A TEST METHODOLOGY FOR EVALUATING COGNITIVE RADIO SYSTEMS

THESIS

Jared J. Thompson, Capt, USAF

AFIT-ENG-14-M-77

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-14-M-77

A TEST METHODOLOGY FOR EVALUATING COGNITIVE RADIO SYSTEMS

THESIS

Presented to the Faculty
Department of Electrical and Computer Engineering
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Electrical Engineering

Jared J. Thompson, B.S.A.E.
Capt, USAF

March 2014

AFIT-ENG-14-M-77

A TEST METHODOLOGY FOR EVALUATING COGNITIVE RADIO SYSTEMS

Jared J. Thompson, B.S.A.E.
Capt, USAF

Approved:

//signed//	28 Feb 2014
Kenneth M. Hopkinson, PhD (Chairman)	Date
//signed//	28 Feb 2014
LTC Robert J. McTasney, PhD (Member)	Date
//signed//	28 Feb 2014
Maj Mark D. Silvius, PhD (Member)	Date

AFIT-ENG-14-M-77

Abstract

The cognitive radio field currently lacks a standardized test methodology that is repeatable, flexible, and effective across multiple cognitive radio architectures. Furthermore, the cognitive radio field lacks a suitable framework that allows testing of an integrated cognitive radio system and not solely specific components. This research presents a cognitive radio test methodology, known as CRATM, to address these issues. CRATM proposes to use behavior-based testing, in which cognition may be measured by evaluating both primary user and secondary user performance. Data on behavior-based testing is collected and evaluated. Additionally, a unique means of measuring secondary user interference to the primary user is employed by direct measurement of primary user performance. A secondary user pair and primary user radio pair are implemented using the Wireless Open-Access Research platform and WARPLab software running in MATLAB. The primary user is used to create five distinct radio frequency environments utilizing narrowband, wideband, and non-contiguous waveforms. The secondary user response to the primary user created environments is measured. The secondary user implements a simple cognitive engine that incorporates energy-detection spectrum sensing. The effect of the cognitive engine on both secondary user and primary user performance is measured and evaluated.

Table of Contents

List of Figures

Figure Page

List of Tables

List of Acronyms

Acronym	Definition
ACRO	AFIT Cognitive Radio Laboratory
AI	artificial intelligence
BER	bit error rate
BW	bandwidth
CE	cognitive engine
CN	cognitive network
CR	cognitive radio
CRATM	Cognitive RAdio Test Methodology
CRN	cognitive radio network
CRS	cognitive radio system
CUT	component under test
d-OFDM	dis-contiguous OFDM
DSA	dynamic spectrum access
DYSE	dynamic spectrum environment emulator
DySPAN	Dynamic Spectrum Access Networks
FFT	Fast Fourier Transform
FCC	Federal Communications Commission
FPGA	field programmable gate array
FSK	frequency shift keying
OFDM	orthogonal frequency division multiplexing
OODA	Observe-Orient-Decide-Act
OSA	Opportunistic Spectrum Access
PER	packet error rate

Acronym	Definition
PU	primary user
QoS	Quality of Service
REM	radio environment map
RF	radio frequency
RKRL	Radio Knowledge Representation Language
SDR	software defined radio
SPEC	Standard Performance Evaluation Corporation
SER	symbol error rate
SU	secondary user
SUT	system under test
USRP	Universal Software Radio Peripheral
WARP	Wireless Open-Acess Research Platform

A TEST METHODOLOGY FOR EVALUATING COGNITIVE RADIO SYSTEMS

I. Introduction

First introduced in 1999, cognitive radio (CR) refers to the emergence of technology that combines software defined radio (SDR), dynamic spectrum access (DSA), networking, and artificial intelligence (AI) techniques. SDR refers to radios that use software to define some or all physical layer functions [1]. The key attribute of SDR is that through software, changes can be rapidly made to the operating characteristics of the radio. For example, waveform modulation can be changed by issuing a software command as opposed to physically changing hardware components. The capabilities enabled by SDR lead to DSA. DSA is the ability to use spectrum that is available in time, frequency, or space [2]. CR capitalizes on both SDR and DSA by enabling intelligent use of spectrum through networks and devices via cognition. Cognition is primarily accomplished through the use of AI to make decisions in support of user policy.

There are several benefits of using CR that solve current real-world problems. First and foremost, CR addresses the issue of spectrum congestion through its use of DSA. Spectrum congestion is the result of spectrum being allocated by a central governing authority; some frequency bands are assigned to users that rarely, if ever, use their allocated portion of the spectrum while other frequency bands are over-utilized by many users. Spectrum congestion has become a more visible problem in recent years due to the explosive growth in the number of wireless devices. CR helps mitigate spectrum congestion by intelligently utilizing unused portions of the spectrum. Mitigating spectrum congestion improves network operation and improves end-user data rates. A second problem CR addresses is tactical in nature. CRs, as intelligent agents, can utilize their knowledge of the spectrum to self-organize, covertly communicate, coordinate action against the enemy, and provide resiliency and redundancy against enemy action. Both of these problems that CR address are stated in the Air Force Technology Horizons report [3].

1.1 Problem Statement

The objective of this research is to develop a framework by which CRs can be evaluated as a complete system. Because CR spans several disciplines, no standardized test methodology has yet been developed for use in test and evaluation of CRs [4]. Furthermore, because CR contains AI, the long term behavior and performance of a CR may be unknown as the AI evolves in reaction to its environment, making test and evaluation of a CR more difficult. According to Zhao, "how to effectively yet trustfully validate a CR device under varying known or even unknown scenarios is an open issue to address" [5].

However, the performance and behavior of a CR must be well understood if CRs are to be certified and used outside of the laboratory. For example, Federal Communications Commission (FCC) compliance requires devices to only transmit on assigned frequencies. Traditionally, this compliance is measured through hardware verification and validation. However, a CR may have the hardware capability to transmit outside of assigned frequencies, and only software prevents it from doing so. Ensuring that AI software meets the same level of compliance as traditional hardware verification and validation is an area of on-going research. As a result, there is no method of verification and validation for CRs. Methods have been proposed to evaluate certain components of a CR, but none offer a comprehensive evaluation of a complete CR system. *The lack of a framework to test CRs as a complete system across multiple technology domains is the primary problem this research seeks to address.*

Specifically, the goals of this research are to:

- develop a test methodology to evaluate CRs as a complete system,

- validate the developed test methodology,
- develop a CR prototype to include a spectrum sensor for use in testing and development.

Note that this research applies only to physical CR devices, not to theoretical or simulated CR devices. The assumption is that the CR under test is a real-world device. As a real-world device, the CR can be seamlessly moved from the test environment to being used in day to day life. Stated another way, a CR is considered to be the combination of the hardware and software that enables cognitive functionality, not simply the software. This research treats the words CR, cognitive radio system (CRS), and cognitive radio network (CRN) as extensions of the same technology. While similar, this research does not directly address cognitive network (CN)s. This is explained further in Section 2.1.

This research effort partially achieved all goals. A test methodology was developed and then validated and shown to be effective in evaluating CRs within the constraints of this research. However, the scope of the validation does not justify the developed test methodology in its entirety. A working CR prototype was developed for use in testing. Also, a spectrum sensor was developed and implemented in hardware and may be used for future research efforts.

1.2 Contributions

This research provides three contributions to the CR field. The first is a test methodology that can be used to standardize CR evaluations, regardless of CR architecture or hardware-specific implementation. Secondly, this research provides a new way of approaching testing cognition by measuring device behavior. A behavior-based approach to testing and evaluating cognition is shown to be effective and negates the need to evaluate cognition on a component level. The results of this research serve as a “litmus test” in establishing the utility of behavior-based testing. Thirdly, this research provides a CR prototype using the Wireless Open Access Research Platform (WARP) board from Rice University. The CR prototype is capable of sensing the environment, reacting to the environment, and communicating with other devices using multiple waveforms. These research contributions apply not only to CR but also to DSA and Opportunistic Spectrum Access (OSA) fields as well.

1.3 Overview

Chapter 2 provides a detailed background on existing research related to test and evaluation of CRs. The chapter begins with formal definitions of what a CR is and then examines multiple CR architectures that have been proposed. Following that, the chapter presents relevant CR testbeds and prototypes. Finally, the chapter examines CR performance metrics, benchmarks, and methodological aspects that pertain to development of a test methodology.

Chapter 3 presents the developed test methodology, known as Cognitive RAdio Test Methodology (CRATM). CRATM is designed to address the issues presented in Chapter 2. A comparison of CRATM with the literature is also presented.

Chapter 4 presents the methodology to evaluate the developed test methodology. Due to scope, CRATM is not evaluated in its entirety. Instead, one key component behind CRATM, behavior-based testing, is investigated. Research experiments are described in this chapter.

Chapter 5 presents the results of the experiments. Empirical results are collected on an implementation of CRATM using a CR prototype in regards to behavioral-based testing.

Chapter 6 presents a summary of the experimental results and of the developed test methodology. Conclusions and future work are also presented in Chapter 6.

II. Related Work

The CR field is diverse and with few formal systemic categorizations and classifications. However, the boundaries of a system must be well understood if it is to be effectively evaluated. This chapter presents a current assessment of the CR field and how it is classified. The various definitions of CRs are examined first, followed by the features of an assortment of CR architectures. After that, testbeds, prototypes, and test methodologies for CRs are presented. The information presented in this chapter is intended to provide context for the developed test methodology in Chapter 3.

2.1 Cognitive Radio Definitions

This section on definitions answers “what does a CR do?” as opposed to the following section on architectures, which answers “how does a CR work?”. In this section, the use of the term ”lower” layers refers to the physical layer and data link layers while “higher” layers refers to the network through application layers.

As an immature field of research, CR has varying definitions on what it means to be a CR. The lack of agreement on the definition of a cognitive radio has direct implications on how a CR is tested. A methodology that does not test the intended system under test (SUT) will not be very useful.

Prior to presenting specific definitions on CRs, the relationships between CRs, CRSs, CRNs, and CNs must be understood. A CR is a single device, which may be used in a CRN. A CRN is a collection of cognitive and non-cognitive devices that may also feature cognition on the network layers. By contrast, a CN requires cognition at the network level, but does not require the use of CRs. A CRS is used in this research as an all-encompassing term to describe either a CR or a CRN. The test methodology developed in this research is proposed to be extensible to CNs; however, its applications to CNs are not directly investigated. Instead this research investigates only CRs and CRNs as CRSs.

The remainder of this section details historical or prominent CR definitions.

2.1.1 Mitola, 1999.

The term "cognitive radio" was initially coined by Joseph Mitola to describe wireless devices and networks intelligent enough to detect user communication needs and provide wireless services in response to those needs [6]. A CR is a SDR that makes use of the radio environment intelligently through automated reasoning about the needs of the user. In Mitola's framework, the radio environment is shared and represented through a Radio Knowledge Representation Language (RKRL). The RKRL provides a standard language through which all device knowledge can be shared with other devices, including knowledge of the spectrum, local policy, network information, device information, and user needs. In this framework, CRs employ a cognition cycle based on the Observe-Orient-Decide-Act (OODA) loop [7], but with the addition of plan, learn, and act stages such that it becomes a OOPDLA cycle [6].

It is important to point out that Mitola views CR from the perspective of a singular CR node interacting with existing networks and devices, regardless if those networks and devices are cognitive or not. The primary means of cognition for Mitola is the use of the RKRL in conjunction with a priori knowledge to change the behavior of the radio at low layer levels [6]. In other words, CR is an extension of SDR in that it intelligently uses the PHY/MAC layers. Higher layer cognition is not expected.

2.1.2 Kantor, 1999 and Thomas, 2005.

By contrast, a fellow doctoral student with Mitola, Theo Kantor, presented the concept of a cognitive network as a network with memory at the same time that Mitola presented his concept of CR [8] (interestingly, both shared the same advisor). The idea of a cognitive network is formalized in [9] as a network that has cognitive processes that sense, plan, decide, learn, and act on network conditions while taking into account end-to-end goals. This cognitive network acts across all layers, not just the PHY/MAC layer. In essence, a CN takes the properties of a CR and implements those properties across all layers. The key difference between a CN and a CR is that the CN takes into account overall end-to-end goals as opposed to node-specific user objectives [10]. CNs may or may not include low layer level cognition such as CRs [11].

2.1.3 Cognitive Radio Networks.

From the foundations of CRs and CNs, the idea of a CRN has emerged in which each end-user node in the network is a CR [12]. In this model, both a cognitive network and a cognitive radio can co-exist, allowing cross-layer optimization depending on the level of cognition of the system. This model allows integration of node-specific CR techniques with the end-to-end goals of CNs.

2.1.4 Haykin, 2005.

Haykin in [13] presents the cognitive radio as an intelligent wireless communication system that is aware of the environment, uses understanding-by-building to learn from the environment, and adapts to RF stimuli with the objectives of highly reliable communication whenever and wherever needed and to efficiently utilize the spectrum. Haykin envisions a CR as adapting its operating parameters, such as transmit power, carrier frequency, and modulation in order to fulfill these objectives. Haykin presents the *spectrum hole* as the means in which a CR operates around existing users. A spectrum hole is a frequency that is not being used by the primary user (PU) at a specific time and geographic location. Note that Haykin, like Mitola, is primarily focused on improving communications at the physical layer by efficiently using the spectrum. Haykin also proposes using interference temperature (discussed further in Section 2.4.2) as a performance metric and *radio scene analysis* as a means of evaluating detection of spectrum holes.

2.1.5 IEEE 802.22 and DySPAN.

The IEEE has long been working towards developing a standard for a real-world CR. There has been two parallel tracks–one that is developing generic supporting standards and another developing a specific implementation called IEEE 802.22. The first committee to work towards developing supporting standards was the IEEE 1900 standards committee. This committee evolved into the Standards Coordinating Committee 41 (SCC41), which in turn evolved into the Dynamic Spectrum Access Networks (DySPAN) committee [14, 15]. DySPAN's working definition of a CR is a "radio in which communication systems are aware of their environment, internal state, and location and can make decisions about their radio operating behavior based on that information" and "utilizes software defined radio, adaptive radio, and other technologies to autonomously adjust its behavior or operations to achieve the desired objectives" [16].

The IEEE 802.22 standard [17–19] is working towards a CR implementation that follows the standards presented by DySPAN. The IEEE 802.22 standard proposes to utilize unused portions of the television spectrum without causing interference to existing television receivers. A *spectrum broker*, or *spectrum manager*, is inherent to the operation of IEEE 802.22 [19]. A spectrum broker is a device responsible for dynamic assignment of channels to secondary devices. IEEE 802.22 assumes that there are base stations which facilitate spectrum coordination by acting as spectrum brokers. Individual sensing nodes may feed local spectrum information to the base stations [17]. The base station spectrum broker has a geolocation database of known licensed transmitters, but supplements spectrum knowledge by using localized spectrum sensing [19].

