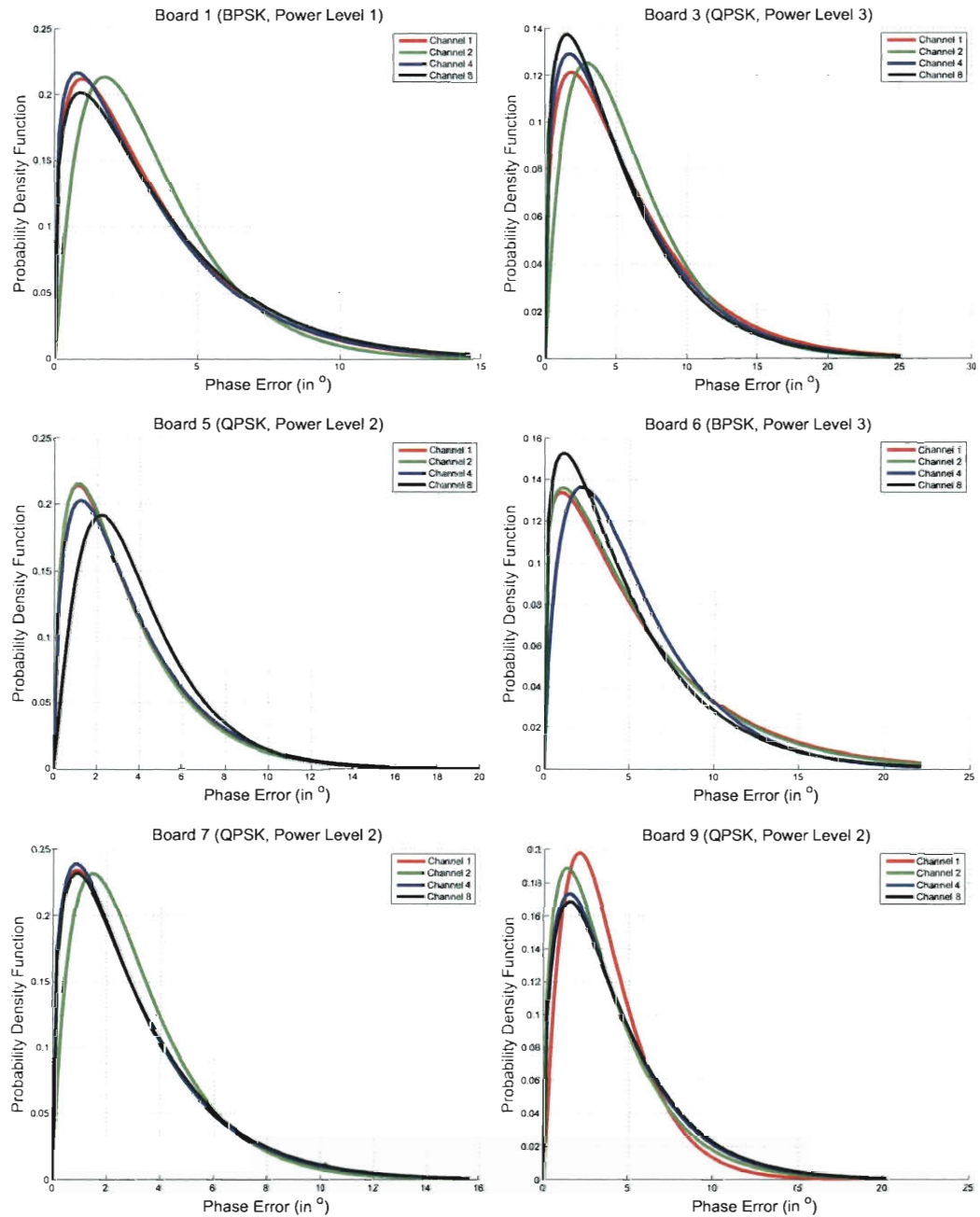


Figure 4.9 : Frequency response for Phase Error

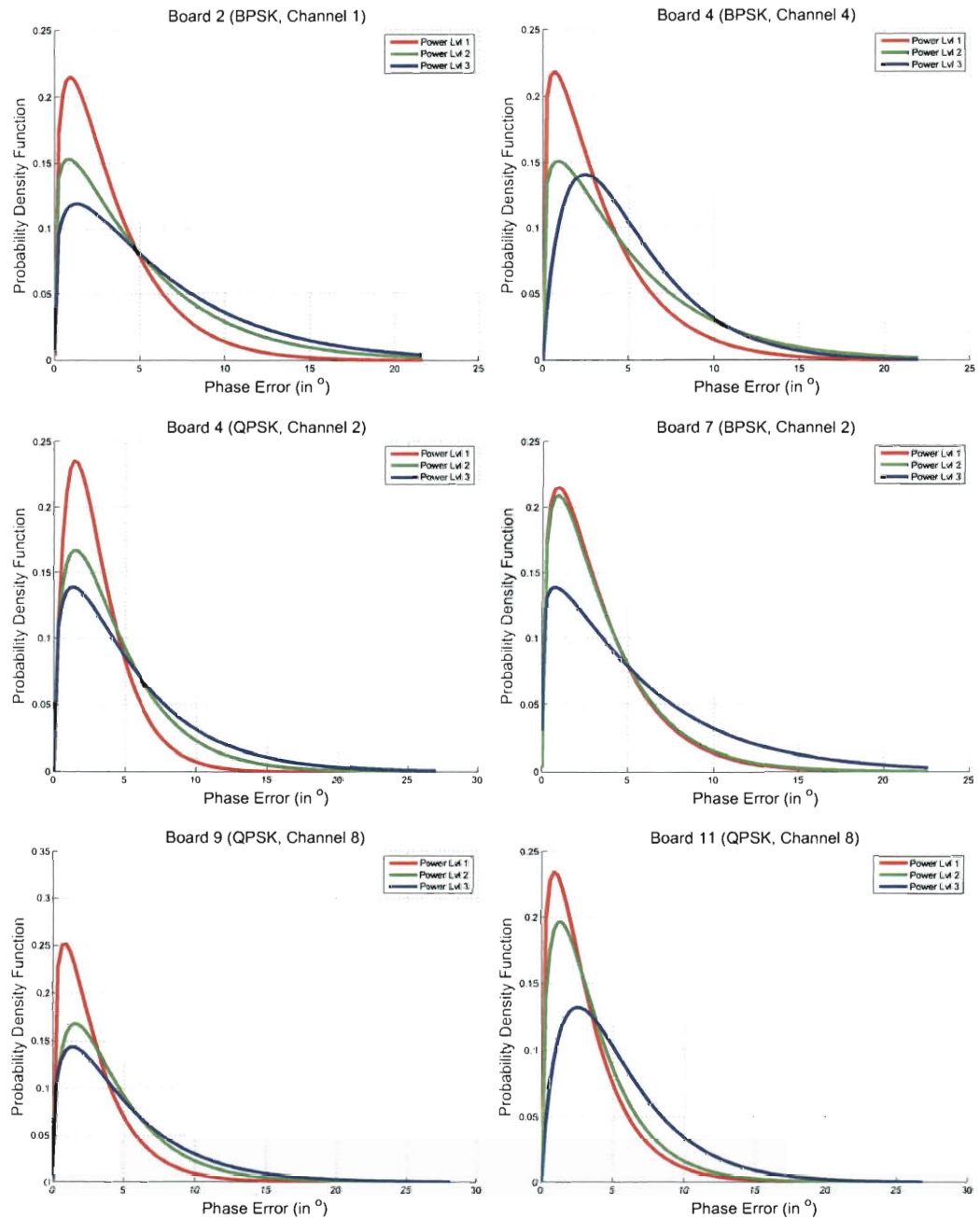


Figure 4.10 : Power response for Phase Error

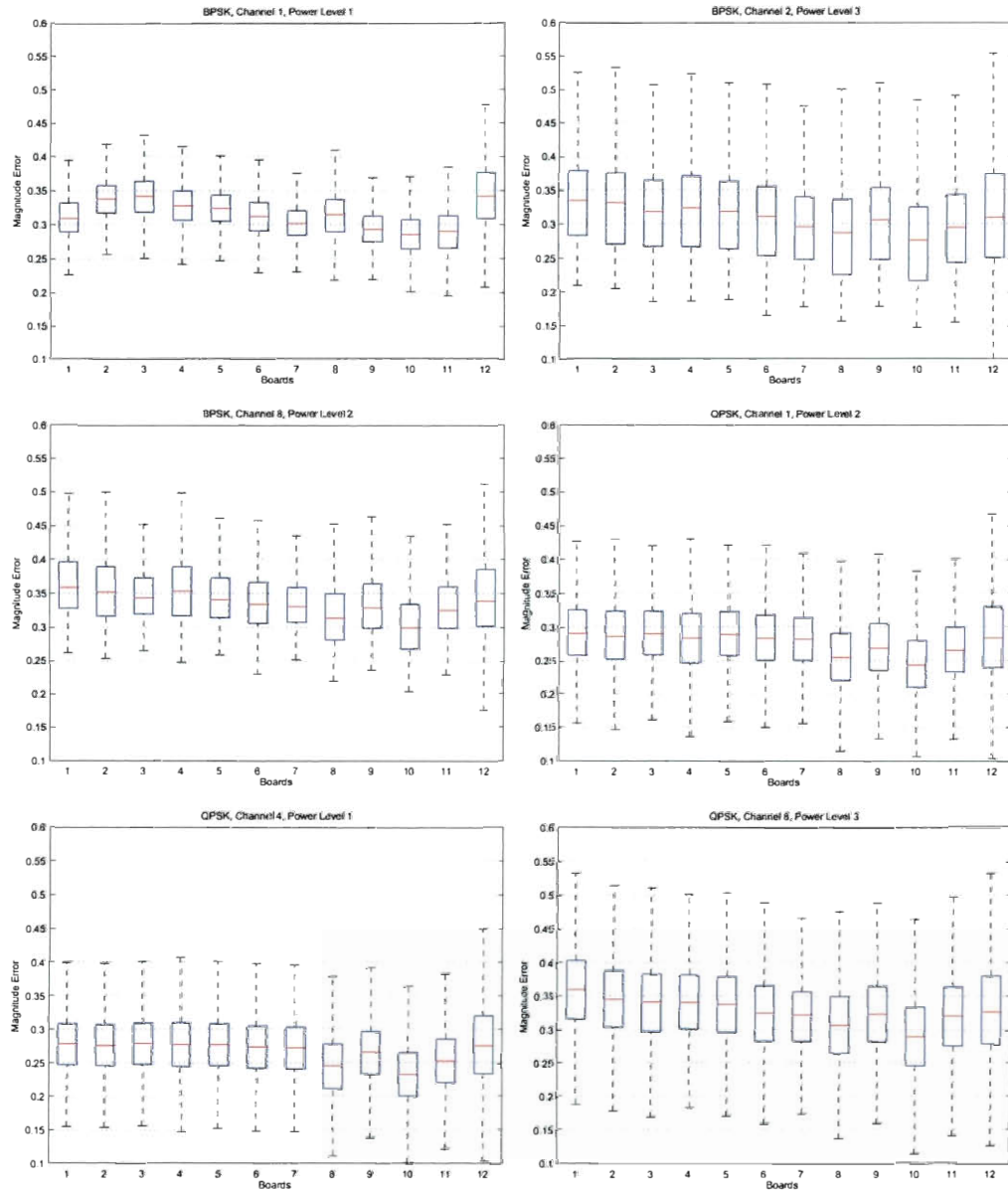


Figure 4.11 : Boxplots over the different configurations of Magnitude Error

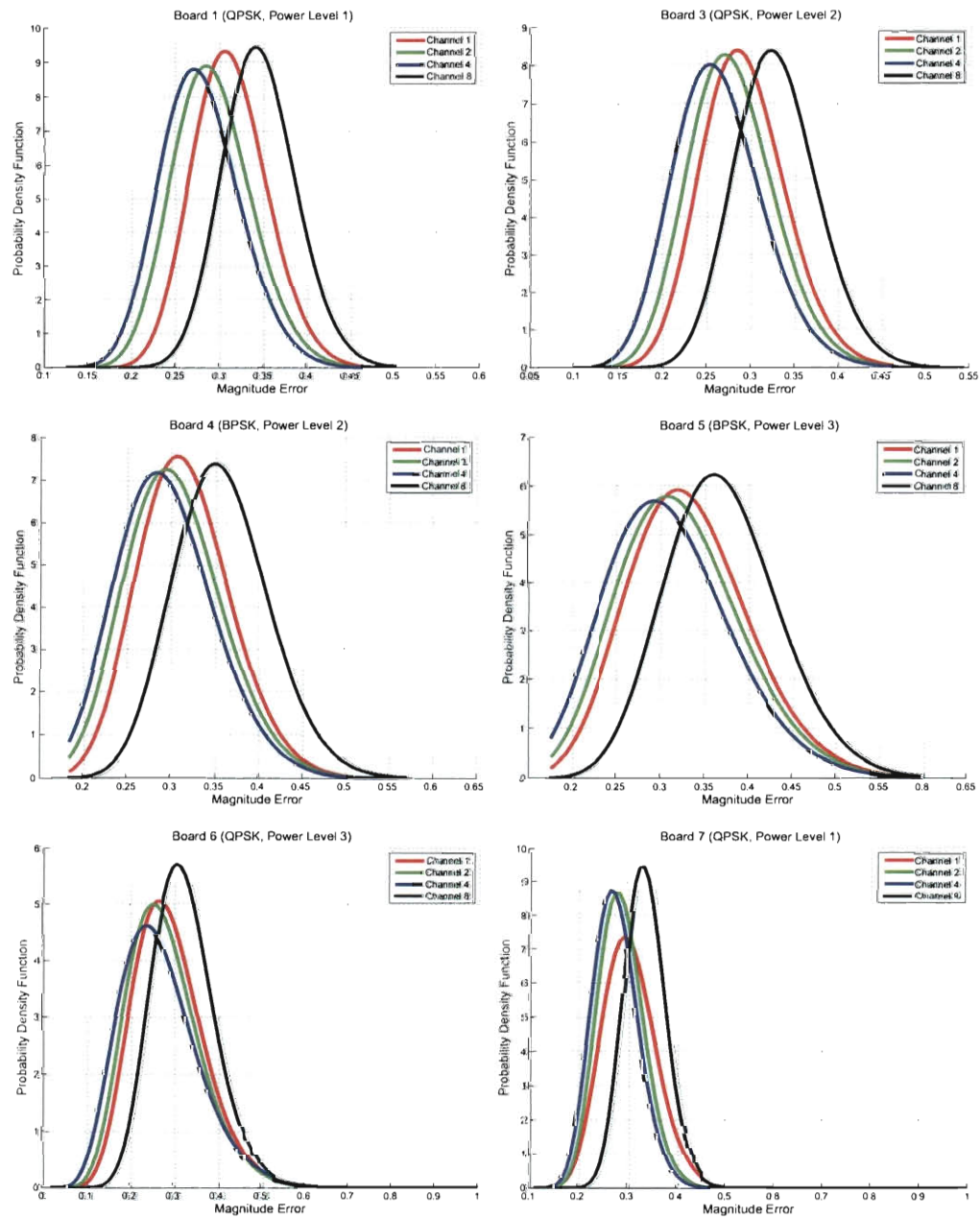


Figure 4.12 : Frequency response for Magnitude Error