Both the IEEE 802.22 and DySPAN standards do not actually require (nor currently have standards for) cognition at the CR level. Instead, each node on the network only needs to implement the policies dictated by the spectrum broker [15, 20]. Sensing nodes are required to feed spectrum knowledge back to the central base station, though this act does not require cognition. Instead, all cognition occurs via a *cognitive engine (CE)* located at the base station. Non-intelligent devices consult a spectrum database, which is created and maintained by the intelligent CE [19, 20].

DySPAN also presents two models of spectrum allocation and usage: owned and common. In the *owned* model, sole ownership of the spectrum rests with a user or agency that has bought that portion of spectrum. In the *common* model, the spectrum is available to all users where the users agree upon an etiquette to operate by [16].

2.1.6 Federal Communications Commission.

The FCC has a loose definition of cognitive radio that is better described as a definition of software defined radio. At the present, the FCC does not address cognitive radio apart from SDR as it does not believe that the fundamental nature of spectrum policy is ready to change [21]. The FCC defines a CR (or SDR) as "a radio that includes a transmitter in which the operating parameters of frequency range, modulation type or maximum output power... can be altered by making a change in software without making any changes to hardware components that affect the radio frequency emissions." [21] The FCC proposed using *interference temperature* as a metric for evaluating the co-existence of unlicensed users with licensed users [22].

2.1.7 Wireless Innovation Forum.

The Wireless Innovation Forum, formerly known as SDRForum, presents a cognitive radio as simply a radio that is "capable of making decisions and selecting or modifying the operating parameters of a radio" [1]. The SDRForum views a CR as a SDR that is controllable at the physical layer while all other layers, except application, are part of the cognition process. The goal of the SDRForum framework is to make the cognition operation transparent to higher layers [1].

2.1.8 DARPA XG Program.

Completed in 2006, the next-generation (XG) program used DSA techniques to operate without causing interference to pre-existing non-cooperative users [23]. The XG program does not claim to use a CR per se, but as an actual implementation of a DSA device, utilizes many CR techniques that are part of a basic CR. The three XG success criteria of relevance to CR applications are:

- Not cause harm to existing users
- Form and maintain connected networks
- Add value by efficiently using the spectrum

Taken together, these three criteria form a definition on what a CR does [23].

2.1.9 Cognitive Radio Definitions Summary.

In general, these definitions agree that CRSs should be able to both sense the environment and autonomously adapt to changing conditions but differ as to to the depth of cognitive functionality, situation awareness, and where cognition takes place [24]. A summary of cognitive radio definitions as presented in this section is shown in Table 2.1.

2.2 Architectures

The architecture of a CR directly affects how it is implemented, and therefore tested. This section presents an overview of key CR architectural features, as well as two representative CR architectures. In general, a CR requires a means of sensing the spectrum, sharing spectrum knowledge, combining spectrum knowledge with the knowledge of other devices, establishing spectrum cooperation with other devices, and communicating with the network. Implicit in these activities is a CE, which forms the nucleus of the AI. The CE implements user and external policy to accomplish user goals. Another implicit feature is network topology.

Table 2.1: Different Capabilities On What Is Necessary To Be Called A Cognitive Radio Adapted from [24]

	Capability												
	Adapts (Intelligently)	Autonomous	Sense Environment	Adaptive Transmitter	Adaptive Receiver	Environment ”Aware”	Goal Driven	Learn the Environment	Capabilities ”Aware”	Negotiate Waveform	No Harmful Interference	Cognition at Radio	Cognition at Base Station
Mitola	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Haykin	✓	✓	✓	✓	✓	✓	✓	✓				✓	
IEEE 802.22 / DySPAN	✓	✓	✓	✓	✓	✓					✓		✓
FCC	✓	✓	✓	✓									
SDRForum	✓	✓	✓	✓	✓	✓	✓		✓			✓	
DARPA XG	✓	✓	✓			✓	✓				✓		

All CR architectures are governed by the OODA loop. A representative version of the OODA loop as it pertains to CR is shown in Figure 2.1[1]. This cognition cycle forms the simplest possible framework for understanding the reaction of the CR to the environment. Note that though it is called a cycle, stages can occur concurrently. For instance, the *observe* stage may occur while the *act* stage is underway.

In CRs, the *observe* stage corresponds to spectrum sensing and spectrum cooperation. The *orient* stage corresponds to applying user policy based on observations. The *decision* stage corresponds to choosing a course of action to adapt to the observed environment, and

---

[1] Reproduced from [25]

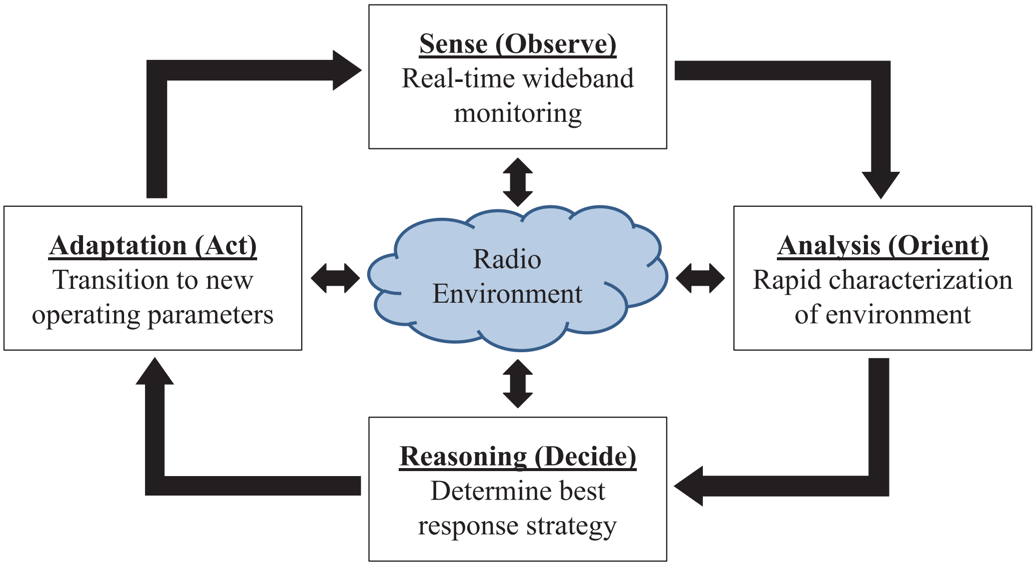


Figure 2.1: Cognitive Radio Cognition Cycle [26]

the *act* stage corresponds to the implementation of the decision and performing normal operations.

In terms of CR, the OODA loop can be illustrated as follows. The spectrum is broken up into 10 channels. While currently operating on channel 2, a CR senses the spectrum and detects a PU on channel 3. This information is shared with other CR devices on the network (*observe*). A master node receives this spectrum information and combines the spectrum information with its spectrum database (*orient*). The master node notices that the PU on channel 3 will be interfered with by several nodes based on their relative proximity. The master node decides that communication is best served by changing to channel 6 (*decide*).

The master node commands the CRs to jump to this new channel and to continue operation (*act*).

2.2.1 Spectrum Sensing, Sharing, and Cooperation.

The issue of spectrum sensing, sharing, and cooperation is well established in the literature. Readers are referred to [2, 12, 27–29] for a comprehensive treatment of spectrum sensing and cooperation. For the purposes of developing a test methodology, it suffices to state that there are many means to choose from to detect PUs and form spectrum knowledge. Spectrum sensing algorithms range from simple algorithms that only look for increased power relative to noise floor to algorithms that detect features beneath the noise floor. Each spectrum sensing algorithm offers a different degree of resolution with corresponding computational complexity.

Two general means of recording and sharing spectrum knowledge have been proposed. The first is to utilize a language-based representation such as the RKRL proposed by Mitola [6]. The second is to utilize databases. A prominent database form of spectrum knowledge is the radio environment map (REM). In the REM, database elements include geographical information, services and networks, regulations and policy, the activity profile of radio devices, and learned experience [30]. If a CE is the brain, the REM is the memory the brain draws upon. The IEEE 802.22 / DySPAN standards combine elements from both language-based representations and database forms of spectrum knowledge [19, 20].

Spectrum information sharing and cooperation is closely related to the CR topology presented in Section 2.2.2.

2.2.2 Topology.

There are two primary CR topologies, centralized and distributed [12, 28]. In the centralized topology, a central node acting as a base station or access point serves to facilitate communication with secondary user (SU)s. In general, the base station controls all of the SUs within transmission range. SUs feed spectrum sensing data forward to the base station. The centralized topology may or may not require two communication channels– one for the primary communication/observation channel and one for reporting spectrum information to the base station.

The second topology is the distributed topology. This topology can be thought of as an ad hoc or mesh network. In this topology, SUs communicate directly with each other without any central node. Spectrum data is shared between SUs and each SU can independently make a decision on the appropriate communication protocol. Optionally, one node can be designated as a head node with spectrum decision authority. As with the centralized topology, a distributed topology may or may not require two communication channels.

2.2.3 Cognitive Engine.

The CE is the intelligence behind a CR [31]. Research into CEs overlaps with AI research. Technologies behind CEs include genetic algorithms, case-based reasoning systems, expert systems, or combination of these approaches [24, 31]. According to [5], there are two approaches to implementing a CE within a CR. The first is a low-complexity CE approach. In this approach, CRs use REMs or similar database both locally and globally for situational awareness, efficient learning, and fast adaptation. The goal with this approach is fast adaptation for low cost and complexity. The second approach is to use high performance computing. In this approach, CRs utilize either multi-core CPUs or offload data computations to the cloud. Both of these approaches assume a centralized topology.

When developing a test methodology, it is important to have a framework that is not limited to testing only one permutation of the CE, but is all-inclusive.

2.2.4 Miscellaneous Features.

Other features of a CR architecture are derived from real-world technology constraints. For instance, a working CRS may need a common control channel other than the primary communications channel. This common control channel serves as a means to share spectrum information or issue commands to nodes in the event that the primary communications channel is not useable. The common control channel is usually thought of as a low bandwidth alternative. Another architectural feature is whether or not the CRS is homogeneous or heterogeneous. These features are often implicitly assumed in some CR definitions or architectures. The test methodology developed in this research seeks to avoid tying itself to any particular CR definition and instead aims to be flexible enough to accommodate all definitions and architectures.

2.2.5 IEEE 802.22 Architecture.

The IEEE 802.22 proposal serves as one type of CR architecture. The IEEE 802.22 standard [17, 19] is intended to allow SU devices to co-exist with television stations PUs by using white spaces (i.e. spectrum holes). IEEE 802.22 is a wireless regional area network (WRAN). The chief problem is ensuring that various SU devices do not interfere with PU receivers. To do this, IEEE 802.22 uses a spectrum broker located at the base station. The spectrum broker polls CR nodes for spectrum information and compiles that information into a global REM database. Then, the spectrum broker allocates spectrum to each node based on the global REM database and the policy goals of the network. In the IEEE 802.22 architecture, nodes do not have to be cognitive so long as they can implement the instructions of the spectrum broker. However, there is a requirement for spectrum sensing to occur away from the base station at the nodes.

2.2.6 AFIT Cognitive Radio Lab Architecture.

The AFIT Cognitive Radio Laboratory (ACRO), based on work by McLean [32], puts forth an ad-hoc CR architecture. In the ACRO architecture, the nodes follow the seven step process shown in Figure 2.2. In this framework, individual nodes sense the spectrum and share REMs with each other using a protocol such as multicast. The REM in this architecture is simply a binary vector, where each position in the vector is a 1 or 0

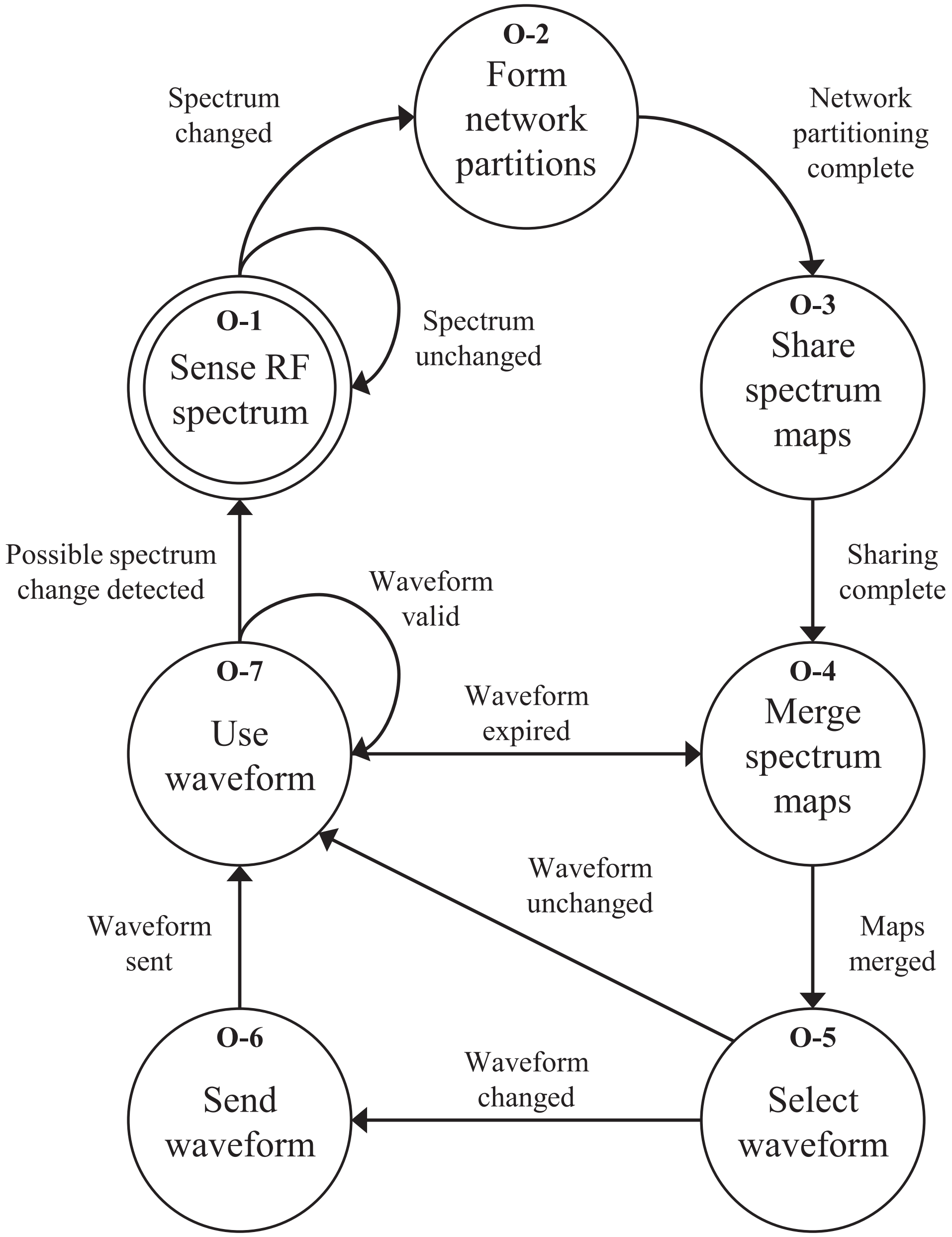


Figure 2.2: AFIT Cognitive Radio Lab Architecture System Functional Diagram [32]

depending on if the corresponding channel is occupied or not. Implicit in the architecture is a master node that compiles distributed REMs and redistributes a shared REM back to the rest of the nodes. The architecture does not specify if the REM sharing occurs on the primary channel or on a common control channel. The architecture uses frequency hopping when selecting waveforms.

2.3 Testbeds

Testbeds are a critical tool in the evolution of CR technology from theoretical analysis to real-world hardware that will be used outside of the lab [26]. Several have been developed for use in prototyping CRs. These testbeds generally combine real hardware devices with emulation, in which the real hardware devices can communicate through a controlled wireless or wired physical layer. However, some testbeds feature no emulation and offer no direct control of the radio frequency (RF) environment. According to [33], every CR testbed should provide the ability to support multiple radios (both PUs and SUs) that are fully controllable, the ability to connect various different front-ends over different frequency ranges, the ability for physical/link layer adaptation and fast information exchange, and support of rapid prototyping. Problems with current testbeds include a lack of large scale capabilities, a tendency to focus on 802.11 networks, limited integration of hardware, and systems that are either too complex or too simple [34].

This section presents two approaches to building and operating CR testbeds. Note that the specific testbeds described in this section are predicated on testing a physical radio device. Systems that simulate or emulate the system under test are not considered. Also, systems that playback traces of the environment, such as in [35], are not considered in this research. Trace based testing is excellent for running realistic simulations but does not lend itself towards the testing of physical devices. Furthermore, testbeds that do not enable testing of unique hardware are not considered in this research. The majority of testbeds in the literature are of this type–these testbeds have a fixed hardware RF front-end and are focused on algorithm development, and cannot be used to test generic hardware configurations.

The testbeds presented in this section are also predicated on offering some control over the RF environment. According to [36], "research aimed at evaluating and improving wireless network protocols and applications is hindered by the inability to perform repeatable and realistic experiments." If experiments are to be repeatable, control over the RF environment is necessary. However, this must be balanced by the requirement for a realistic RF environment. The ideal wireless experimentation device would offer repeatable results, controllability, cross-layer realism, run real experiments, be quickly configurable, and test a large number of nodes. Emulation offers the easiest way to meet these criteria.

2.3.1 Emulation-based Testbeds.

Emulation offers a means of tightly controlling the variables in the experiment while still offering a well-defined dynamic environment which can be simultaneously realistic, fully controllable, repeatable, and diverse [37–39]. Emulation also enables the outputs of one sub-system to be fed into another sub-system. According to Borries et al., emulation is particularly well suited to device and link characterization for controlled studies in the early stages of a project into physical layer effects [39]. They also claim that emulator capabilities are unique in supporting experiments that are simultaneously realistic, fully controllable, repeatable, and diverse.

To emulate the RF environment, there are three main approaches [40]. The first approach is to scale the radio signals between wireless devices via attenuation. The second approach is to emulate just the channel path between wireless devices using a hardware-based channel emulator. The third approach is to emulate the wireless device and channel effects. The goal of this research is to investigate how real wireless devices behave in a controlled environment; therefore, the use of a hardware-based channel emulator (the second approach) is most appropriate. Existing hardware-based channel emulator testbeds include the CMU Emulator, WHYNET, BEE2, and the DYSE. Alternative testbeds, ORBIT and MiNT, utilizes the first approach.

2.3.1.1 CMU Emulator.

The Carnegie Mellon University (CMU) Emulator [39] is a 15 node emulator that has been in use since 2007. The emulator supports the full 2.4 GHz ISM band. The wireless nodes are laptops with 802.11b interfaces, though the system has the capability to incorporate other RF devices. A Linux server acts as the environment controller and controls the channel parameters that are computed by field programmable gate array (FPGA). The output of the wireless nodes is fed to FPGAs for channel path processing. The FPGAs emulate the desired signal effects such as attenuation, fading, multi-path, and interference. The FPGAs also combine the processed signals into an output signal, after which it is converted to RF and streamed back to the wireless nodes.

2.3.1.2 WHYNET.

The Wireless Hybrid Network [40], WHYNET, is a framework that enables integration of physical, simulation, and emulation components. The system is built to understand cross-layer protocol interactions by exploiting physical layer flexibility. The system is designed to provide a realistic, scalable, flexible, and cost-effective evaluation environment. The physical components of WHYNET consist of wireless devices that communicate over real wireless channels. The emulated components consist of a hardware-based channel emulator and a wireless network emulator. The hardware-based channel emulator computes the emulated signal between wireless nodes. The wireless network emulator emulates the devices and combines both device and channel path behavior into software models. WHYNET also has simulation capabilities. The testbed infrastructure includes a variety of communications testbeds and devices.

2.3.1.3 BEE2.

The Berkeley Emulation Engine 2 (BEE2) [33] is a an FPGA-based channel emulator that supports up to 18 wireless nodes. The BEE2 designers envisioned the platform to serve as a testing ground for CRs, enabling both PUs and SUs to operate in the same space for testing. The PUs are directly controllable and precise interference measurements can be made. The BEE2 unit contains five Virtex-2 Pro 70 FPGAs, of which four are used for channel emulation and one is used for control. The BEE2 accepts up to 20 MHz of RF bandwidth and can operate within the entire 2.4 GHz ISM band.

2.3.1.4 ORBIT and MiNT.

The Open Access Research Testbed for Next-Generation Wireless Networks (ORBIT) is a two-tier laboratory emulator and field trial network testbed designed to achieve reproducibility of experimentation while also supporting real-world settings. The testbed has over 400 802.11 a/b/g nodes laid out on 20 by 20 grid. Research at ORBIT is focused on higher-layer network protocols with limited physical layer research [41, 42]. ORBIT emulates various environments by attenuating and mixing signals. The Miniaturized Mobile Multi-Hop Wireless Network Testbed (MiNT) [43] is a similar testbed to ORBIT, though it focuses on 802.11b networks.

2.3.1.5 CREW.

Cognitive Radio Experimentation World (CREW) Federated Testbed [37] consists of five geographically separated laboratory testbeds located throughout Europe. The CREW testbeds are intended to be able to perform experimentally driven research across all layers and to capitalize on the strengths of individual testbeds for different applications. CREW is focused on cognitive networks and dynamic spectrum access (DSA) research. The CREW approach uses traces recorded in one testbed to play back in another to enable a form of emulation. The chief limitation of this approach is that discrete testbeds will not cause RF interference with each other, which means that the cognitive system is not fully characterized.

2.3.1.6 DYSE.

Dynamic Spectrum Environment Emulator (DYSE) [44] is a machine that tests wireless devices by emulating channel path conditions in near real-time for both real and emulated wireless devices. For each node, controllable parameters include movement, delay, and antenna gain patterns. The DYSE has predefined channel settings that include air-to-air, air-to-ground, rural, suburban, urban, and dense urban. The DYSE offers control of channel parameters for path loss, fading, and multipath. Real world terrain data is fed into the system via the Google Static Maps API service or through Terrain Integrated Rough Earth Model (TIREM) data. Therefore, wireless paths between nodes can be calculated for a wide variety of realistic test conditions, making it well suited for diverse experimentation with DSA radios.

The DYSE operator specifies the location of each node and whether it is a transmitter or receiver via a GUI. The operator also specifies the antenna gain patterns and movement of each node through the GUI as well. The channel path characteristics are automatically populated using the altitude and location of each node from the TIREM dataset unless the operator chooses to specifically define the channel path characteristics.

The DYSE can operate with all virtual nodes (all nodes are emulated), with only physical nodes attached, or with a combination of both physical and virtual nodes (mixed mode). In virtual-only mode, the DYSE can emulate up to 100 nodes at once. In this mode, the system operates in snapshots since real-time operation is not necessary. By contrast, the physical-only mode requires and operates with full real-time streaming. The gross-processing delay is non-time varying but is fixed for a specific scenario. For a nominal 4 physical unit scenario (16 channels), the gross processing delay is approximately 500 μs. The mixed virtual and physical mode is the most complex. Real-time operation is capable for simulations with low complexity; however, at a certain point the system can no longer support real-time operation and continuity of signals cannot be guaranteed. This crossover threshold is scenario dependent. Two major factors affecting real-time operation in mixed mode are the number of virtual nodes and the complexity of virtual nodes.

The DYSE hardware consists of an workstation computer hosting a graphics processing unit (GPU) and an RF interface tray for interfacing with physical devices. The workstation utilizes a single 6-core, 3.33 MHz Intel i7-980x CPU with 12GB of triple channel DDR3 DRAM. The workstation runs on Linux Ubuntu. The GPU utilizes the Nvidia®CUDA API. The GPU is responsible for signal propagation computation. The RF interface tray consists of four Universal Software Radio Peripheral (USRP) SDRs from Ettus Research. Each USRP performs digital-down/up conversion between RF and baseband. The RF interface tray is limited to 12.5 MHz of bandwidth per USRP. The USRPs operate over the frequency range from 50 to 2200 MHz.

2.3.2 Non-Emulation Testbeds.

There are a large number of non-emulation testbeds used by the CR community. Since this research is focused on testbeds that offer control of the physical layer, they are not presented here. The reader is referred to [5, 34, 37, 39, 42, 45] for information on non-emulation testbeds.

2.4 Performance Metrics

Performance metrics for CR may measure the performance of a CR component, an individual CR node, or the CR system as a whole. Some performance metrics, such as those that apply to spectrum sensing, may apply across these domains. The use of uniform performance metrics is necessary for standardized evaluation. The most common performance metrics are presented here; however, the authors of [24] provide an exhaustive treatment of performance metrics for CRs.

2.4.1 Spectrum Sensing.

Research on spectrum sensing has focused on the ability to detect a PU. The following are established performance metrics.

- *Probability of Primary User Detection:* This metric is the probability that the PU is properly detected. This detection has both a false negative and a false positive component. The false negative component is the probability of not detecting the PU, while the false positive component is the probability that a PU is detected that is not present [33]. The false positive is also referred to as the probability of false alarm. Mathematically [26], the null hypothesis is H_0, or the hypothesis that there is no PU on the channel. The test hypothesis is H_1, or that there is a PU on the channel. The test statistic from the spectrum sensing unit is T, and is compared to a predetermined threshold λ. The probability of false alarm, P_F, denotes the probability that the hypothesis test chooses H_1 while it is actually H_0. The probability of detection, P_D, denotes the probability that the test correctly decides H_1.

  $$P_F = P(T > \lambda | H_0) \tag{2.1}$$

  $$P_D = P(T > \lambda | H_1) \tag{2.2}$$

  The probability of PU detection is best assessed at the component level, unless absolute spectrum knowledge of both the SUT and the environment is known.

- *Time to Detect a Primary User:* T_{DETECT} is the time between the initial PU signal transmission and the time the CRS is updated with correct spectrum knowledge [33]:

  $$T_{DETECT} = T_{REM} - T_{PU,TX} \tag{2.3}$$

  As with the previous metric, the time to detect a PU is difficult to measure in hardware and is best suited for laboratory or simulating testing.

2.4.2 Avoiding Causing Interference.

The following metrics pertain to the ability of a CR to avoid causing harmful interference to existing users. According to Kolodzy, "the design and operation of

RF equipment including communications and emitting non-communications devices are predicated upon preventing and/or mitigating electronic interference" [46]. In accordance with this statement, this research assumes that the ability to avoid causing interference is a mandatory capability of CR.

- *Time to Evacuate a Channel:* T_{EVAC} is the time it takes a CR, or CRS, to detect a signal (T_{DETECT}) plus the time it takes for the node to process the spectrum knowledge, decide to cease transmitting, and to ultimately cease transmitting (T_{CEASE}) [23, 33]:

  $$T_{EVAC} = T_{DETECT} + T_{CEASE} \tag{2.4}$$

  This metric was used in the XG program, where success was defined as a channel abandonment time of less than 500 ms [23].

- *Interference Temperature:* Interference temperature was first proposed by the FCC as a way to set a limit on how much interference an unlicensed user causes to licensed users [22]. Interference temperature is a measure of how much RF power is available at a receiving antenna to be delivered to a receiver that incorporates the power generated by other emitters and noise sources [46–48]. The equation for interference temperature is:

  $$T_I(f_c, B) = \frac{P_I(f_c, B)}{kB} \tag{2.5}$$

  In this equation, $P_I(f_c, B)$ is the average interference power in Watts for a center frequency f_c and bandwidth B measured in Hertz. Parameter k is Boltzmann's constant. Interference temperature is similar to noise temperature [47].

  Interference temperature was developed to address CRs utilizing the underlay paradigm (see Section 2.8) [22]. However, it is suitable for use in other paradigms as well.

- *Probability of Collision:* Instead of measuring interference temperature, interference can be measured as the likelihood that the SU transmits simultaneously with the PU [2]. Due to the binary results of this metric, it does not offer fidelity or adequate resolution for assessing the impact of the SU on the PU. However, it is simple to measure in a hardware SUT device.

2.4.3 Cognitive Radio Artificial Intelligence.

For CRSs, Zhao et al [5] propose to use "IQ" and "EQ" to test CRs, where the IQ is the intelligence of a single CR node while the EQ is the intelligence of a cross-node/network collaboration of the CRS. EQ is also the capability of obtaining global environmental awareness through collaborative sensing with other nodes or through network infrastructure. CR AI performance may also be viewed as a set of tasks to be performed, and evaluated as such [13].

2.4.4 Communication Performance.

Legacy performance metrics for evaluating communication performance are sufficient for CRS applications. Legacy performance metrics include measuring bit error rate (BER), packet error rate (PER), throughput, or network Quality of Service (QoS) for the CRS. These metrics are well established and not discussed further here.