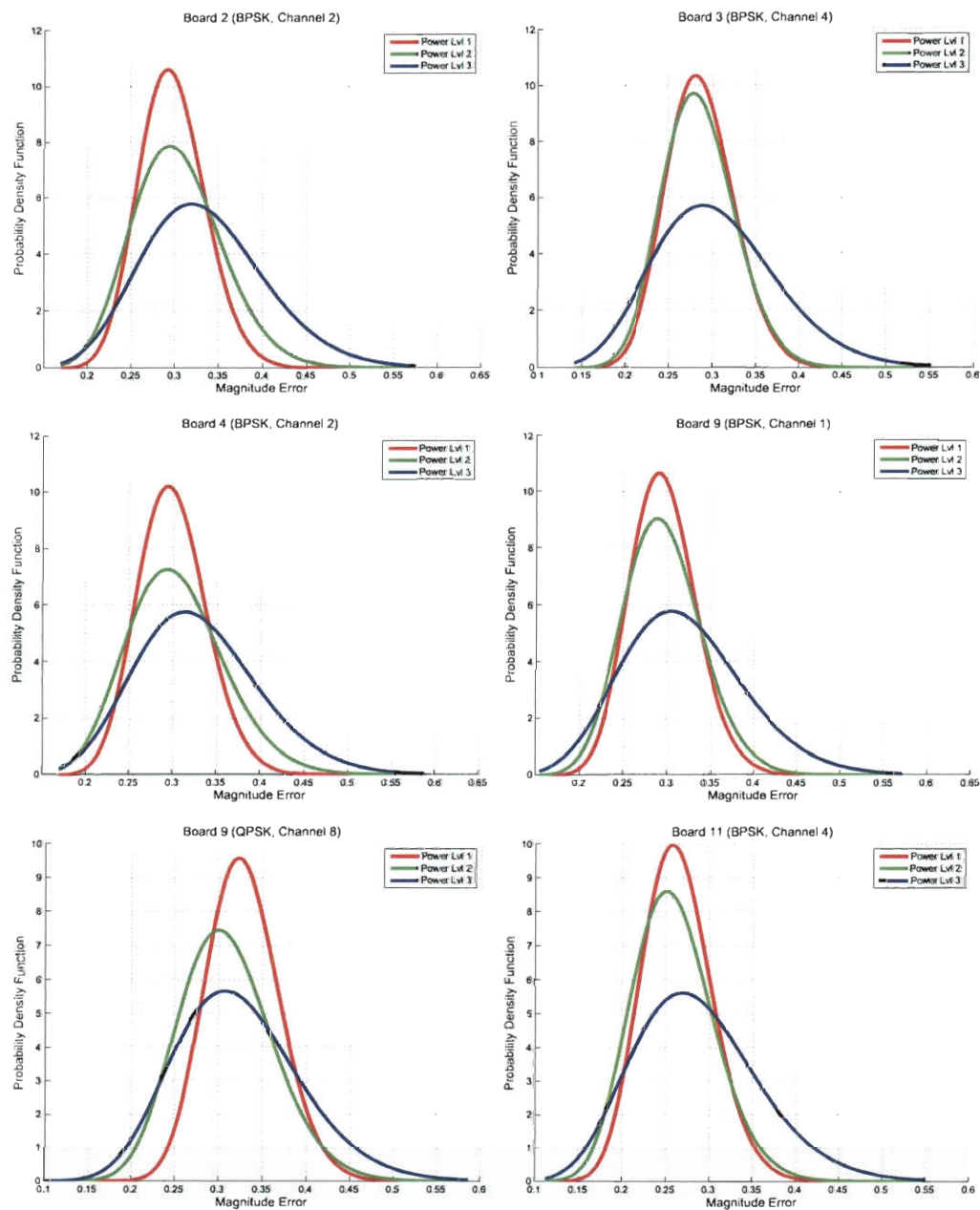


Figure 4.13 : Power response for Magnitude Error

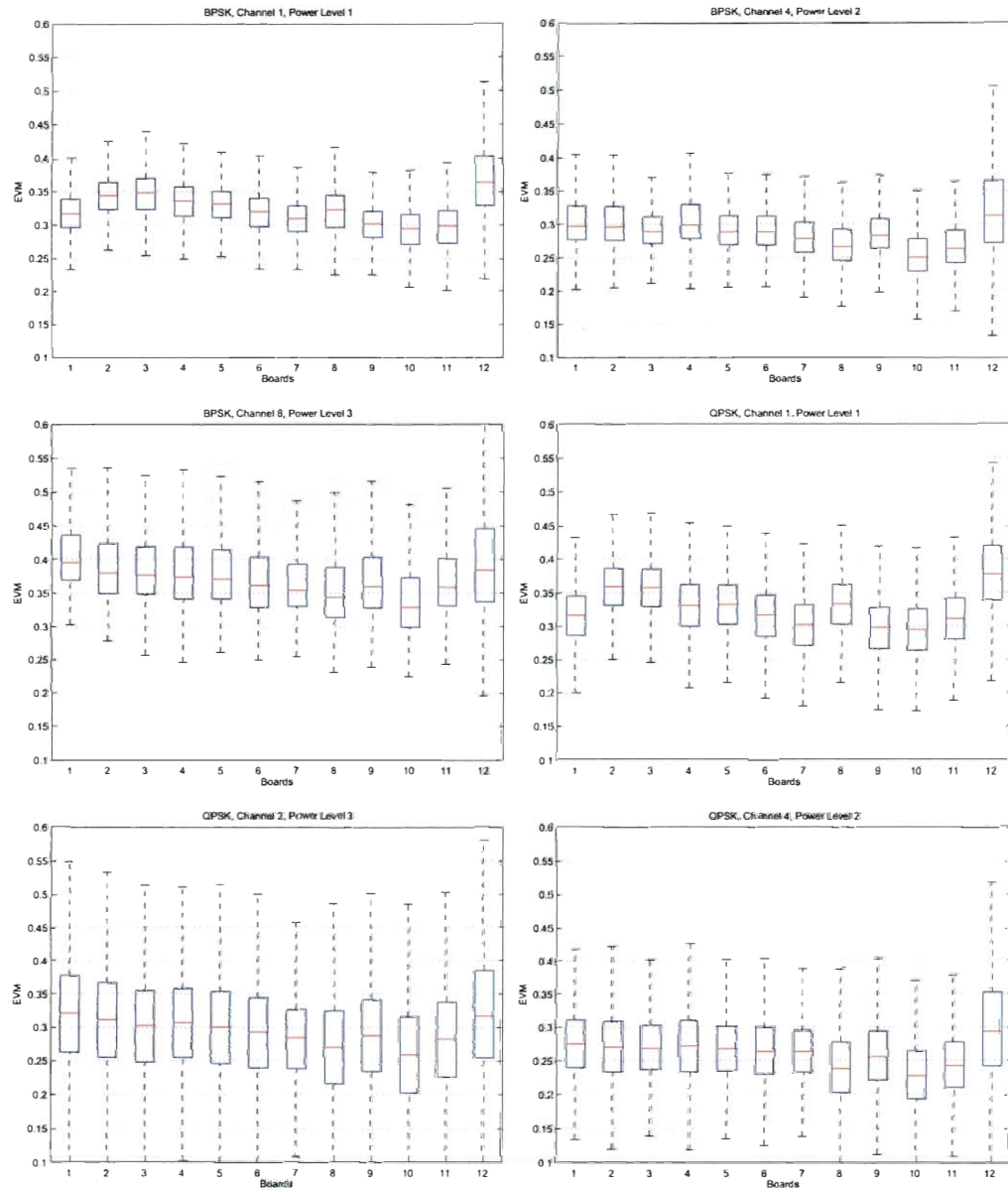


Figure 4.14 : Boxplots over the different configurations of EVM

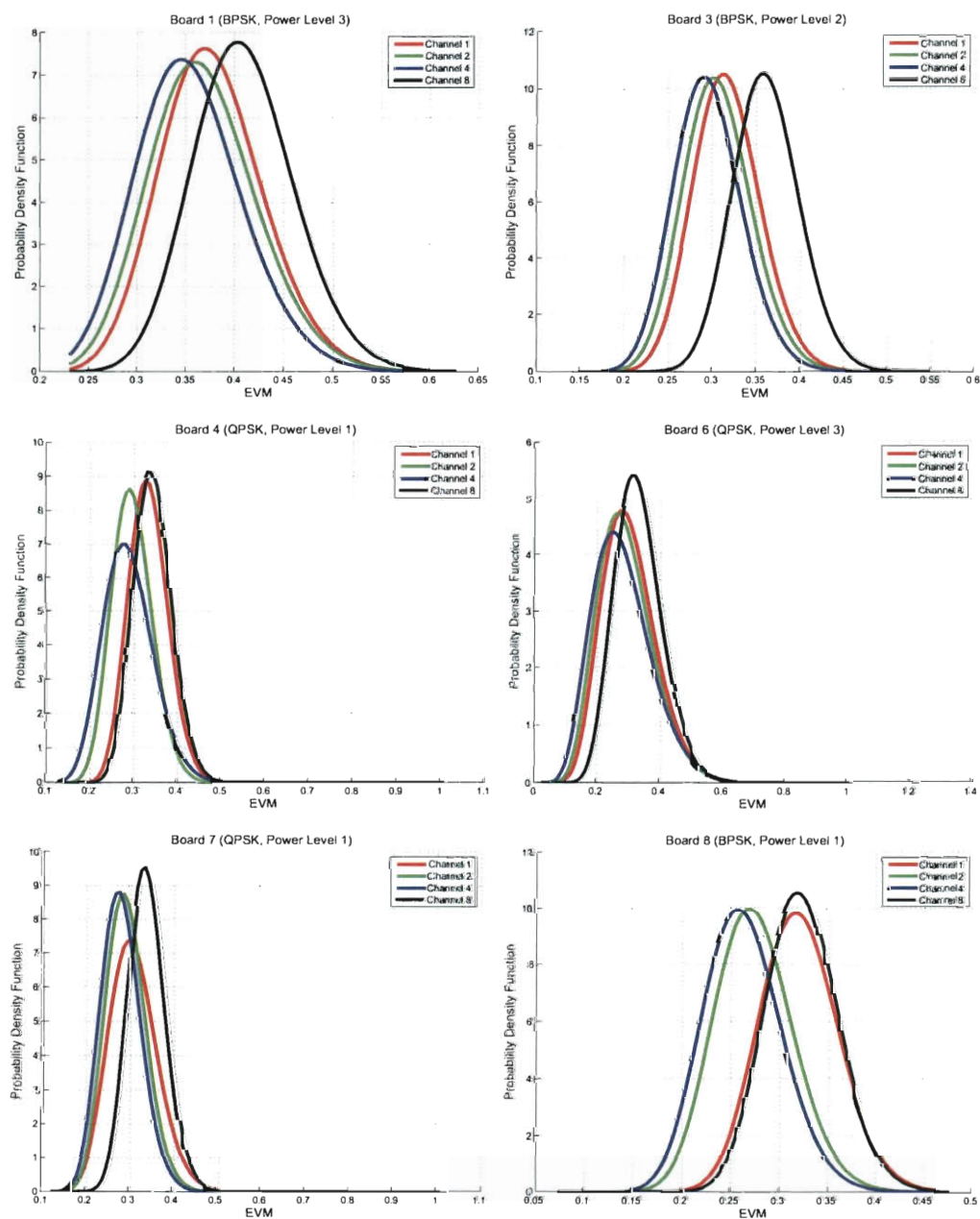


Figure 4.15 : Frequency response for EVM

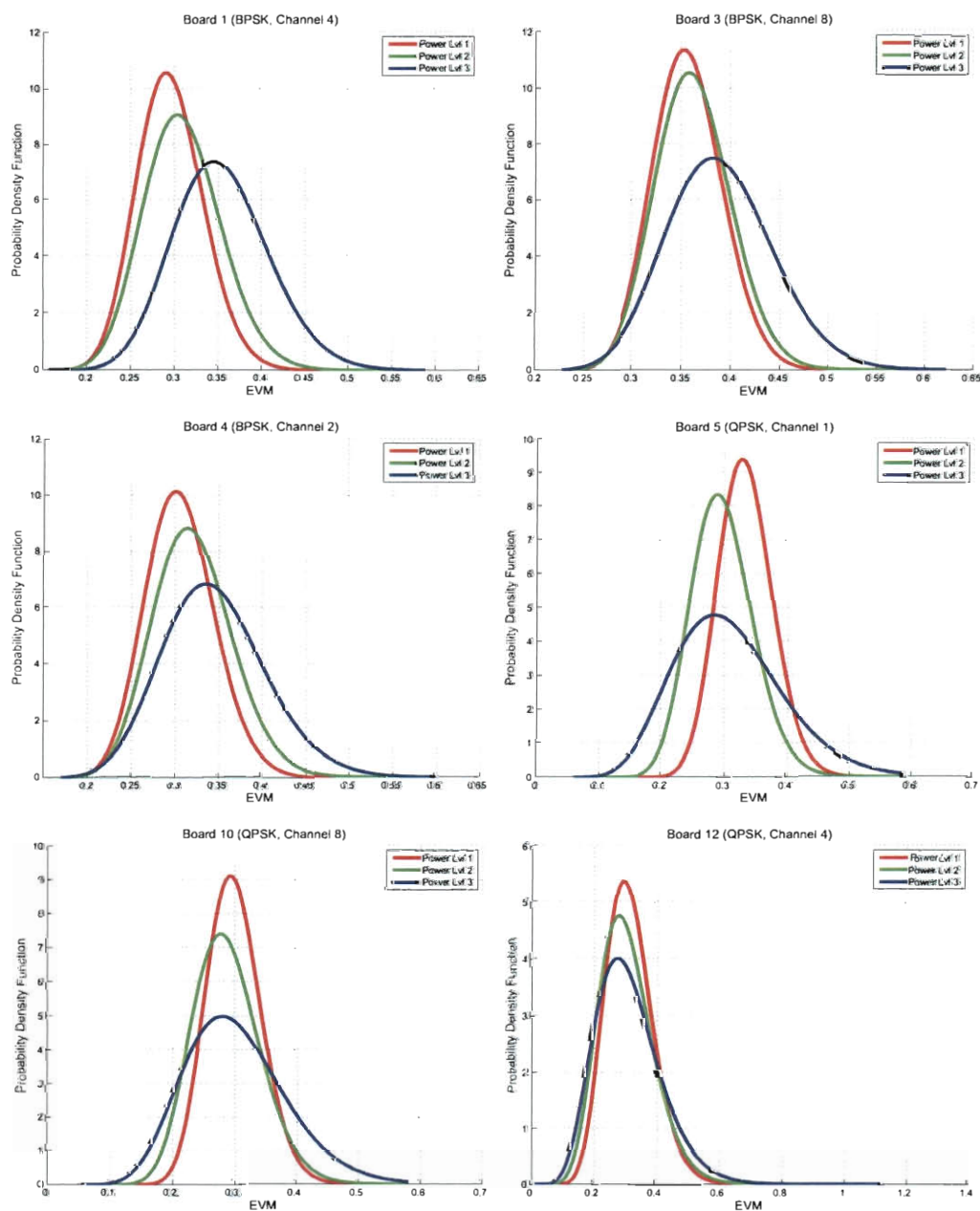


Figure 4.16 : Power response for EVM