2.5 Benchmarks

Benchmarks offer a means to provide repeatable experiments and to provide comparability between various CRs being tested. The dynamic interaction of environments, goals, and capabilities of CRs means that creating a generic benchmark is non-trivial [24]. Furthermore, the performance of a CR may change over time as it adapts and learns. However, the benefits of benchmarking include providing a basis for spectrum regulators to certify and regulate CR, vendors for type approval testing, and service providers for implementation [24]. The authors in [33] concur, stating that it is crucial for a common set of representative test cases to be used that every CR has to pass.

The workload offered by a benchmark is the RF environment. Because the RF environment is naturally variant, multiple replications are needed to achieve a suitable level of statistical confidence and to realize comparability [37].

The consensus among researchers is that the best way to test CR systems is to use representative cases of the radio environment [5, 33, 37]. CREW [49] researchers propose having various reference scenarios (home, office, public buildings) and wireless technologies characterized and then use these reproducible reference scenarios in tests. Both the radio and network can be tested and metrics collected this way. However, erroneous results may be indicated when there is observed external RF interference. This entire process can eventually be automated to sweep the entire range of interest [37].

2.6 Cognitive Radio Test Methodologies

A test methodology describes the overall procedure in how a CR is evaluated. Here, strategies in implementing that overall procedure are presented. These strategies are all suitable for benchmarking.

2.6.1 Radio Environment Map Scenario Driven Testing Approach.

Zhao et al. propose the use of REMs as the means of testing CR systems [30]. These REMs represent scenarios, and REM-based radio scenario testing (REM-SDT) can be performed. This approach is transferable across different systems, and each system can incorporate varying levels of the REM as necessary. For instance, the nominal REM is a large database containing radio environment data including geographical information, service and networks, policy, device activity, and so forth. A primitive CE may only use frequency information from the REM in the form of a binary vector, while an advanced CE may use policy and geographical information in addition to frequency information. REM-SDT enables both systems being tested to use the same input and be evaluated to the same benchmarks. Furthermore, these benchmarks are tied to both specific cases such as IEEE 802.22 and to and general case DSA environments, such that a benchmark suite is used to test CRs over a full range of situations [24].

Also in [24], the authors divide CR performance metrics into three levels: node, network, and application. Each level provides different levels of insight to the user, whether they are a regulator, standards organization, or CR developer. The authors also present the idea of using “score-cards” for comparability and standardization at each level. The authors next present a means of evaluating CRs using game theory and utility functions.

The approach presented in [30] and [24] is well suited for an IEEE 802.22-style approach, but is not necessarily useful for other CR architectures. This is primarily because the test methodology revolves around evaluating the CE and how it processes REMs. Actual RF performance is not measured.

2.6.2 Psychometric Approach.

Since radio cognition is an analogue to human cognition, it is proposed to evaluate CR using approaches developed for testing human cognition [50]. Psychometric testing, as described by Dietrich et al., utilizes item response models (IRMs) to assess CR performance [50]. IRMs have been used in psychological and educational fields, and depict how latent traits govern behaviors such that the observed behaviors can be used to estimate the levels of those traits. For example, the latent trait of a CR may be the underlying CE algorithm, while the observed trait is the CR itself. CRs and CEs are assessed by administering tasks to perform. IRMs are constructed and applied to the CR to elicit the desired response from the CR. At present, the psychometric approach has been limited to evaluation of a subset of a CE and not to an entire system.

As with REM-SDT testing, the psychometric approach as presented by [50] is geared towards an IEEE 802.22-style approach as it focuses on evaluating the CE.

2.7 Cognitive Radio Prototyping Platforms

There are a plethora of prototypes used in CR research. These prototypes range from purely software/simulations to hybrid hardware/software devices. Of interest are devices that are built using hardware, as this research is focused on testing real physical devices. Three main hardware prototyping platforms are in use by researchers today and they are presented here. All of these platforms are similar in that they enable researchers to rapidly prototype spectrum sensing, CE, and other CR components as well as offering control over the physical layer. In general, these platforms simply act as an RF interface in which the bulk of the processing is done either on an attached PC, though some contain an embedded processor.

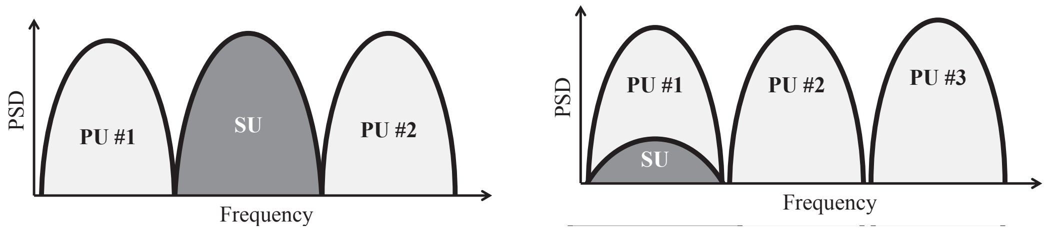
- *USRP:* The USRP is a SDR platform intended to be low-cost yet effective [51]. The USRP is ideal for experiments with large numbers of nodes. The USRP offers a flexible range of RF front-ends that enable operation from 0 to 6 GHz. There are several USRP variants available; in general, all contain an on-board FPGA for data processing. USRPs require the use of an external computing device, such as a PC, for full operation.

- *WARP:* The Wireless Open-Acess Research Platform (WARP) board [52, 53] is a high-performance SDR that enables user control over all layers of operation. The WARP is designed such that the researcher has full control over programming and operation, thus suiting a wide range of research needs. The WARP board has up to four RF interfaces, though only two are normally used. The WARP board transmits using the 2.4 or 5 GHz ISM band. WARPs can fulfill all processing and communication requirements on the board; however, it is also possible to offload processing and communication requirements, such as modulation, to an external computing device.

- *Nutaq:* Nutaq offers several diverse SDRs for use in CR applications [54]. The Nutaqs take a middle position between the WARP and the USRP: the radios are in general more capable than USRPs; however, they do not offer the full level of control and ease of use that the WARP does. Nutaq offers radios that can be used in embedded or development applications. The Nutaq radios operate on a wide range of frequencies and not solely on the 2.4 / 5 GHz ISM band. Nutaq products are used in large-scale CR testbeds.

2.8 Cognitive Radio Spectrum Paradigms

In DSA, spectrum is used in one of two ways [2]. The first is to utilize spectrum overlay. Spectrum overlay is the basis for OSA and is when SUs use spatial and temporal spectrum holes. In spectrum overlay, the SU is free to transmit so long as there is no PU transmitting at the same time. By contrast, spectrum underlay allows the SU to transmit concurrent with the PU so long as the SU does not interfere with PU. Spectrum overlay and underlay are shown in Figure 2.3.



(a) Spectrum Overlay

(b) Spectrum Underlay

Figure 2.3: Spectrum Overlay and Underlay Paradigms

Both spectrum overlay and underlay will likely be used by CRSs, and thus both should be accounted for when developing a CR test methodology.

III. Proposed Test Methodology

The proposed test methodology, hereafter referred to as CRATM, developed by this research is presented in this chapter. A methodology for evaluating CRATM is presented in the next chapter, followed by the results in the subsequent chapter.

3.1 Overview

CRATM is designed to be an over-arching framework that applies to evaluating a CRS. CRATM applies classes of benchmarks to stimulate behaviors in the SUT (e.g. a CRS). The CR response to the stimuli is then measured using performance metrics. The classes of benchmarks enable the testing to be flexible, yet allow comparisons between platforms. The radio environment is ideally controlled using emulation. A unique characteristic of CRATM is that the impact to the PU by the presence of a SU can be measured directly.

The simplest way to view CRATM is that it is similar in style to computer benchmarks such as the Standard Performance Evaluation Corporation (SPEC) suite [55]. In SPEC, multiple standard programs are used to comprehensively evaluate computer performance. Specific benchmarks are comprised of programs that are known to stress computer performance while being representative of expected workloads. The results from the benchmarks are not necessarily used to gather data on absolute real-world performance; rather, the results enable relative comparison between platforms. However, because the benchmarks are based on real-world problems, the results provide some insight into real-world behavior.

Similarly, the benchmarks proposed in CRATM are not intended to measure absolute real-world CRS performance. Rather, the benchmarks are proposed to allow an accurate relative comparison between CRSs as well as to stimulate the CRS under test to exhibit potentially undesired behavior. This is done by modeling the benchmarks on real-world radio environments.

The means of controlling the radio environment with CRATM, and thus enabling benchmarks, is to use emulation. Physical CRS test devices are used; however, the signal paths, PUs and background environment are generated virtually. The limitation to this approach is that antennas must be virtualized as well; however, the emulation environment offers flexibility in the implementation of antennas. The use of an emulated RF environment is not required for CRATM, though it is highly desired.

To test a range of devices from a single CR node to a network of CR nodes, scaling is offered by the flexible use of performance metrics. The performance metrics correspond to particular benchmark classes. For example, an evaluation of multiple CR nodes operating as a network may use network level performance metrics, while the evaluation of single CR node may utilize node level performance metrics. Scaling is important to allowing broad evaluation of CRS platforms while still allowing comparison between platforms that are similar.

The underlying assumption behind CRATM is that cognition is measured from behavior and not by direct examination of the system. This assumption is key to enabling a flexible yet specific means of evaluating CRSs.

3.2 Cognitive Radio Behavior-Based Evaluation

As stated previously, CRATM assumes that a CRS can be evaluated based on its behavior, without knowledge of the specific cognition processes taking place. CRATM treats a device as cognitive if it performs the following functions:

1. *Improves performance by responding to the environment.*
2. *Avoids causing interference to existing users.*
3. *In performing 1 and 2, implements user policy and goals*

This definition can be restated as, “does the device improve communications while not degrading others’ ability to communicate?”. The two key elements of this definition are that the device improves throughput while not causing harm to existing users. Cognition comes into play when implementing these two objectives. Therefore, cognition does not need to be measured explicitly as it can be measured through the performance of improving throughput and avoiding causing interference. Note that this definition only applies to DSA environments in which spectrum adaptation is essential. The cognition of a CRS cannot be tested if it is explicitly instructed to transmit on open-frequencies from an external decision making authority. In other words, a CRS should not be tested using this definition as if it were a licensed primary user.

CRATM treats cognition as occurring on a continuum as opposed to occurring in discrete steps. Researchers can focus on evaluating CRS performance without needing to justify if a device is in fact cognitive, or to what relative cognitive level it is. With CRATM, cognition becomes apparent when observing performance. A more intelligent cognitive CRS should perform better than a less intelligent CRS. The actual cognition levels of the CRSs is secondary to performance. Stated another way, *cognitive ability only matters if it is manifested in performance*. Otherwise, cognition is a theoretical abstraction with little relevance to the real world. This distinction is important as it enables the evaluation of CRSs without being tied down to specific definitions of cognition. Furthermore, this distinction means that CRS devices can evolve from current systems without requiring a discrete jump to CRS capability. Treating cognition as a continuum is a feature of behavior-based testing.

Behavior-based testing is similar to the psychometric testing presented in [50]. However, where psychometric testing applies stimuli based on models of the underlying cognitive behavior, behavior-based testing applies stimuli based on the expected real-world environment. Both measure the observed behavior of the SUT and use that information to extrapolate the performance of the underlying cognitive processes. If behavior-based testing is proven to work, it negates the need to require specific definitions for CR.

3.3 Performance Metrics

Based on the definitions found in Section 3.2, the behavioral based performance metrics available for use fall into two categories: improving performance and avoiding causing interference. Within these categories, the applicable performance metric depends on the class of benchmark as presented in Section 3.4.

The baseline performance metrics for use in measuring both improving performance and avoiding causing interference are to measure BER and/or throughput. For instance, the SU BER may be used as a metric of improving performance while PU BER may be used as a metric of avoiding causing interference. Both SU and PU performance may be baselined by collecting data without the presence of the other user. The basic equations for BER and throughput are:

$$BER = \frac{bit\ errors}{total\ bits\ transmitted} \tag{3.1}$$

$$throughput = \frac{bits\ successfully\ transmitted}{time\ to\ transmit\ bits} \tag{3.2}$$

Other performance metrics that may be used for evaluating both SU and PU performance are packet loss/PER, end-to-end delay, delay-variation, QoS, or bandwidth.

It is important to note that specific developers may have need for a specific performance metric besides the BER and throughput of the PU and SU. Just as with computer benchmarks, the most important benchmark is one that encompasses the envisioned end-use of the platform. Likewise, the most important performance metrics (and benchmarks) for a CRS derive from the envisioned end use of the CRS. If power conservation is paramount, a power metric can be added and used. CRATM provides an overarching framework for generic CRS testing; future testers can modify it as they see fit for specific needs.

3.4 Benchmarks

Benchmarks allow CRS testing to be flexible, accurate, and repeatable. CRS behavior is dependent on the test scenario. Similarly, the capabilities of the CRS limit applicable benchmarks. For example, the benchmark should stimulate the capabilities of a CRS to sense the spectrum. However, the benchmark should not be used to evaluate a CRS if the signal of interest to detect lies outside of the capabilities of the CRS receiver test device.

To allow for AI learning to occur, benchmark scenarios should be programmed to be time-driven as opposed to task-driven. In other words, the scenarios are not to be executed as fast as possible by the CRS; instead, the scenarios operate on a fixed timeframe and the performance metrics of the CRS are measured at predetermined time intervals. For example, a CRN may be evaluated at one minute into a scenario, five minutes in, and thirty minutes in. This provides time for the CRN to adapt to the environment. A non-adaptive system may outperform the adaptive system initially, but the longer timespan provides an opportunity for adaptation to occur. In the future, extended testing could feature scenarios that are on the scope of months to years. For instance, a cognitive engine may “learn” that certain days of the year have different usage profiles and adapt accordingly.

Specific classes of benchmarks enable a wide variety of CRS architectures to be evaluated. For example, an IEEE 802.22 CRS architecture should be evaluated under benchmarks that pertain to its architecture. Various architectures are executed under the same benchmark to enable direct comparison. If an architecture cannot be executed under a specific benchmark, then it cannot be directly compared with other architectures based on that benchmark. This prevents incorrect comparative conclusions from being drawn.

The benchmarks are split into eleven characteristics, which combined, form a unique benchmark. The top four characteristics can be combined to form a *benchmark class*, or a grouping of similar benchmarks. Multiple benchmark classes are combined into a *benchmark suite*.