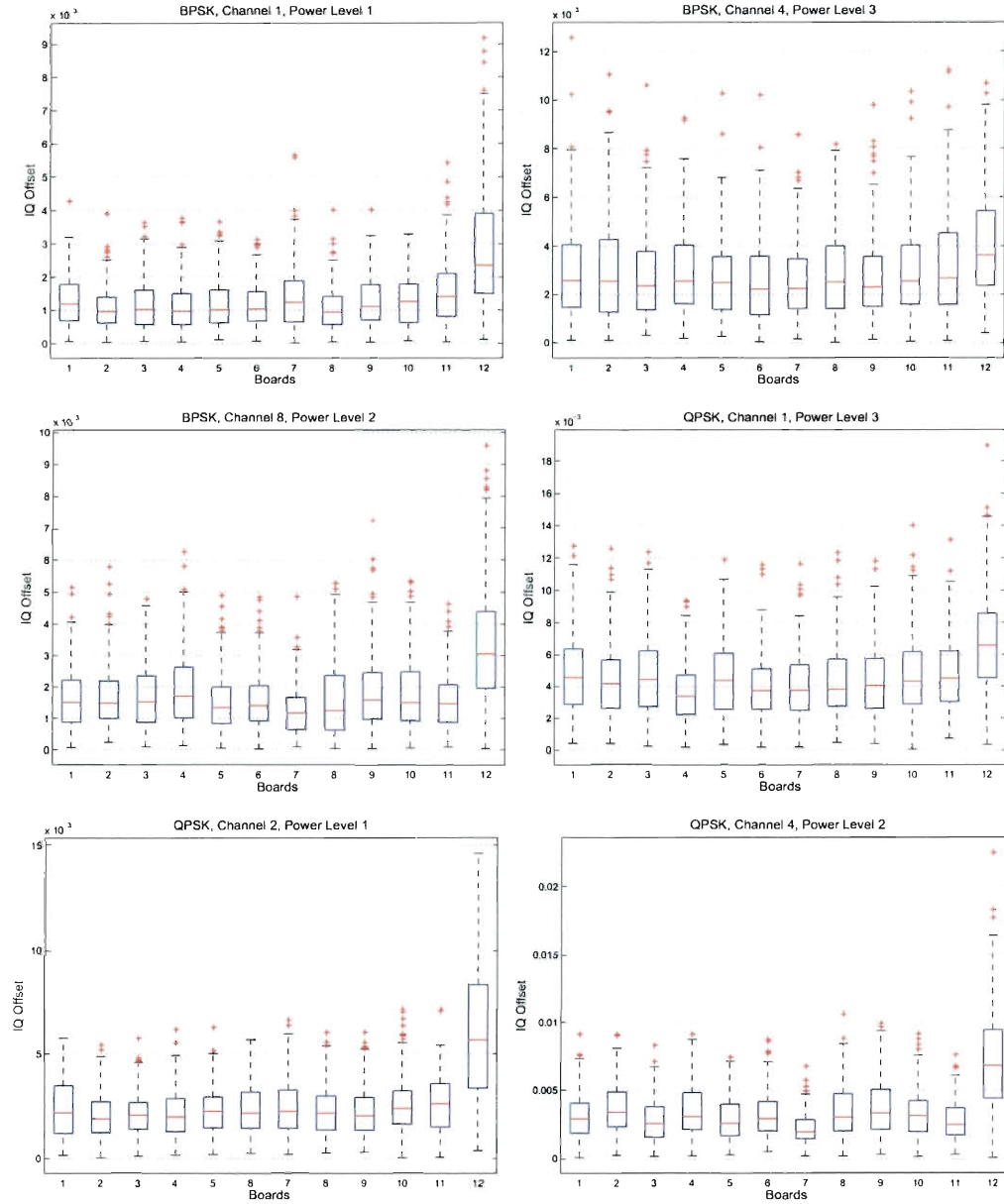
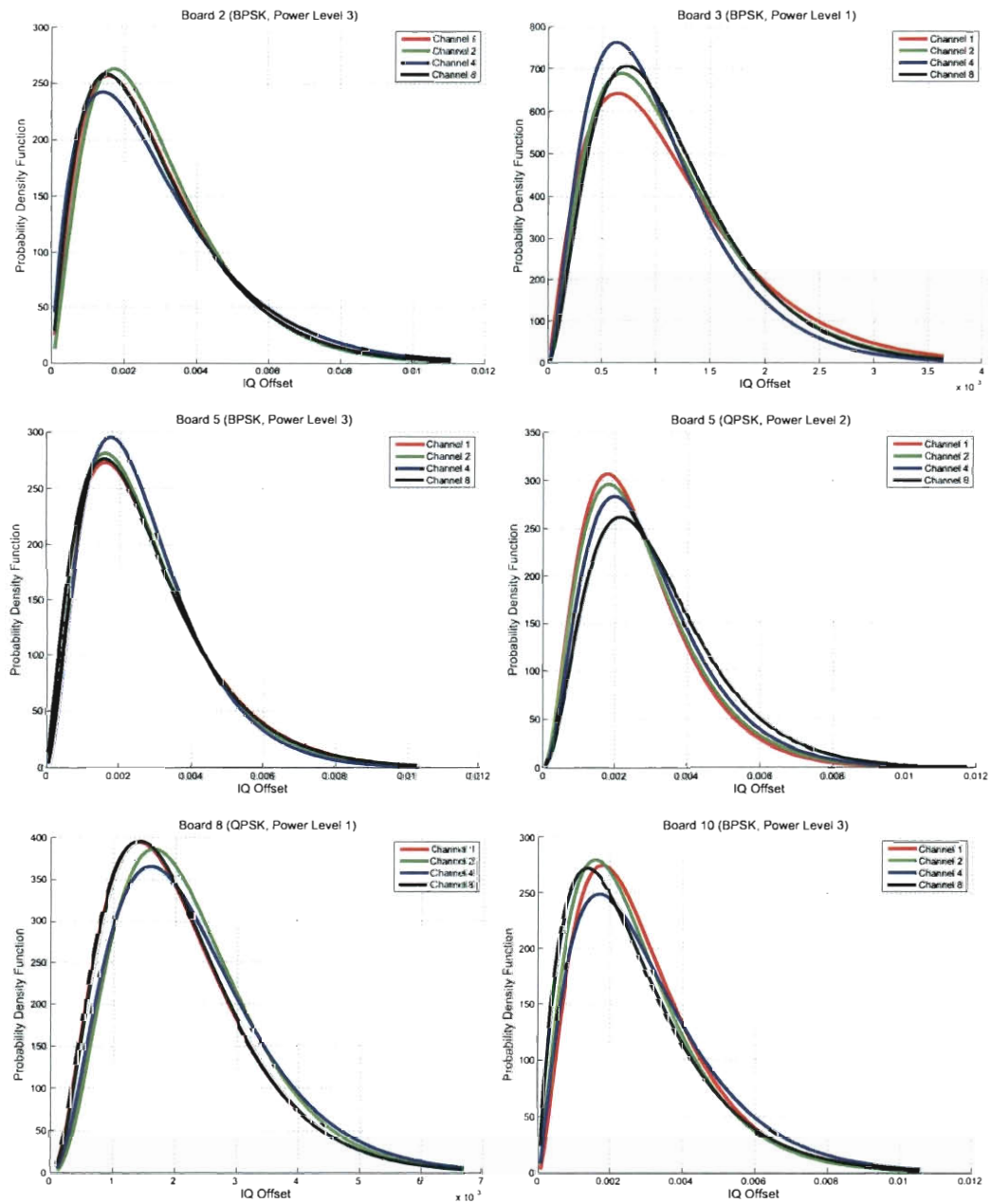


Figure 4.17 : Boxplots over the different configurations of I/Q Offset

Figure 4.18 : Frequency response for I/Q Offset

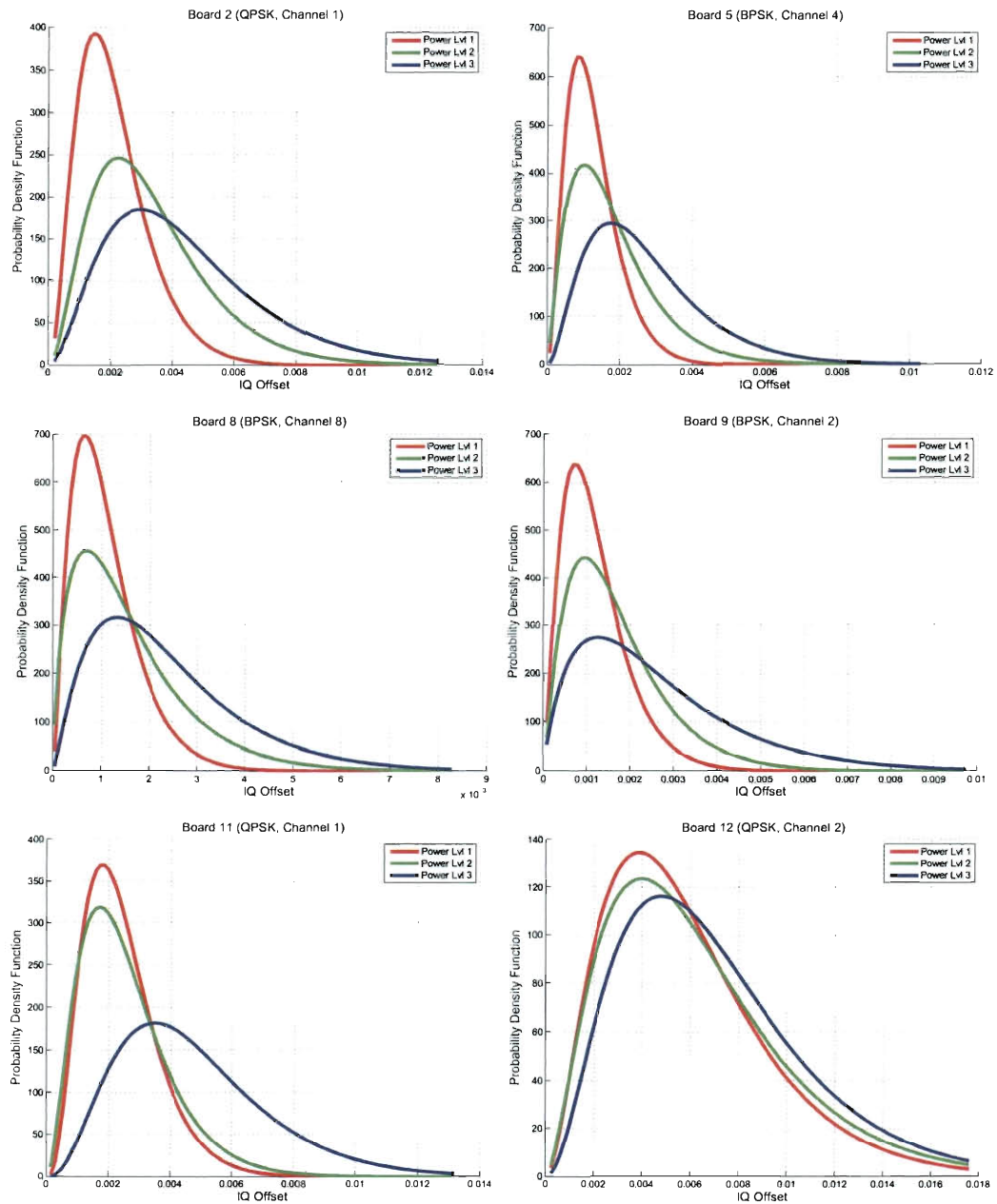


Figure 4.19 : Power response for I/Q Offset

4.3.5 Frequency Offset

Frequency offset, which can be inferred from Figure 4.20, is the best performing classifier among all the classifiers. The boxplots of the boards show variations in terms of shape and magnitude which are completely different from other classifiers. For different channels, frequency offset tends to be constant as in Figure 4.21. Unlike other classifiers, frequency offset is immune to change in power and stays almost constant as illustrated in Figure 4.22.