- *Characteristic 1, Cognitive Radio Type:* Cognitive radio type describes whether the SUT is a specific CR device or a CRN. In the case of a CR device, the SUT is just the CR device. All other components are either virtual or physical, but their behavior is well known and characterized. In the case of a CRN, the SUT is comprised of a network of CR devices. All devices on the network are considered part of the SUT, unless otherwise explicitly stated for the particular benchmark.

- *Characteristic 1a, Number of Nodes:* If the SUT is a CRN, the number of nodes must be specified. To simplify benchmark classification, the number of nodes may be set to common numbers of interest such as 2, 4, 10, or 100 nodes.

- *Characteristic 2, Topology:* The topology drives the type of scenario implemented. The topologies are distributed, centralized, and combined. For example, an IEEE 802.22 CR should not be evaluated in a distributed environment as it not is designed for that environment. Distributed environments imply no centralized base station, while centralized environments do. The advantage in offering this distinction is that it provides flexibility in the type of architectures tested. A CRN may be tested with a virtual base station instead of requiring a physical base station. Making the topology a key characteristic enables specificity when comparing architectures.

- *Characteristic 3, Radio Frequency Environment:* The RF environment describes the emulated RF environment. The options include, but are not limited to, urban, rural, suburban, home, and battlefield environments. The RF environment may also be a combination of these environments.

- *Characteristic 4, Test Band:* The test band lays out the overall frequency range eligible to be reached by the SUT. The emulated RF environment is generated such that PUs appear in this band as specified by the scenario description.

- *Characteristic 5, Motion:* The SUT node(s) may be either static or moving. If moving, the node(s) follow a preprogrammed route that is defined for the benchmark. Otherwise, the node(s) remain fixed at predetermined locations.

- *Characteristic 6, Performance Metrics:* The performance metrics of interest are defined under this characteristic. The metrics are drawn from those defined in Section 3.3.

- *Characteristic 7, Timeframe and Sampling Intervals:* The overall test duration and sampling intervals are defined in this characteristic.

- *Characteristic 8, Geographical Information:* The locations of each node plus virtualized units are described.

- *Characteristic 9, Maximum Power Level:* The maximum power level provides an upper bound on the CR power levels so that performance metrics are not artificially boosted. It is specific to each scenario. In practice, the maximum power levels may be drawn from FCC regulations.

- *Characteristic 10, Primary User Profile(s):* The characteristics of the PUs are specified here. Parameters include frequency, power, location, duty cycle, and waveform, among others. The PU profiles form the core workload of the benchmark.

- *Characteristic 11, Scenario Description:* The final characteristic is a detailed scenario description. Any information not already detailed out above must be categorized here. For example, if a priori spectrum knowledge is assumed, this knowledge should be specified here. Then, that information is passed to the SUT when loading the benchmark.

An example benchmark based on the benchmark characteristics is shown in Table 3.1.

Table 3.1: Example CRATM Benchmark

	Characteristic	Description
1	CR Type	CRN
1a	Number of Nodes	4
2	Topology	Distributed
3	RF Environment	Rural
4	Test Band	2.4 - 2.5 GHz
5	Motion	Static
6	Performance Metrics	SU - BER, PU - Throughput
7a	Timeframe	15 minutes
7b	Sampling Interval	1, 5, 15 minutes
8	Geographical Information	4x4 node centered at 90, 90, 0
9	Maximum Power Level	-40 dBm
10	Primary User Profile	802.11n Wireless Router connected to 4 laptop devices with nominal packet traffic
11	Scenario Description	Ad-hoc test. No apriori signal knowledge. BPSK PU at 2.462 GHz at -50 dBm.

3.5 Means of Testing

The means of testing for CRATM is to utilize an emulated RF environment in which all activity from CR RF output to CR RF input is controlled via the emulation environment. This necessitates using virtualized antennas. Virtual PUs are created within the emulated environment, negating the need for physical PU devices (though they are allowed, if necessary). This enables rapid changeover between benchmarks. All elements, except for the physical SUT devices, are software defined. CRATM may be implemented without using an emulated RF environment; however, flexibility in running different benchmarks is lost.

3.6 Test Framework

The proposed CRATM test framework is as follows. An emulated radio environment is created which consists of PUs with various waveforms that communicate dynamically over the physical layer. These virtual PUs are designed to mimic either expected real-world conditions or to show specific test conditions. This emulated radio environment of virtual PUs acts as a repeatable benchmark so that different CRS radios are compared by running the same benchmark. Physical CRs, shielded from each other, are connected to the test system and evaluated by collecting performance metrics. Full spectrum knowledge is known and is controlled since the emulation happens at the physical layer. Furthermore, since the physical layer emulation is accurate, realistic higher-layer operation of both virtual PUs and the physical test devices is expected. Finally, since the emulated environment is repeatable, test results are comparative between different CRSs. The physical test devices are treated as “black boxes” in that the behavior of the device is measured via device outputs as opposed to monitoring and measuring the actual internal processes of the device. The emulated environment allows a rapid turnaround of the test environment as compared to field testing, while offering more realism than laboratory/simulation testing.

3.7 Comparison to the Literature

CRATM is tested to see if it contains characteristics and attributes of proposed test methodologies as found in the literature. If CRATM does not contain those attributes, then justification is required to show that CRATM is sufficient. In other words, if there are any novel components to CRATM that are not found in published data, they should be justified through further research.

Data for this literature comparison are the characteristics and attributes of proposed test methodologies as collected from published peer-reviewed journals and articles. Secondary data sources are conference workshop proceedings and theses. Since the field of CR is relatively new, data is not excluded unless it is specifically refuted in subsequent published work or is of marginal relevance. Conversely, data is given greater weight if it is cited or repeated by subsequent work.

3.7.1 Key Characteristics.

The collected data is categorized by the following characteristics:

- *Means of Testing:* Since the scope of this research is for testing physical CR devices, the means of testing includes laboratory, emulation, and real-world. Laboratory testing may include testing inside of a controlled environment such as an anechoic chamber. Emulation testing includes testing in which the RF path is emulated. Real-world testing includes testing of devices in uncontrolled and controlled RF environments. The difference between real-world and laboratory testing is one of degree. Laboratory testing allows fine-grain control of the RF environment and near-ideal operation of equipment while real-world testing offers coarse-grain control of the RF environment (if at all) and non-ideal equipment operation.

- *Performance Metrics:* Performance metrics are used to evaluate CR performance at both the node and network levels.

- *Attributes:* The attributes of the data are elements such as reproducibility and accuracy.
- *System Under Test:* The SUT is used to define the test device. These attributes are derived from the various CR architectures.
- *Cognition Evaluation Methodology:* The overall approach and strategy to evaluating cognitive ability.

3.7.2 Results.

The literature was surveyed and the results are summarized in Table 3.2. All references in the bibliography, except hardware and technical specification references, were considered when compiling the categories and attributes in the table.

The following elements were found in the aggregate data that are not found in CRATM: open-air testing, directly measuring spectrum sensing, and directly measuring cognitive ability. Also, neither cognitive methodology, psychometric or REM-SDT, in the literature is utilized in CRATM. Open-air testing can be viewed as an either/or requirement; that is, if emulation meets the requirement, then open-air testing is not required. Thus, this does not need to be justified. However, the remaining discrepancies do need to be justified.

CRATM does not directly measure spectrum sensing or cognitive ability. As argued in Section 3.2, behavior-based testing is proposed to take the place of measuring spectrum sensing and cognitive ability directly. Behavior-based testing also takes the place of psychometric testing and REM-SDT testing. A methodology to evaluate the hypothesis that behavior-based testing measures cognitive ability is presented in the next chapter.

3.8 Summary

CRATM offers a means to evaluate the performance of a CRS, acting as a SU, in a controlled yet realistic environment. Workloads offered to the test device are classified into

Table 3.2: Comparison Between CRATM and Published CR Test Methodology Attributes

Category	Attribute	Does CRATM Contain?
Means of Testing	Emulation	Y
	Open-Air	N
Performance Metrics	Avoid Causing Interference	Y
	Spectrum Sensing	N
	Cognitive Ability	N
	Network/Device Performance	Y
Attributes	Reproducability	Y
	Accuracy	Y
	Uses Benchmarks	Y
	Test Overlay	Y
	Test Underlay	Y
SUT Characteristics	Single Device	Y
	Heterogeneous Network	Y
	Homogeneous Network	Y
	Centralized Architecture	Y
	Distributed Architecture	Y
Cognitive Methodology	Psychometric	N
	REM-SDT	N

benchmarks in a manner similar to SPEC benchmarks used with modern computers. CRSs are evaluated using benchmarks so that the results are comparative across multiple types of systems. Ideally, the PU is created in an emulated RF environment to allow greater flexibility in testing, though it is not necessary. The key feature that CRATM offers is the capability to measure SU cognition by measuring its behavior and performance. CRATM assumes that cognition is defined as the ability to improve SU performance while not negatively impacting the PU, all while implementing user goals and policy. However, this is yet to be proven. Furthermore, CRATM is unique in that it proposes to measure the impact to the PU directly by measuring PU performance.

IV. Methodology

4.1 Problem Background

The key systemic problem facing CR developers and researchers is how to validate CR performance such that its real-world behavior is testable and knowable outside of the real-world. CR performance, by definition, is subject to change and evolve over time. A derivative problem facing CR researches is how to test real hardware CRSs in order to evaluate end-to-end CR performance. The means of testing the CR must be repeatable, accurate, and meaningful so that results are extendable to real-world applications. CRATM does not address the former problem; however, it does address the latter.

4.1.1 Goals and Hypotheses.

The goal of this research is to develop a test framework such that given a hardware CRS to test, the CRS can be successfully evaluated so that the evaluation results are comparative to other CRSs (CRSs may be comprised of either a single CR node or a network of CR nodes as defined in 2.1). CRATM is proposed to fulfill this research goal. As there is currently no standardized method of testing CR systems, CRATM serves as a proof of concept. CRATM cannot be evaluated in its entirety by measuring its utility against established test methodologies, nor can it be evaluated by exhaustively gathering data on all elements of CRATM due to a shortage of CR prototypes. Instead, the novel component of CRATM, that SU cognition may be determined by measuring behavior, is evaluated. This behavior-based testing component is critical to the success or failure of CRATM. Thus, if it shown to be false, CRATM is not sufficient in its current form. However, if it is not shown to be false, then further development and testing of CRATM is warranted. The formal hypothesis to be evaluated is as follows:

- **Hypothesis**: *The cognition of cognitive radio systems can be successfully tested by measuring cognitive radio behavior of the entire system, without knowledge of the*

> *behavior of the underlying cognitive components*. The null hypothesis is that the cognition of cognitive radio systems can only be evaluated by direct measurement and analysis of the cognitive components.

The hypothesis tests the key underlying assumption of CRATM. This underlying assumption is that a CR is defined to be cognitive if it performs three specific tasks in a DSA environment: 1) improves performance, 2) avoids causing interference to existing users, and 3) implements user policy and goals when performing 1 and 2. If this assumption is true, then it is possible to measure CRS behavior without needing to directly evaluate the cognition of specific components. The hypothesis is tested by comparing CR prototypes with non-CR prototypes, with the expectation that CRs with the full OODA cycle (sense the spectrum, react to it, and change waveforms as required) will perform better than devices without the full OODA cycle.

4.1.2 Approach.

The first step in this research is to develop a CR prototype to serve as a CRS SUT. The CR prototype performs all four actions of the OODA cycle as described in Section 2.2. After basic CR device performance has been validated, the hypothesis is tested as follows.

A PU network is implemented alongside a SU network. The PU is used to create the environment sensed by the SU. The SU has a CE with spectrum sensing capability. The PU and SU are setup to operate simultaneously, and performance metrics are collected for both. Various desired waveforms (user goals) are passed to the SU to serve as a baseline when the CE is off. These experiments are repeated with the SU CE on. The results should show that the PU and SU performance improves when the SU CE is on as compared to when it is off. Furthermore, the PU and SU performance should improve, or stay constant, as greater levels of cognitive adaptability are added to the SU. Cognitive adaptability for this experiment is increased waveform choices and bandwidth for the SU. It is assumed for this experiment that greater adaptability in waveform and bandwidth is correlated with greater cognitive ability.

The PU implements several different waveforms in order to provide diversity in spectrum opportunities. For example, one experiment workload may be a PU that uses a constant wide-band transmission, while the next experiment workload may be a PU that uses a frequency hopping narrow-band transmission. A range of PU waveforms are used so that the SU response to the PU created environment may be adequate characterized.

All experimental configurations are repeated for the case when the opposing network is off. In other words, the experiments are repeated with the SU off and PU on and then with the PU off and the SU on. These experiments serve as a control.

4.2 System Boundaries

The SUT, shown in Figure 4.1, includes a wireless radio node pair as part of a small communications network. One node is a designated transmitter while the other node is the receiver. For this experiment, the SUT is limited to WARP nodes [52] running WARPLab [56]. In WARPLab, the radio node only serves as a transmitter or receiver and all modulation or demodulation of the waveform takes place on board a PC connected via Ethernet.

The component under test (CUT) is the CE, or the ability of the radio to execute the OODA cycle. For the baseline non-cognitive system, the CUT is the unmodified WARP design, which has no observe, orient, or decision part of the OODA cognition cycle. For the cognitive system, the WARP design is modified to use an energy-detection spectrum sensing algorithm to implement the full OODA cycle. Both the cognitive and non-cognitive configurations use the same hardware configuration. The software is identical with the exception of an additional function, the CE, added to the CR to sense the environment and cognitively choose a waveform.

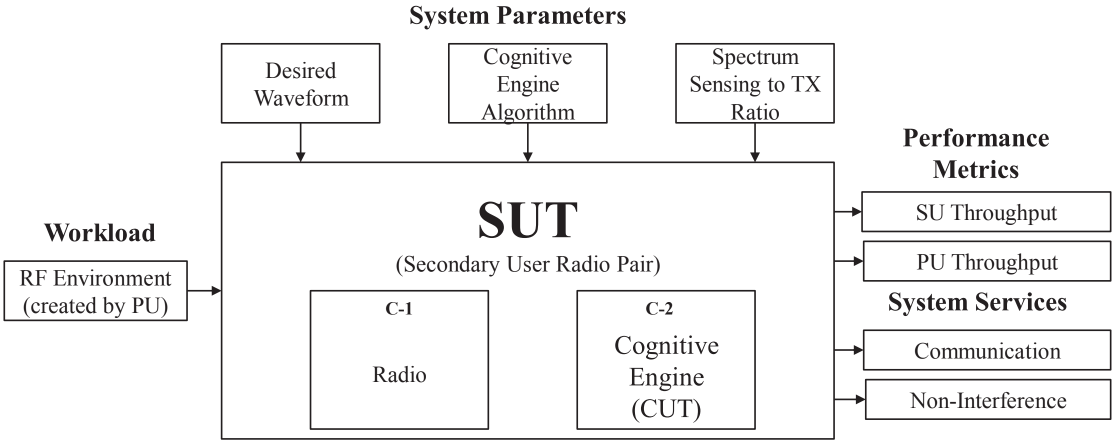


Figure 4.1: DSA Testbench SUT

The SUT is limited to transmission on a single 2.4 GHz ISM band channel. Within this 20 MHz channel, the spectrum is broken down into 128 sub-channels, or frequency bins. The SUT uses only 64 of those bins, for a bandwidth of 10 MHz.

For the purposes of this experiment, the SUT is also referred to as the SU. The scope of the SUT is limited to the cognitive engine presented in Section 4.6 using WARP hardware.

4.3 System Services

The SUT offers two services: communication and non-interference. These services are derived from the hypothesis. Communication applies to the SU, while non-intererence applies to the impact of the SU on the PU.

Communication (or throughput) is defined as the successful transmission of data from one node and its successful receipt by other nodes. Due to the scope of the SUT, only physical layer data communication is tested and the data of interest being transmitted is only bits. The outcomes of this service are communication success, partial success or failure. Success occurs when all data is successfully transmitted between the transmitter and receiver. Partial success occurs when data is transmitted between the transmitter and receiver but is degraded due to interference from the RF environment. Failure occurs when data is not transmitted due to interference.

The non-interference service is the ability of the SU nodes to not interfere with a PU. Interference occurs when there is simultaneous transmission of a SU and a PU on the same frequency bin. There are two types of interference: closed channel interference and open channel interference. In closed channel interference, a channel is occupied by a PU prior to the transmission of interference by a SU. Successful avoidance of closed channel interference means that the SU does not transmit during any time when the PU is transmitting for a channel. Failure is when the SU does transmit on-top of an existing PU. In open channel interference, a channel is open and used by a SU, but is then occupied by a PU. The SU will inadvertently cause interference in this case, so the goal is to minimize the time it takes for the SU to detect that it is interfering with a PU and abandon the channel.

4.4 Workload

The workload for the SUT is PU activity. The PU activity creates the radio environment that enables the SUT to react and adapt to in accordance with its CE. The workload is generated by two WARP radio nodes running WARPLab. One PU node is the transmitter and the other PU node is the receiver. There is only one PU pair (transmitter and receiver) for this experiment. The PU transmitter is used to create a total of five environments.

The SU may use spectrum holes created by the PU, as shown in Figure 4.2. In this figure, the 10 MHz spectrum bandwidth is broken up into 64 frequency bins that the SU may use to transmit on. Spectrum holes are indicated as light areas on the REM plot. An arbitrary spectrum sensor threshold line is also shown on the chart to signify how the SU determines what frequencies are occupied or not.

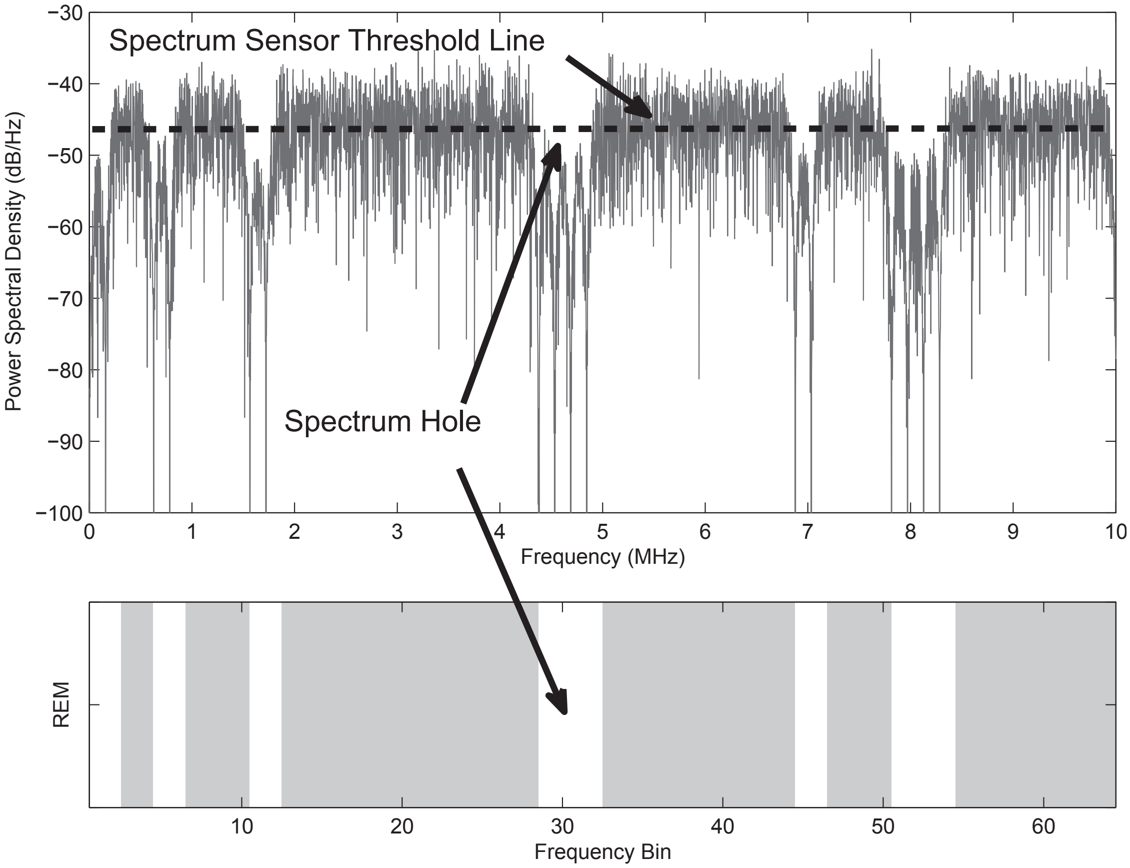


Figure 4.2: PU Created Spectrum Holes

4.5 Performance Metrics

The performance metrics are derived from the system services and are chosen in accordance with the metrics found in Section 3.3. For this experiment, both SU communication performance and PU non-interference are measured using Equation 4.1.

$$Throughput = \frac{Number\ of\ Attempts\ to\ Receive\ *\ Bits\ Correctly\ Demodulated\ *\ WARP\ Buffer\ Size}{WARP\ Clock\ Frequency} \tag{4.1}$$

The units of Equation 4.1 are in bits per second (bps). The throughput of the SU and the throughput of the PU are measured independently.

A secondary performance metric used in this experiment is BER, as shown in Equation 4.2. All modulation schemes used in this research use binary modulation. Therefore, BER is used as opposed to the more generic symbol error rate (SER).

$$BER = \frac{Number\ of\ Bits\ in\ Error}{Total\ Number\ of\ Bits\ Transmitted} \tag{4.2}$$

BER is used to corroborate throughput data. For greater SU cognition ability, there should be an increase in throughput for both the PU and SU. If the throughput cannot increase due to a bandwidth limitation, then the BER should be seen to decrease or stay constant. BER is not used as the primary metric as it does not fully characterize performance. For example, if the CE user goal is to minimize BER, then a minimal bandwidth waveform will be chosen. However, the throughput will be drastically reduced, limiting its desirability for use in real-world communications.

4.6 System Parameters

The system parameters are characteristics that affect the performance of the SUT. Initial pilot experiments reduced the number of key system parameters down to the following list.

- *Desired Waveform:* The desired radio waveform type impacts the communications capabilities of the SUT as well as the ability of the SUT to avoid interfering with PUs. Three specific waveforms are used in this experiment: frequency shift keying (FSK), orthogonal frequency division multiplexing (OFDM), and dis-contiguous OFDM (d-OFDM). All waveforms are strictly physical layer only and no MAC is used. The bandwidth of the waveform is directly correlated to a number of frequency bins. For example, an OFDM waveform using 16 bins (contiguous) will have an overall bandwidth of $16 * \frac{F_s}{256} = 2.5$ MHz, where F_s is the sampling frequency of 40 MHz. In the case that the CE is off, the desired waveform is in fact the waveform that is used in the experiment. In the case that the cognitive engine is on, the cognitive engine will try to maximize its performance by selecting the desired waveform; however, due to environmental constraints, a lower performing waveform may be utilized. FSK uses binary modulation, and the OFDM waveforms use binary phase-shift keying modulation.

- *Transmission Power:* The radio node transmission power inversely affects communications performance and non-interference. In general, increasing transmission power to improve communications performance will negatively impact the non-interference capability. For this research, the transmission power is kept constant and is not varied by the cognitive engine. The transmit gain on the MAX2829 transceiver on the WARP is set to 0. The receive gain on the MAX2829 transceiver is set to 15.

- *Cognitive Engine Algorithm:* The CE allows a radio node to optimize communications performance and non-interference. The cognitive engine for this experiment is a simple bandwidth optimizing algorithm. Based on the sensed environment, the algorithm chooses the best waveform that maximizes SU throughput while avoiding transmitting on frequency bins occupied by the PU. For instance, the user may desire a OFDM waveform with 16 bins of bandwidth. The sensed spectrum, however, reveals that there are only 8 contiguous frequency bins available. As 8 bins is the transition between FSK and OFDM, the algorithm prescribes downgrading the waveform to FSK (as it has superior throughput for bandwidth of 8 bins or less). The algorithm is presented in Figure 4.3.

It is important to note that this CE algorithm is intended to represent the baseline functionality of a CE and not necessarily to be a fully functional, state of the art CE. In Mitola's cognition levels [6], this CE algorithm would likely lie around Cognitive Level 1 or 2. For reference, Cognitive Level 1 is Goal-Driven, meaning that the CR chooses a waveform according to a goal. Cognitive Level 2 is Context Awareness, meaning that the CR has knowledge of what the user is trying to do.

Based on pilot experiments, the system is most sensitive to the CE status (on or off) as well as to the choice of desired radio waveform.

4.7 Workload Parameters

The workload parameters are workload characteristics that affect the performance of the SUT.

- *Power and Position:* The transmission power, as well as relative position, of each PU impacts the perceived spectrum of each individual SUT node. To simplify the experiment, the PU transmission power remains constant and equal for all PUs. Furthermore, the position of the PUs to the SUs is fixed on a 2 × 2 rectangular grid with dimensions 18 inches by 36 inches. This orientation was chosen primarily for a configuration that can be replicated in the DYSE, but also due to equipment constraints and to minimize the proximity of the antennas to the radios. It should be noted that pilot experiments did not show an appreciable difference for various antenna configurations. The PU transmission power is set equal to the SU transmission power of 0 dB gain for the MAX2829 chip on the WARP radios.

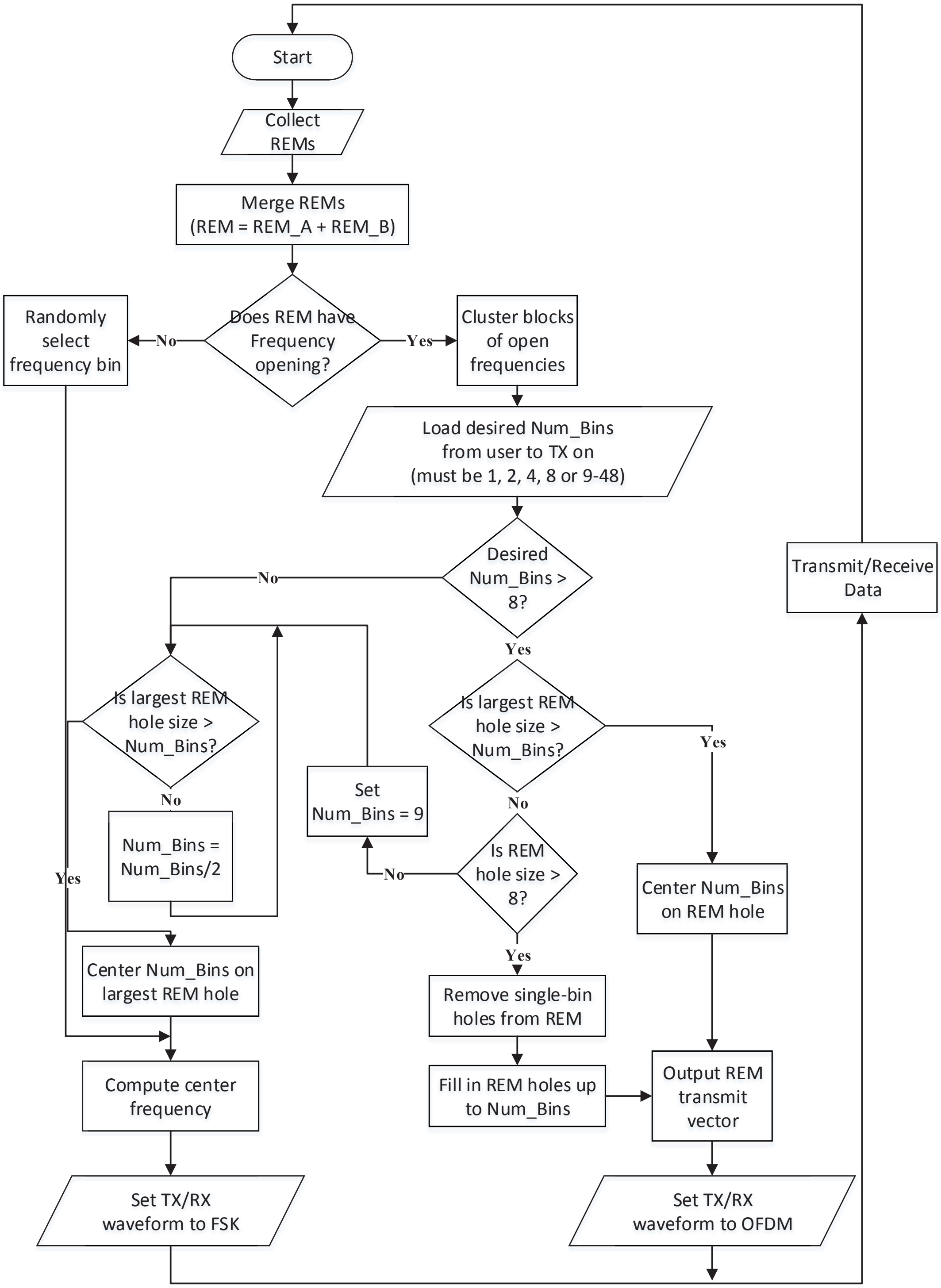


Figure 4.3: Cognitive Engine Algorithm Flowchart

- *Waveform:* The PU waveform is either a wideband, narrowband, or dis-contiguous (non-contiguous) signal to create a range of RF environments. These specific waveforms are created using FSK, OFDM, or d-OFDM. The overall waveform type stays constant within an experiment, though the frequencies it transmits on may change.