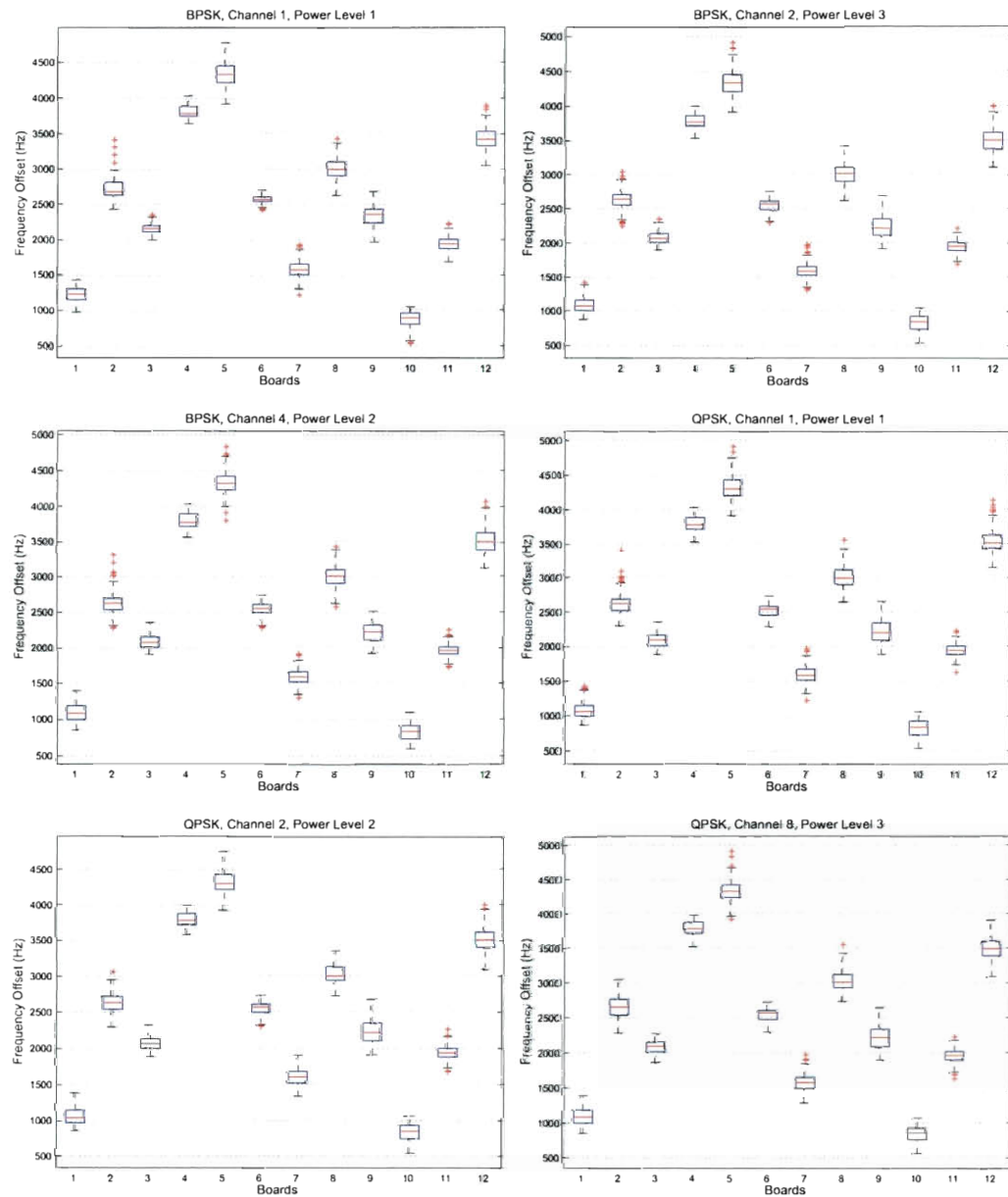


Figure 4.20 : Boxplots over the different configurations of Frequency Offset

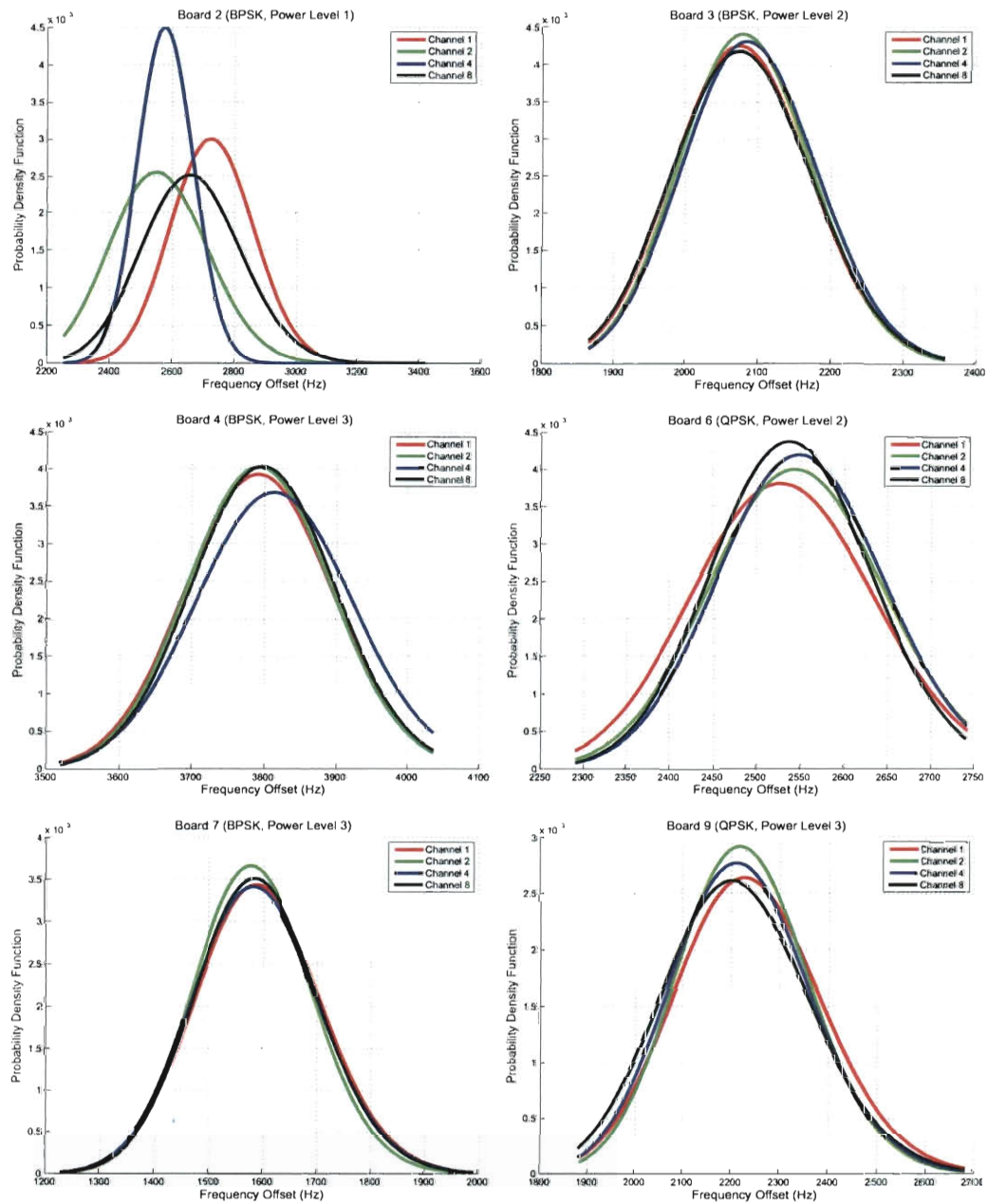


Figure 4.21 : Frequency response for Frequency Offset

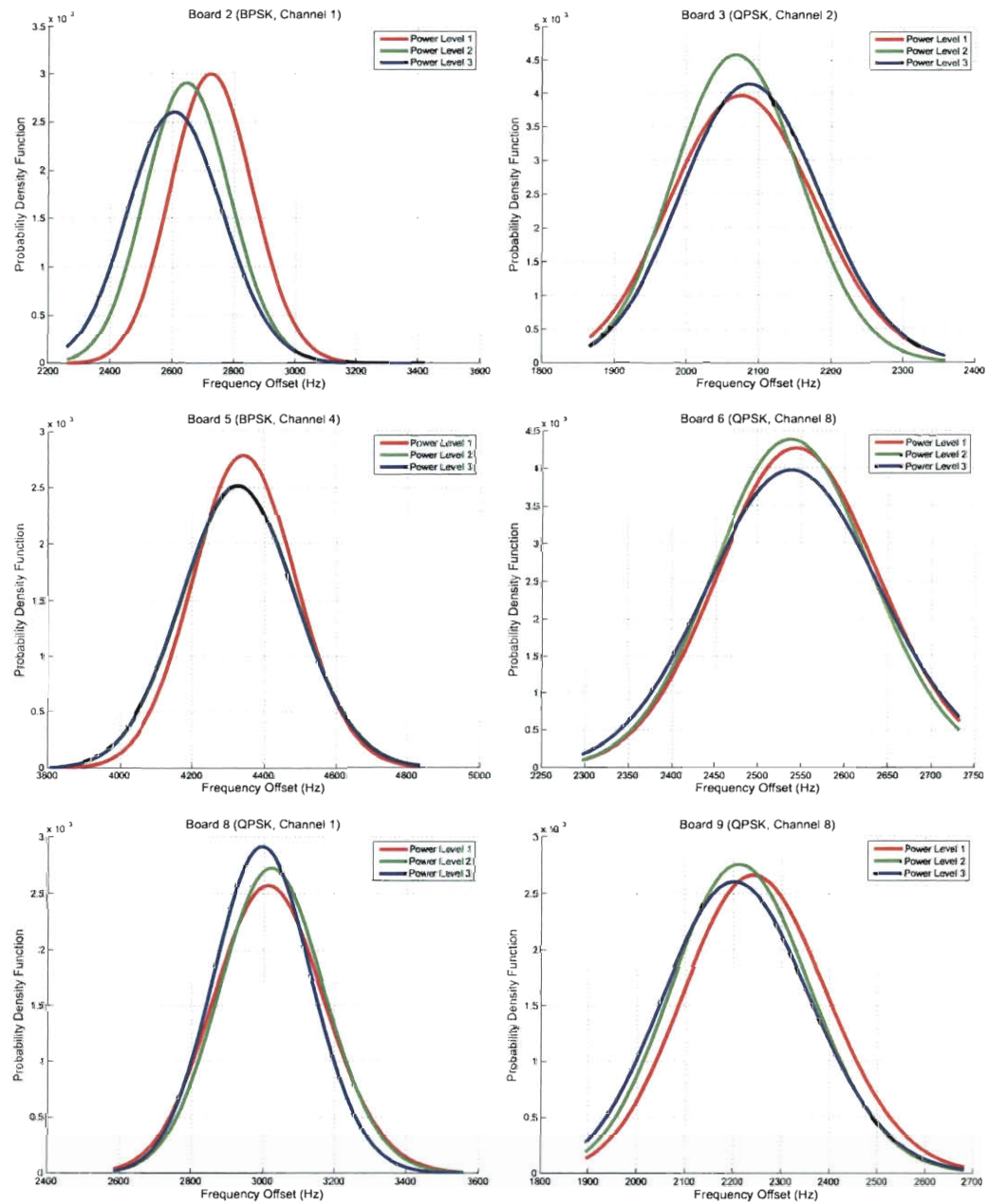


Figure 4.22 : Power response for Frequency Offset

Chapter 5

Classification

In this chapter, we will introduce our fingerprinting mechanism for cognitive radios. First, flow of the mechanism will be explained. Then a new classification method will be proposed for signature generation. Finally the performance of the classification method will be analyzed for both static channel and in-door office environment.