- *Transmission Duration:* For each experiment, the PU is set to transmit for a minimum duration. The minimum transmission duration acts as a discrete step size. For instance, the total experiment time may be 60 seconds and the minimum transmission duration is 1 second. If the PU waveform utilizes frequency hopping, then the frequency will change once every second.

  The total time to transmit one WARPLab buffer over the air is approximately 410 μs. Data processing was found to take two to three orders of magnitude longer. Therefore, to minimize the total experimental time and to drastically reduce the number of experiments required, the PU transmitter is set to continuously transmit. The transmit buffer is updated for each new BER calculation cycle and the transmit waveform is updated at each transmission duration interval.

- *External RF Environment:* The RF environment is either emulated or over-the-air. For RF environment emulation, testing takes place in the DYSE. All antennas are assumed to be dipole antennas. For over-the-air, testing takes place on a lab bench. Antennas are located on a plane and are separated by at least 3 feet from the nearest WARP board. Antennas are dipole antennas.

To simplify the experiment, the workload parameters are combined into a single workload factor, which is referred to as the Environment. This is discussed further in the following section.

4.8 Factors

Factors are the parameters that are varied during the experiment and are derived from the parameters. The factors for this experiment are summarized in Table 4.1.

Table 4.1: SUT Factors

Type	Factor	Levels
System	Desired Waveform	FSK-1 Bin BW, FSK-8 Bins BW OFDM-16 Bins BW, OFDM-48 Bins BW
System	Cognitive Engine	Off, On
Workload	Environment	1, 2, 3, 4, 5
Workload	External RF Environment	Lab, Emulated

4.8.1 System Factors.

- *Desired Waveform:* Four different waveforms levels are available for this experiment. These are a FSK waveform with 1 bin of bandwidth, a FSK waveform with 8 bins of bandwidth (BW), a contiguous OFDM waveform with 16 bins of bandwidth, and a contiguous OFDM waveform with 48 bins of bandwidth. For readability, waveforms are referred to as the modulation-bandwidth as in FSK-8. These four levels serve as maximum performance boundaries for when the CE is on. For instance, the CE when on may choose a lower performing waveform for use than the desired waveform. The performance without interference, however, will not be any greater than the desired waveform with the CE off. OFDM-48 is the highest performing waveform. Discontiguous OFDM has lower performance than contiguous OFDM.

- *Cognitive Engine:* The SU operates with the CE on or off. If off, the SU randomly generates a new waveform (in accordance with the desired/specified waveform) in place of each CE cycle. For example, consider a case where the desired waveform is FSK with 1 bin bandwidth. With the CE on, the CE may be programmed to take 0.25 seconds per sense-process-communicate (i.e. OODA) cycle. With the CE off, a new frequency bin to transmit on is selected every 0.25 seconds.

4.8.2 Workload Factors.

The workload parameters, except external RF environment, are combined to form a single workload factor, referred to as the environment.

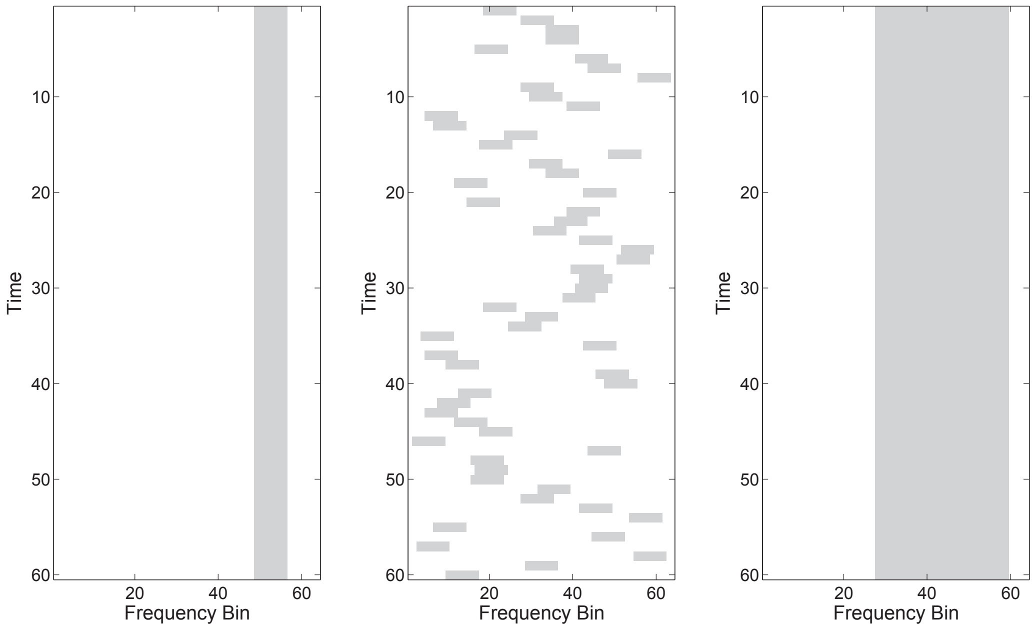
- *Environment:* Five distinct Environments are created for this experiment: a constant FSK tone with 8 bins BW, a hopping FSK tone with 8 bins BW, a contiguous OFDM transmission with 32 bins BW that has a constant bandwidth, a non-contiguous OFDM transmission with 48 bins BW that changes waveform every 5 transmission cycles, and a randomized non-contiguous OFDM waveform with 16 bins of bandwidth. These environments are depicted graphically in Figure 4.4.

  Note that the frequency bins utilized for transmission are randomly generated for each experiment. The transmission duration is fixed at 1 second and the total experiment time is fixed at 60 seconds.

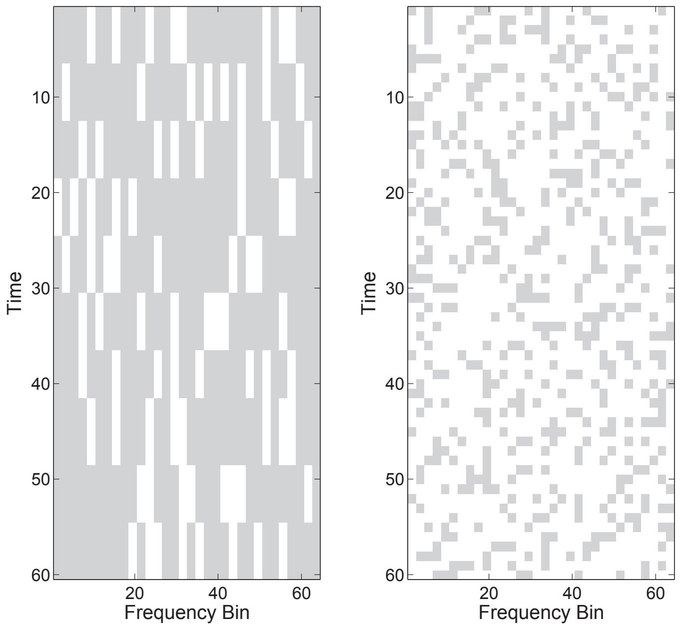
- *External RF Environment:* The environment is either the lab environment or the emulated environment of the DYSE.

4.9 Evaluation Technique

The experiment is run on four WARP v2 hardware platforms. Two radios are designated for use as the PU pair, while the others are designated for use as the SU pair. The antennas and radios are placed such that each pair communicates across the diagonal as shown in Figure 4.5. Each radio pair is controlled by one HP Laptop running MATLAB R2012b and WARPLab v7.3. Both laptops are identical images operating on 64-bit Windows 7. The laptops are HP Compaq 8510p with Intel Centrino processors. Both



(a) Environment 1 (b) Environment 2 (c) Environment 3



(d) Environment 4 (e) Environment 5

Figure 4.4: PU Created Workload Environments. Clockwise from top-left: FSK-1, FSK-8, OFDM-32 Contiguous, OFDM-48 Non-contiguous, and OFDM-16 Randomized. Note that dark areas indicates frequencies that the PU transmits on.

laptops and the radios are connected via Ethernet to a Cisco Gb switch. The RF channel is approximately white Gaussian noise.

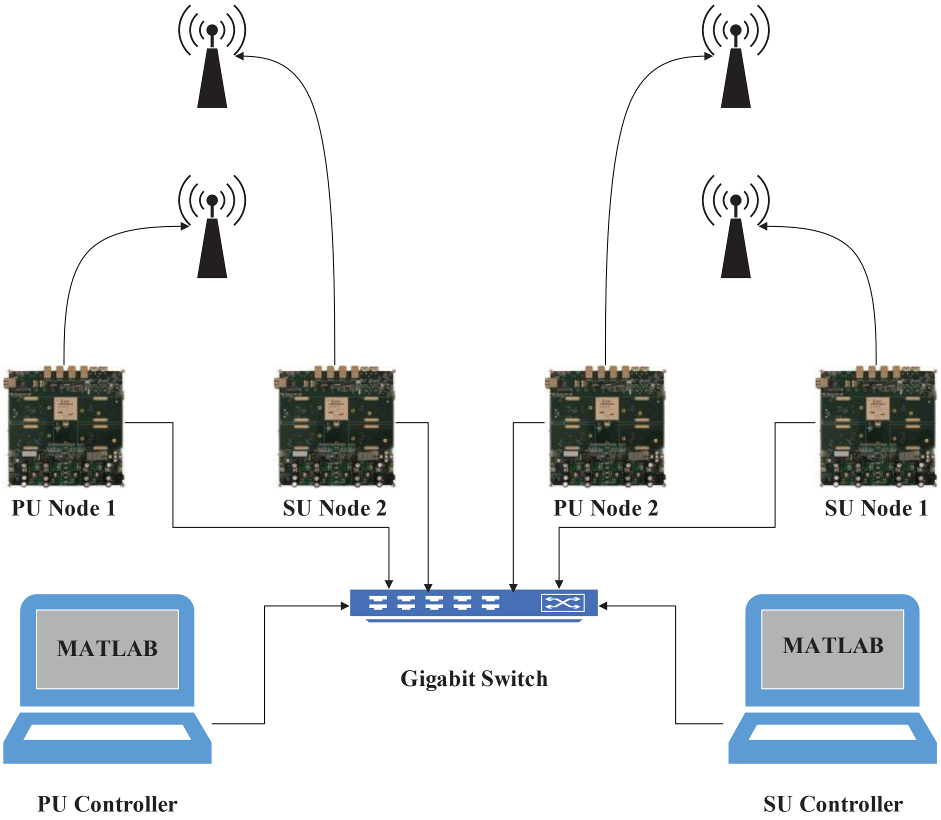


Figure 4.5: Experimental Setup

The PU laptop acts as a controller for the experiment. A list of experimental configurations are loaded onto the PU laptop. Then, the order of experiments is randomly sorted. The PU laptop shares the current experiment configuration with the SU laptop. Prior to each experiment, the spectrum sensing thresholds for the SU are calibrated in order to minimize environmental effects. Calibration is performed by having the PU transmit a known waveform and then lowering the threshold level on the SU spectrum sensor until the measured REM occupies at least all of the bins for the PU transmitted waveform. Then, the PU transmits a second known waveform and the threshold level is checked for validity. If there are a greater number of REM bins occupied by interference than by the desired PU signal, the calibration process is repeated.

Following calibration, the laptops synchronize and then the PU laptop sends a trigger signal to the SU laptop, upon which both laptops begin executing the experiment for a specified duration. Once the time for each experiment has expired, output parameters and performance metrics are saved and the process begins again.

Each experiment is programmed to be 60 seconds in length. Pilot experiments evaluated experimental durations of 10, 30, 60, and 120 seconds. These experiments showed that 60 seconds was sufficient to minimize variance of the results.

The results of this experiment serve as a proof of concept. However, the results are also validated since BER is calculated alongside throughput. Since the transmission powers and ambient noise are known via use of a spectrum analyzer, the performance of the SU and PU pairs may be validated against known data. If the computed BER is outside expected bounds, then the experimental results may not be valid.

The experimental setup as used with the DYSE is identical except that rather than communicating over the air, the WARP boards are connected via SMA cable to the DYSE RF inputs. Additionally, the SU transmit node is replicated in software in the DYSE.

4.10 Experimental Design

A full factorial experimental design is used, excluding the RF environment factor. That is, there are 4 x 2 x 5 = 40 experimental trials. Additionally, baseline performance data is collected for each factor, which adds another 8 SU factor levels + 5 PU factor levels = 13 trials. Therefore, there are a total of 53 trials conducted for one experiment set. A minimum of 20 repetitions are accomplished in order to sufficiently reduce 95% confidence intervals.

The results are assessed by utilizing both a two-sided *t*-test as well as by using a scoring system. In both cases, the results should show that for all PU workload environments, the SU configuration with the cognitive engine on improves both PU and SU throughput when compared to when the cognitive engine is off. Furthermore, the results should show that PU and SU throughput is improved as greater levels of cognition are added (i.e. using higher performing waveforms). If these results are found, then the hypothesis will have been shown to not be false. This research is not intended to prove that the hypothesis is true; rather, the research is intended to see if the hypothesis is false.

4.11 Methodology Summary

The hypothesis is designed to test whether or not a cognitive radio can be evaluated by measuring its behavior. The SUT, which is the SU, is a WARP radio pair that has cognitive capability. The SU is evaluated against a PU workload, which drives the SU. The throughput of both the SU and PU are gathered to assess communications throughput of the SU and non-interference to the PU.

V. Results and Analysis

The results of the experiments described in the previous chapter are discussed here. An analysis demonstrating that the collected data is valid is presented in Appendix A.

5.1 Evaluating Behavior Based Cognitive Radio Testing

The hypothesis is designed to test whether or not a cognitive radio can be evaluated by measuring its behavior. The results of evaluating this hypothesis are presented in this section. The expected results are that the SU throughput is increased when the CE is on relative to when the CE is off, that the PU throughput is increased when the SU CE is on relative to when the SU CE is off, and that greater CE functionality will result in greater throughput for both the SU and PU. CE functionality is assumed to be based on the ability of the SU to adapt. Since all SU configurations with the CE on use the same underlying algorithm, the adaptation component comes from the choice of the waveform. As presented in Section 4.8.1, OFDM waveforms are considered more adaptive and capable than FSK waveforms, meaning that the OFDM waveforms are considered more cognitively useful for the purposes of this experiment. The order of cognition goes from FSK-1, FSK-8, OFDM-16, OFDM-48.