5.1 Fingerprinting Mechanism

In this section, we will use the classifying variables described in Chapter 4 to build a statistical model for each of the classifiers. Figure 5.1 shows the overall flow of our signature extraction (learning phase) and signature matching (testing phase) method. Signature extraction process (demonstrated on the upper row on Figure 5.1) for each radio can be defined as follows. First, a predefined number of training message frames are generated then transmitted by each radio card. Next, we define the single characteristics (G_m) that can be used to identify source of the messages. We use data-driven density formation and pdf distance metrics for representing and computing each G_m . As we mentioned earlier, each of the characteristics would be a weak classifier in the sense that we may get a high prediction error. The last step in signature extraction is to combine the results of the M classifiers. We select weighted voting as the committee formation method. Our reason for this choice is the simplicity and good performance of this method compared to other alternatives [37].

The signature matching phase (shown on the lower row on Figure 5.1) is much simpler. Upon arrival of a batch of F frames, the characteristics of the frames are evaluated against the M single classifier's extracted signatures. Next we combine the results of the M classifiers. The result would be identification of the radio source of the incoming batch of signals.

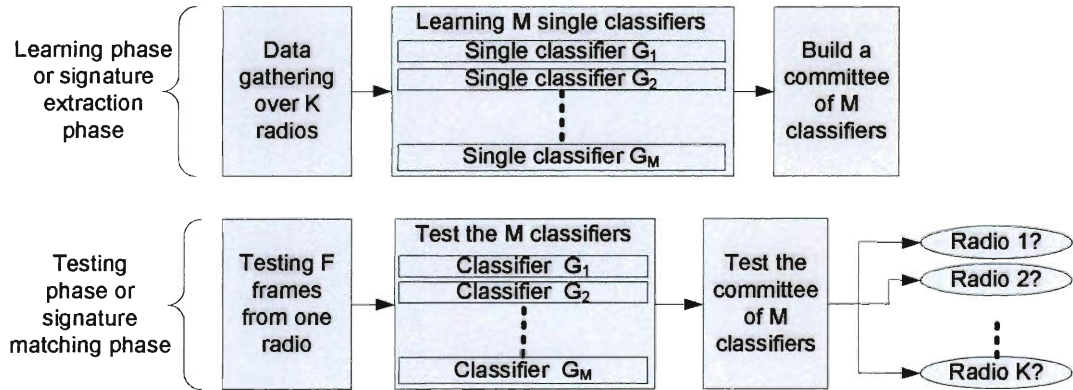


Figure 5.1 : The flow of signature extraction and signature matching approach.

5.1.1 Signature Extraction

For the signature extraction phase (learning phase) we choose random 100 frames for the training set among the 200 frames transmitted which is explained in Section 4.2 (Remaining frames will be used later for testing phase). For each classifier, first histogram data is generated, then the data is used to estimate probability density function (pdf) with gamma distribution parameters, namely scale and shape parameters. After forming the training set, we look at the weak classification using a single classifier. Randomly chosen frames are tested with the training set to find probability detection (P_D) and probability of false alarm (P_{FA}) for each classifier. The pdf of

the test frame and training set is compared to identify the source of the signal. We perform the comparison via distance/similarity measures between two pdfs, namely test frame and training set. Kullback-Leibner (KL) divergence (Equation 5.1) is chosen as our metric for measuring the distance between two pdfs. Since KL divergence is non-symmetric we will use $d_{KL} = d_{KL}(P||Q) + d_{KL}(Q||P)$ as our distance metric which is symmetric.

$$d_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (5.1)$$

P_D and P_{FA} of each classifier are then computed as follows: first we choose a random frame transmitted from one board as test frame, then distance between the test frame and training set is evaluated via KL divergence. Figure 5.2 presents a simple scenario for the distance computation for the frequency offset classifier. Board 8 is selected as the target device and a sample frame is chosen from corresponding training set. Sample frame is then compared with the training set of several boards. Only four training sets are displayed for brevity. It is trivial to identify visually the source of the signal as board 8. Also, the distance measurements from each training set which are given in Table 5.1 confirms the source of the signal as board 8, since it has the minimum distance to sample among all boards.

Based on KL distance metric we define P_D and P_{FA} as follows:

- $P_{Di} = P[D_i|H_i]$, given a test frame is transmitted from board i (H_i), training set of board i has the minimum distance (D_i) with the test frame.
- $P_{FAi} = P[D_i|H_k]$, given a test frame is transmitted from board k (H_k), training set of board i has the minimum distance (D_i) with test frame where $i \neq k$.

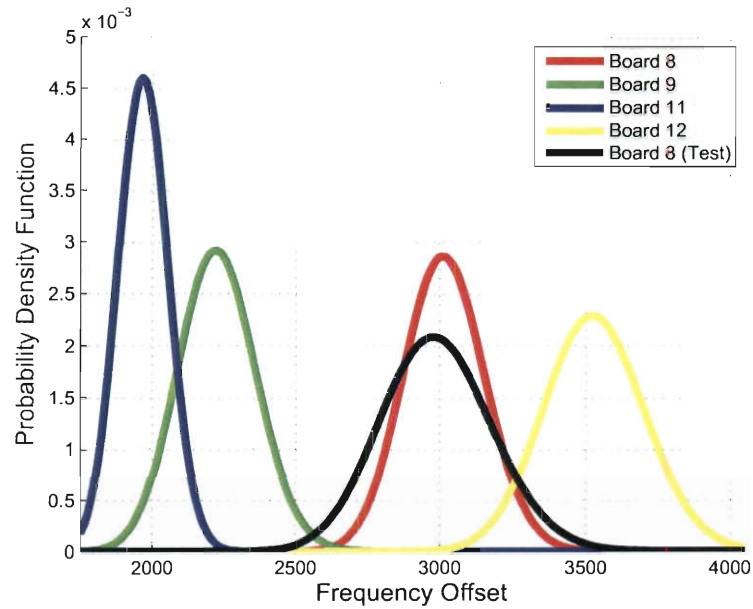


Figure 5.2 : Probability density function of frequency offset classifier

	B_8	B_9	B_{11}	B_{12}
KL dist	0.005	0.401	0.856	0.163

Table 5.1 : KL distance of boards

We evaluate the P_D and P_{FA} values of each classifier for each configuration by choosing random frames from the test set and comparing with the training set. Table 5.2 presents P_D and P_{FA} values for BPSK modulation. Each column represents BPSK classifiers which are listed from column 2 to 9 as: Phase Error over symbol 0 and 1, Magnitude Error over symbol 0 and 1, Error Vector Magnitude over symbol 0 and 1, I/Q Offset and Frequency Offset. We note that all classifiers except frequency offset have low P_D and high P_{FA} values which provides a weak means of prediction. This is an expected result since in Chapter 4 we show that boards have similar distribution

for PE, ME, EVM and I/Q Offset. We also note that values in these tables are average values for all power and frequency configurations.

	PE ₀	PE ₁	ME ₀	ME ₁	EVM ₀	EVM ₁	I/Q	FREQ
P_D	57.1	58.9	68.6	68.7	63.6	63.8	63.7	94.3
P_{FA}	42.9	41.1	31.4	31.3	36.4	36.3	36.3	5.7

Table 5.2 : P_D and P_{FA} for BPSK identifiers

P_D and P_{FA} values for QPSK modulation are also presented in Table 5.3. Table format is similar to BPSK tables but PE, ME and EVM are defined over four symbols instead of two. P_D values of frequency offset is still better than rest of the classifiers and also it has low P_{FA} values compared to the rest of classifiers. Both tables demonstrate that frequency offset is the best single classifier among all the classifiers.

	PE ₀	PE ₁	PE ₂	PE ₃	ME ₀	ME ₁	ME ₂	ME ₃	EVM ₀	EVM ₁	EVM ₂	EVM ₃	I/Q	FREQ
P_D	33.7	34.4	34.4	33.8	66.6	66.3	66.0	66.8	67.4	66.9	66.8	67.8	63.1	93.8
P_{FA}	66.3	65.6	65.6	66.2	33.4	33.7	34.0	33.3	32.6	33.1	33.2	32.3	36.9	6.3

Table 5.3 : P_D and P_{FA} for QPSK identifiers

5.1.2 Combining Classifiers

The last step of our procedure is to combine the several weak classifiers computed earlier to form one stronger committee of the classifiers. To make the committee, our first method is to perform a weighted voting. In weighted voting, we find the probability of detection for each of the classifiers G_m , $m = 1, \dots, M$. Next, we assign normalized weights α_m to each weak classifier G_m based on its probability of

detection that can be learned by using standard statistical validation methods. In such validation methods, the probability of detection can be found during the learning phase by setting aside a part of the learn data and then testing the prediction ability of the built signature from the first part of the data on the second part of the data (the set aside part) [36]. The normalization is such that the sum of the weights is 1, i.e., $\sum_{m=1}^M \alpha_m = 1$. In our evaluations, we find the non-normalized value of α_m , denoted by α'_m using the following formula for each of our weak classifiers G_m , $m = 1, \dots, M$:

$$\alpha'_m = \overline{P_D(G_m(.))} - \overline{P_{FA}(G_m(.))} \quad (5.2)$$

where $\overline{P_D(G_m(.))}$ is the average probability of detection (derived using the statistical validation methods) for the weak classifier G_m over all the radios and $\overline{P_{FA}(G_m(.))}$ is the average probability of false alarm computed like $\overline{P_D(G_m(.))}$. α_m can be easily found by normalizing the α'_m s.

We will have total of 24 set of α_m 's, one for each power, frequency and modulation configuration. Table 5.4 and 5.5 shows the effects of varying frequency and power on α values for BPSK and QPSK modulation. Rows of the tables represent the classifiers for each modulation. Columns 2 to 4 show the α values for different power levels for channel 1, while columns 4 to 7 represent the α values for different channels on high power setting. It can be observed from these tables that changing power on the same channel changes the α values dramatically. The weights of frequency offset and I/Q Offset start decreasing with increasing power level. This behavior is related the observations made in Chapter 4 where we show the signal characteristics of boards in Figure 4.2 and 4.3. Signals transmitted with low power form more uniform clusters which decreases the contribution of EVM related classifiers, yet with the increasing power levels, clusters change its shape and form rather unique formation. Thus, increasing power results in different EVM results which in turn decrease the weights

of frequency offset and I/Q Offset. This behavior is discussed also in Section 4.3, where each board acts different for power levels shown from their pdfs. Changing the channel for same power setting on the other hand, does not result in critical change on α values which confirms the observations made in Section 4.3 where pdf's of classifiers tend to stay same with changing frequency for the same power level.

To form the committee, we also map the classification results from each weak classifier to a value in the set $\{-1,1\}$. If the weak classifier G_m identifies the radio R_1 as the transmitter, then $G_m(R_1) = 1$, otherwise, $G_m(R_1) = -1$. Let G_{vote} denote the final voting function for a radio. The following voting function is used for weighting the votes of the different classifiers for one radio R_k :

$$G_{vote}(R_k) = \sum_{m=1}^M \alpha_m G_m(R_k). \quad (5.3)$$

The radio with the highest G_{vote} would be selected to be the target radio. In our experimental results, we compare the performance of our classifier against combining the results of the M classifiers by summing up their KL distances (Equation 5.4). This time, the radio R_k with the minimum KL distance (G_{MD}) would be the target radio.

$$G_{MD}(R_k) = \sum_{m=1}^M \text{KL dist}(G_m(R_k)) \quad (5.4)$$

We present a simple case for voting and MD classifier for BPSK modulation in Table 5.6. Columns of the table represents weak classifiers identification for corresponding board's test frame. Weak classifiers are listed between row 2 and 9, the final two rows correspond to decision of voting and MD classifier. For boards 1, 3, 5 and 6 both classifiers identify boards correctly as expected since majority of the weak classifiers detect the boards correctly. The results for boards 2, 4 and 7 are interesting

	C_{1,P_L}	C_{1,P_M}	C_{1,P_H}	C_{2,P_H}	C_{4,P_H}	C_{8,P_H}
PE ₀	0	0.083	0.105	0.090	0.088	0.027
PE ₁	0	0.11	0.100	0.105	0.073	0.098
ME ₀	0.077	0.141	0.160	0.138	0.177	0.192
ME ₁	0.083	0.142	0.154	0.160	0.185	0.199
EVM ₀	0.062	0.031	0.119	0.143	0.103	0.105
EVM ₁	0.071	0.030	0.118	0.130	0.106	0.123
I/Q	0.172	0.136	0.060	0.015	0.051	0.007
FREQ	0.536	0.327	0.183	0.220	0.217	0.249

Table 5.4 : α values for different configurations (BPSK)

	C_{1,P_L}	C_{1,P_M}	C_{1,P_H}	C_{2,P_H}	C_{4,P_H}	C_{8,P_H}
PE ₀	0	0	0	0	0	0
PE ₁	0	0	0	0.003	0	0
PE ₂	0	0	0	0	0	0
PE ₃	0	0	0	0	0	0
ME ₀	0.022	0.064	0.112	0.106	0.109	0.114
ME ₁	0.034	0.066	0.101	0.112	0.108	0.104
ME ₂	0.051	0.054	0.107	0.108	0.112	0.106
ME ₃	0.032	0.065	0.110	0.102	0.111	0.107
EVM ₀	0.013	0.096	0.110	0.097	0.096	0.102
EVM ₁	0.034	0.091	0.090	0.100	0.101	0.096
EVM ₂	0.013	0.079	0.098	0.101	0.103	0.098
EVM ₃	0.019	0.092	0.101	0.098	0.103	0.098
I/Q	0.096	0.111	0.034	0.054	0.025	0.036
FREQ	0.686	0.282	0.136	0.119	0.133	0.140

Table 5.5 : α values for different configurations (QPSK)

to demonstrate the importance of voting-based classification. Majority of the weak classifiers produce false alarm for these boards. Yet, frequency offset and I/Q offset classifier identifies the boards correctly as they have a greater weight than rest of the classifiers. For this reason, voting based classifier could identify the boards correctly while MD based classifier fails to detect actual board.

	B ₁	B ₂	B ₃	B ₄	B ₅	B ₆	B ₇
PE ₀	1	2	3	6	7	4	7
PE ₁	4	4	3	6	2	4	7
ME ₀	1	1	5	12	4	6	12
ME ₁	1	1	5	12	5	6	12
EVM ₀	1	1	5	12	5	9	12
EVM ₁	1	1	2	12	5	6	12
I/Q	1	2	3	4	9	5	7
FREQ	1	2	3	4	5	6	7
G_{vote}	1	2	3	4	5	6	7
G_{MD}	1	1	3	12	5	6	12

Table 5.6 : Voting Example for BPSK

5.2 Performance of Classifiers

We evaluate the performance of the classifiers by choosing 5 random frames from each board's test set and looking at the classifiers response. This procedure is repeated for 100 times to compute P_D and P_{FA} values for each configuration. Table 5.7 and 5.8 present average P_D and P_{FA} values of two classifiers for all configurations with BPSK and QPSK modulation.

The results are really promising and clearly show the superior performance of the voting-based classifier. We have an average P_D of 97.7% and 96.8% for BPSK and QPSK respectively, while average P_{FA} is very low, only 2.3% and 3.1% for BPSK and QPSK. The results of the voting-based method clearly benefit from the frequency offset classifier which has much better prediction rate compared to other classifiers. As a result, this classifier gets a higher value for its weight α_m , providing accurate prediction.

MD classifier on the other hand has 75.9% P_D and 24.1% P_{FA} for BPSK and 76.1% P_D and 23.9% P_{FA} for QPSK. These results confirm the effectiveness of weighting for generating stronger classifier.

	B ₁	B ₂	B ₃	B ₄	B ₅	B ₆	B ₇	B ₈	B ₉	B ₁₀	B ₁₁	B ₁₂
$G_{vote}(P_D)$	100	90.3	95.1	94.8	99.2	97.3	97.2	100	99.8	99.9	98.5	100
$G_{MD}(P_D)$	79.5	65	76.5	66.5	74.5	73	90	89.5	76.5	62.5	57.5	100
$G_{vote}(P_{FA})$	0.1	4.5	1.3	5.2	6.9	6.7	1.1	0	1.3	0	0.8	0
$G_{MD}(P_{FA})$	25.1	23.7	15.8	16.3	43.7	21.1	19.5	17.6	27.7	39.8	29.8	12.5

Table 5.7 : Combining classifiers: Voting and ML (BPSK)

	B ₁	B ₂	B ₃	B ₄	B ₅	B ₆	B ₇	B ₈	B ₉	B ₁₀	B ₁₁	B ₁₂
$G_{vote}(P_D)$	99.3	92.5	96.7	96.8	87.1	96.2	99.3	100	96.3	100	98.5	100
$G_{MD}(P_D)$	77.5	54	76	87.5	86	84.5	57	56	94.5	72.5	67.5	100
$G_{vote}(P_{FA})$	0.8	4.8	10.6	3.8	1.8	7.8	2.7	0.5	2.5	0	2.2	0
$G_{MD}(P_{FA})$	19.3	16.3	26.7	16.4	20.5	14.1	29.5	21.7	10.2	21.8	31.4	6.4

Table 5.8 : Combining classifiers: Voting and ML (QPSK)

5.2.1 Results with Office Environment

To analyze the impact of the channel on our classification method, we perform same set of experiments with indoor office environment setting of the Channel Emulator. Unlike a static channel, indoor office environment will have big impact on transmitted signals. Fading and multi-path effects will be observed in the received signal which will cause further deviation from the ideal signal. Table 5.9 and 5.10 show the P_D and P_{FA} values of the classifiers in office environment for BPSK and QPSK modulation. As expected, EVM related classifiers, already weak in the static channel, are severely affected by the office environment. We observe low P_D and high P_{FA} rates compared to static channel for these classifiers. While I/Q Offset and frequency offset is resistant to channel impact and provide similar detection probability with static channel.

	PE ₀	PE ₁	ME ₀	ME ₁	EVM ₀	EVM ₁	I/Q	FO
P_D	16.6	16.5	8.5	8.5	9.6	9.8	54.7	94.9
P_{FA}	83.4	83.5	91.5	91.5	90.4	90.2	45.3	5.1

Table 5.9 : P_D and P_{FA} for BPSK identifiers

	PE ₀	PE ₁	PE ₂	PE ₃	ME ₀	ME ₁	ME ₂	ME ₃	EVM ₀	EVM ₁	EVM ₂	EVM ₃	I/Q	FREQ
P_D	18.8	19	19	18.9	9.2	9.3	9.2	9.2	11.9	12.1	11.8	12	49.1	94.5
P_{FA}	81.3	81	81	81.1	90.8	90.8	90.8	90.8	88.1	87.9	88.2	88.0	50.9	5.5

Table 5.10 : P_D and P_{FA} for QPSK identifiers

The effect of the channel is manifested clearly on the EVM based classifiers which results low P_D and high P_{FA} for these classifiers. The weights of these classifiers (α

values) will be lower than static channel weights, while weights of the I/Q Offset and frequency offset will increase compared to their static channel counterpart.

Even with the indoor office environment weighted voting mechanism will be immune to channel impacts. Table 5.7 and 5.12 present average P_D and P_{FA} values for the two classifiers. We have 96.6% and 95.9% detection rate for BPSK and QPSK respectively. Also P_{FA} values are still low, only 3.4% and 4.1%. The identification rate slightly decreased with the channel presence, yet still we have a high P_D values due to frequency offset classifier. The performance of the MD classification degrades with the channel presence. P_D has dropped to 58.1% and 51.1% while P_{FA} increased to 41.9% and 48.9% for BPSK and QPSK respectively, which further encourages us to use weighted-voting mechanism.

	B ₁	B ₂	B ₃	B ₄	B ₅	B ₆	B ₇	B ₈	B ₉	B ₁₀	B ₁₁	B ₁₂
$G_{vote}(P_D)$	93.1	100	95.8	96.9	96	90.9	100	98	95.8	96.5	97.5	99
$G_{MD}(P_D)$	59.7	59.8	62.2	45.9	65.5	53.5	56.9	48.8	40.1	41.9	62.3	100
$G_{vote}(P_{FA})$	2.5	8.7	1.7	1.3	8.5	7.1	2.5	3.3	0.2	2.3	1.7	0.7
$G_{MD}(P_{FA})$	52.6	43.3	37.3	44.1	45.6	46.8	38.2	40.4	45.2	39.4	40.3	30.3

Table 5.11 : Combining classifiers: Voting and ML (BPSK)

	B ₁	B ₂	B ₃	B ₄	B ₅	B ₆	B ₇	B ₈	B ₉	B ₁₀	B ₁₁	B ₁₂
$G_{vote}(P_D)$	93.3	94.1	98.5	96.4	94.8	95.2	97.4	94.2	97.2	98.3	94.7	97.6
$G_{MD}(P_D)$	45.3	39.4	37.9	42.7	50.7	50.1	52.2	47.3	38.4	45.5	64.3	99.7
$G_{vote}(P_{FA})$	3.9	4.3	4.1	5.3	4.4	3.7	4.1	2.9	1.7	5.5	4.2	5.1
$G_{MD}(P_{FA})$	54.2	57.2	44.6	59.1	51.1	45.5	50.8	61.8	30.5	33.7	62.1	34.2

Table 5.12 : Combining classifiers: Voting and ML (QPSK)

Chapter 6

Conclusion

Ever increasing usage of wireless devices and inefficient utilization of wireless spectrum necessitate a paradigm shift in wireless communication. New technologies are emerging to use wireless spectrum more efficiently. Cognitive radio (CR) is one of the novel technologies that can adjust its parameters adaptively based on communication environment. Still broadcasting nature of wireless communication poses great threat for communication parties independent of the technologies. Many attacks has been listed in the past and impersonation (identity based) attacks is noted as one of the most critical ones due to its ease of implementation. Thus identification of a transmitted signal is one of the most challenging problems for securing wireless communication. Previous research addressed identity based attacks and proposed counter measurement techniques to prevent these type of attacks. However, these techniques are based on single radio adjustment and will no longer be sufficient for CR devices due to its ability to reconfigure parameters based on the environment. To the best of our knowledge, there is no previous work on addressing the security issues for CR. This thesis provides first attempt to generate a robust fingerprinting mechanism for different environment settings and radio configurations to secure CR communication.

We propose a signature scheme for identifying cognitive radio for different configurations (power, channel, modulation). Our signature scheme is based on the effects of the hardware imperfection on the transmitted signal. Due to manufacturing variability, devices have minute imperfections which cause deviations on the transmitted

signal. We analyze these deviations on the modulation domain to generate unique signatures for each device. We define classifying variables based on the characteristics of a signal in modulation domain which are listed as follows: Error Vector Magnitude (EVM) measurements based classifiers (phase error, magnitude error and EVM), I/Q Offset and frequency offset.

We perform experiments with WARP platform and Channel Emulator. Reconfigurable WARP boards are employed to emulate CR and channel emulator is used to cancel out channel effects to observe only the hardware imperfections on the transmitted signal. To simulate a possible CR activity, we choose different power, channel and modulation configurations. A predefined number of frames are transmitted from 12 WARP boards for each configuration. Based on the transmitted frames, we look at the effects of changing parameters on our classifiers. Our analysis show that individual classifiers are weak due to their high prediction error.

Next, we propose a signature scheme to combine our weak classifiers to form a strong classification method. We choose weighted voting to form a committee. Each classifier assigned a weight based on its probability of detection (P_D) and probability of false alarm (P_{FA}). Then we tested our method with random frames chosen from each board. The results are encouraging and we have average 97.7% and 96.8% of P_D for BPSK and QPSK modulation respectively. Also the P_{FA} rates are very low, only 2.3% and 3.2% for each modulation.

Finally, to analyze the impact of the channel on our signature scheme, we perform same set of experiments with indoor office setting of the Channel Emulator. Due to fading and multi-path effects, transmitted signals further deviate from the original signal. The presence of the channel also degrades the prediction rate of EVM based classifiers. However, I/Q offset and frequency offset classifiers are immune to channel

effect and due to weighted voting mechanism we still able to get high prediction rates. With the channel effect, we have an average of 96.6% and 95.9% P_D for BPSK and QPSK modulations. Average P_{FA} rate is only 3.4% for BPSK and 4.1% for QPSK.

Our results are promising and clearly show the superior performance of our classification method. We have high P_D values with low P_{FA} rate for both static channel and indoor office environment. These values are calculated based on observation of 5 frames and prediction rate will improve with the increase of sample set.

Bibliography

- [1] FCC, ET Docket No 03-222 Notice of proposed rule making and order, December 2003.
- [2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.
- [3] "Rice university warp project." <http://warp.rice.edu>.
- [4] K. Remley, C. Grosvenor, R. Johnk, D. Novotny, P. Hale, M. McKinley, A. Karygiannis, and E. Antonakakis, "Electromagnetic signatures of WLAN cards and network security," in *Signal Processing and Information Technology, 2005. Proceedings of the Fifth IEEE International Symposium on*, pp. 484–488, 2005.
- [5] R. Gerdes, T. Daniels, M. Mina, and S. Russell, "Device identification via analog signal fingerprinting:a matched filter approach," in *NDSS*, 2006.
- [6] O. Ureten and N. Serinken, "Bayesian detection of radio transmitter turn-on transients," in *Proceedings of NSIP99*, pp. 830–834, 1999.
- [7] O. Ureten and N. Serinken, "Detection of radio transmitter turn-on transients," *Electronics Letters*, vol. 35, no. 23, pp. 1996–1997, 1999.
- [8] O. Ureten and N. Serinken, "Wireless security through RF fingerprinting," *Electrical and Computer Engineering, Canadian Journal of*, vol. 32, no. 1, pp. 27–33, 2007.

- [9] O. Tekbas, N. Serinken, and O. Ureten, "An experimental performance evaluation of a novel radio-transmitter identification system under diverse environmental conditions," *Electrical and Computer Engineering, Canadian Journal of*, vol. 29, no. 3, pp. 203–209, 2004.
- [10] W. Suski, M. Temple, M. Mendenhall, and R. Mills, "Using spectral fingerprints to improve wireless network security," in *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pp. 1–5, 2008.
- [11] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proceedings of the 14th ACM international conference on Mobile computing and networking*, pp. 116–127, 2008.
- [12] A. Candore, O. Kocabas, and F. Koushanfar, "Robust stable radiometric fingerprinting for wireless devices," *Hardware-Oriented Security and Trust, IEEE International Workshop on*, vol. 0, pp. 43–49, 2009.
- [13] K. Talbot, P. Duley, and M. Hyatt, "Specific emitter identification and verification," in *Technology Review*, 2003.
- [14] L. Langley, "Specific emitter identification (SEI) and classical parameter fusion technology," pp. 377–381, 1993.
- [15] <http://www.decodesystems.com/mt/97dec/>.
- [16] M. Riezenman, "Cellular security: better, but foes still lurk," *Spectrum, IEEE*, vol. 37, no. 6, pp. 39–42, 2000.
- [17] M. Barbeau, J. Hall, and E. Kranakis, *Detecting Impersonation Attacks in Future Wireless and Mobile Networks*, pp. 80–95. 2006.

- [18] D. B. Faria and D. R. Cheriton, “Detecting identity-based attacks in wireless networks using signalprints,” in *Proceedings of the 5th ACM workshop on Wireless security*, pp. 43–52, 2006.
- [19] N. Patwari and S. K. Kasera, “Robust location distinction using temporal link signatures,” in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pp. 111–122, ACM, 2007.
- [20] Y. Sheng, K. Tan, G. Chen, D. Kotz, and A. Campbell, “Detecting 802.11 MAC layer spoofing using received signal strength,” in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pp. 1768–1776, 2008.
- [21] Y. Chen, W. Trappe, and R. Martin, “Detecting and localizing wireless spoofing attacks,” in *Sensor, Mesh and Ad Hoc Communications and Networks, 2007. SECON '07. 4th Annual IEEE Communications Society Conference on*, pp. 193–202, 2007.
- [22] 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*, pp. 33–42, 2006.
- [23] G. E. Suh and S. Devadas, “Physical unclonable functions for device authentication and secret key generation,” in *DAC '07: Proceedings of the 44th annual Design Automation Conference*, pp. 9–14, 2007.
- [24] M. Majzoobi, F. Koushanfar, and M. Potkonjak, “Techniques for design and implementation of secure reconfigurable pufs,” *ACM Trans. Reconfigurable Technol. Syst.*, vol. 2, no. 1, pp. 1–33, 2009.

- [25] Y. Alkabani, F. Koushanfar, N. Kiyavash, and M. Potkonjak, "Trusted integrated circuits: A nondestructive hidden characteristics extraction approach," vol. 5284, pp. 102–117, 2008.
- [26] DARPA XG WG, The XG Architectural Framework V1.0, 2003.
- [27] DARPA XG WG, The XG Vision RFC V1.0, 2003.
- [28] FCC, ET Docket No 08-260 Second Report and Order, December 2008.
- [29] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1998.
- [30] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, 2006.
- [31] S. M. Mishra, D. Cabric, C. Chang, D. Willkomm, B. van Schewick, A. Wolisz, and R. W. Brodersen, "A real time cognitive radio testbed for physical and link layer experiments," in *IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 562–567, 2005.
- [32] Z. Miljanic, I. Seskar, K. Le, and D. Raychaudhuri, "The winlab network centric cognitive radio hardware platform: Winc2r," *Mob. Netw. Appl.*, vol. 13, no. 5, pp. 533–541, 2008.
- [33] Q. Zhang, A. B. J. Kokkeler, and G. J. M. Smit, "Cognitive radio design on an mp soc reconfigurable platform," *Mob. Netw. Appl.*, vol. 13, no. 5, pp. 424–430, 2008.

- [34] P. Murphy, A. Sabharwal, and B. Aazhang, “Design of warp: A flexible wireless open-access research platform,” in *Proceedings of EUSIPCO*, 2006.
- [35] Agilent Technologies. Using Error Vector Measurements to Analyze and Troubleshoot Vector-Modulated Signals.
- [36] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer-Verlag, 2001.
- [37] L. Breiman, “Arcing classifiers,” *Annals of statistics*, no. 26, pp. 801–849, 1998.