The experiments were collected in the ACRO lab. All experimental data was collected using over-the-air RF transmissions. A total of 25 to 30 data points were collected for each experimental configuration. An attempt was made to run the experiments in the dynamic spectrum environment emulator (DYSE); however, this attempt did not produce useful results.

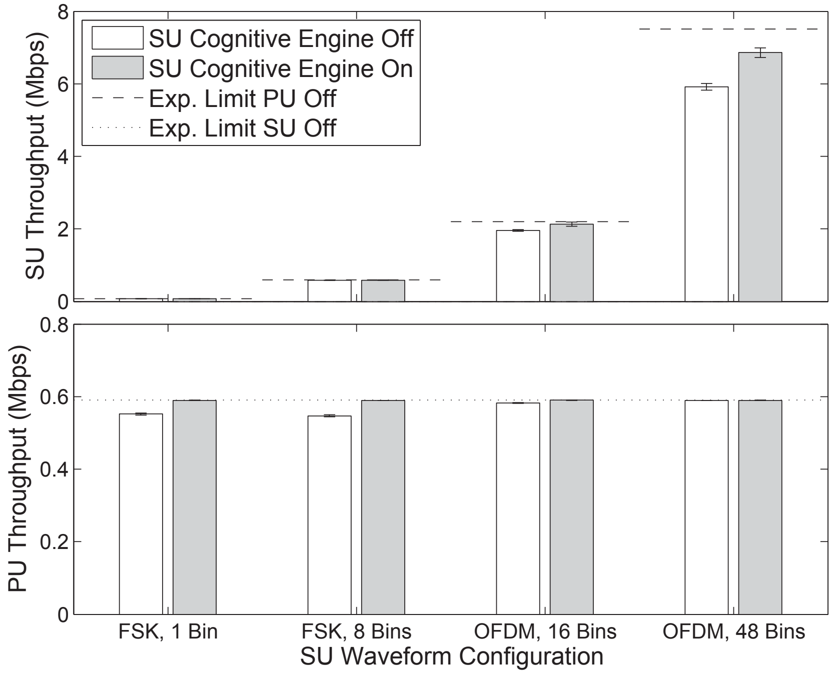
The throughput of testing each PU environment is shown in Figures 5.1, 5.2, and 5.3. The charts are interpreted as follows. The top subplot shows the SU throughput for each CE configuration, while the bottom chart shows the PU throughput. Error bars are shown for each bar in the plot. The magnitude of the SU throughput is primarily affected by the desired waveform and CE status (the x-axis). Smaller deviations in SU throughput are due to it being affected by the specific PU created environment. The PU throughput magnitude is primarily affected by the choice of the PU environment, though the choice of the SU waveform and CE status provides a smaller effect. For the SU throughput, the experimental limit line for each CE configuration is the maximum performance measured for that CE and waveform configuration with the PU turn off. For the PU throughput, the experimental limit line is the maximum measured PU performance with the SU turned off.

The remainder of this chapter examines the effect of the SU CE on SU throughput and BER, the effect of the SU CE on PU throughput and BER, the overall rank of each CE using a scoring system, and an examination on the validity of the data. Finally, this chapter presents the data collection attempt using the DYSE.

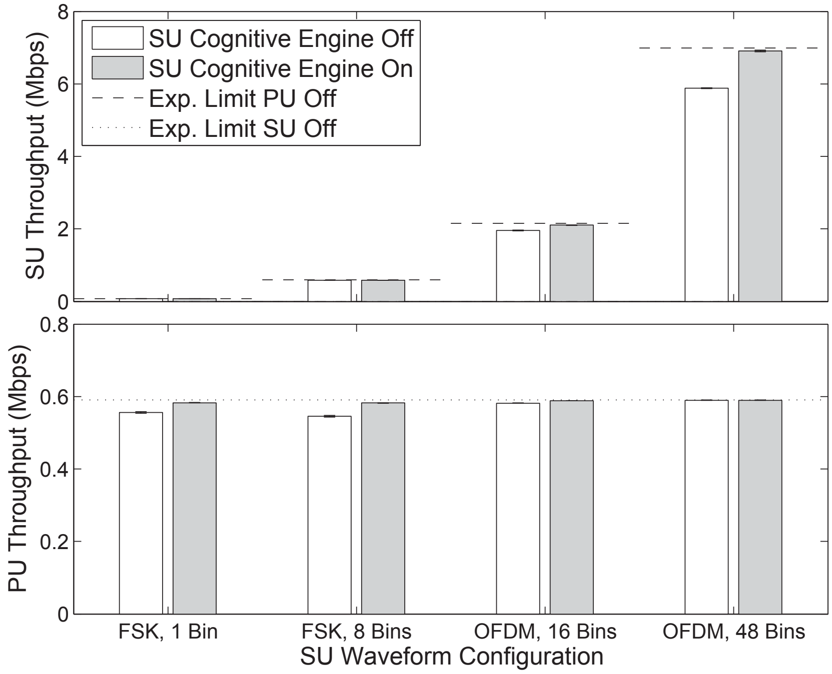
5.2 Effect of Cognitive Engine on Secondary User Throughput and BER

A t-test comparison of the SU throughput values is shown in Table 5.1. The t-test utilized is a single sided t-test to evaluate whether or not the SU throughput with the CE on is higher than when the CE is off. For example, the t-test evaluates whether or not the throughput of OFDM-16 with the CE on is greater than the throughput of OFDM-16 with the CE off. The t-test significance level used is α=0.05.

The CE improves SU throughput in 55% of cases when compared to the corresponding waveform with the CE off. The exceptions are the cases where the PU utilizes a relatively large amount of the bandwidth, as with Environments 3, 4 and 5. For these exceptions, the throughput with the CE off is greater for one primary reason. This is because the CE on condition means that the waveform is adapting to the environment; as a result, the bandwidth available to the SU decreases and so the throughput decreases as well.

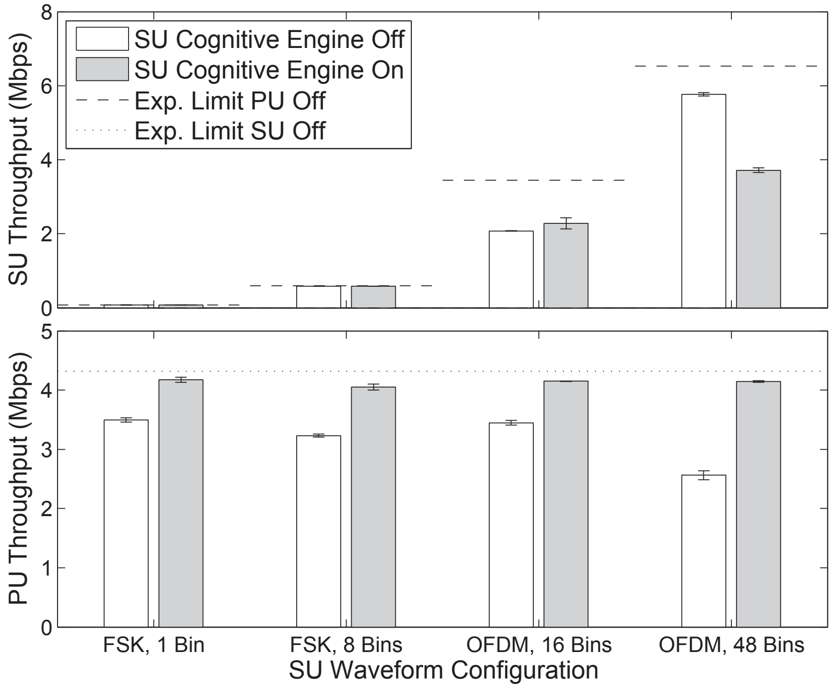


(a) Environment 1

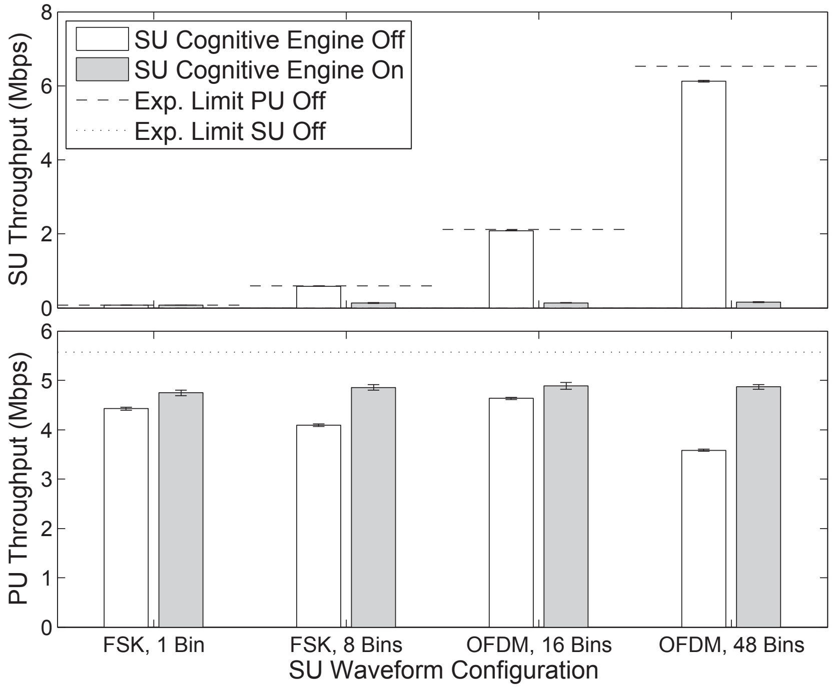


(b) Environment 2

Figure 5.1: PU and SU Throughput, Environments 1 and 2

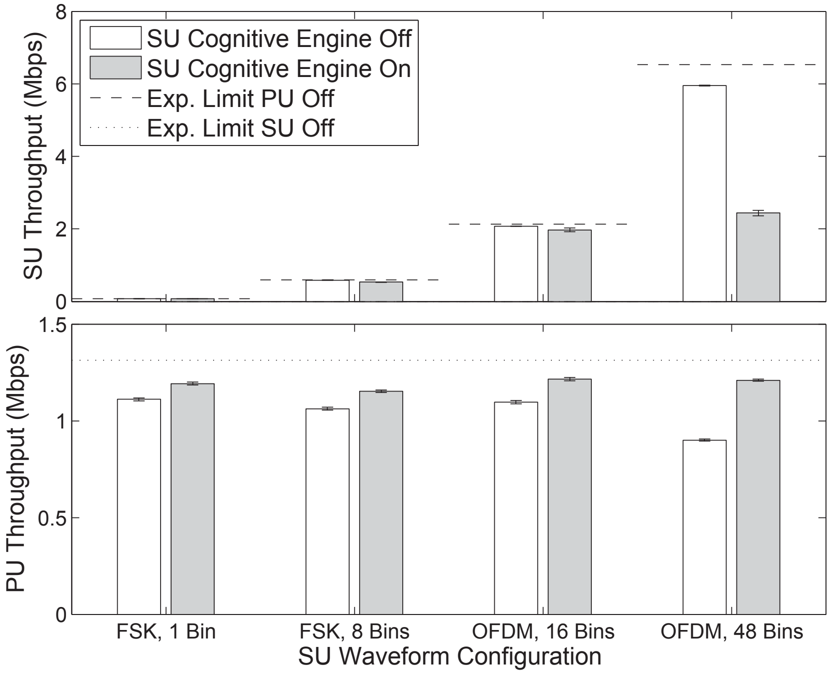


(a) Environment 3



(b) Environment 4

Figure 5.2: PU and SU Throughput, Environments 3 and 4



(a) Environment 5

Figure 5.3: PU and SU Throughput, Environment 5

Table 5.1: Secondary User Throughput

Environment	Waveform	Mean Throughput (bps)		t-test: CE Off less than CE On?	p-value
		CE Off	CE On		
1	FSK 1	72566.5	73181.12	Yes	2.86E-13
1	FSK 8	586072.1	588475.3	Yes	6.97E-06
1	OFDM 16	1959149	2126356	Yes	3.29E-07
1	OFDM 48	5917864	6859417	Yes	6.48E-17
2	FSK 1	72477.17	73178.99	Yes	6.07E-13
2	FSK 8	586129.3	588055.3	Yes	3.58E-07
2	OFDM 16	1957118	2101196	Yes	6.89E-36
2	OFDM 48	5882772	6911066	Yes	3.85E-59
3	FSK 1	72651.49	73178.46	Yes	3.37E-11
3	FSK 8	588245.1	585617.6	No	1
3	OFDM 16	2076489	2282354	Yes	0.000986
3	OFDM 48	5765735	3716908	No	1
4	FSK 1	72416.31	71440	No	0.989317
4	FSK 8	588485.9	131121.8	No	1
4	OFDM 16	2088617	137588.4	No	1
4	OFDM 48	6122345	156274.4	No	1
5	FSK 1	72515.63	73152.62	Yes	3.51E-10
5	FSK 8	588330.6	534137	No	1
5	OFDM 16	2072105	1971300	No	0.999768
5	OFDM 48	5947066	2436589	No	1

By comparison, the throughput of the SU with the CE on is greater for all narrowband interference conditions (Environments 1 and 2).

Next, the performance of the CE is evaluated with respect to relative CE capability. These results only apply to the condition when the SU CE is on. FSK-1 is assumed to be the least cognitive CE configuration, while OFDM-48 is assumed to be the most cognitive CE configuration. For this analysis, each CE configuration is compared to all previous configurations to evaluate whether it is statistically greater or statistically less (using t-test, α=0.05). If the difference is neither greater nor less, then the configurations are considered to have equal performance. The results of this analysis are presented in Table 5.2. Waveform 1 is the waveform with greater adaptation, waveform 2 is the waveform with relatively lower adaptation.

The throughput increases for greater cognition levels in all Environments except for one case. This one exception shows that in PU Environment 4, OFDM 16 has performance that is not statistically different than the FSK 8 performance. In no cases is the throughput less for greater cognition levels.

The BER, which is equivalent to the SER for the waveforms in this experiment, is shown in Table 5.3 along with associated statistics. In this table, the BER statistics are computed between experiments, thus treating each experimental result as a random variable. Note that these statistics only apply to the experimental condition where both the PU and SU are on. The BER of the SU is shown graphically in Figure 5.4.

Relative to the SU waveform with the CE off, the BER decreases at a statistically significant level (α=0.05) when the CE is on 75% of the time. Of the remaining 25%, two cases have p-values between 0.05 and 0.95, indicating that they are not statistically significant. This leaves three of the comparisons, or 15%, to not follow the trend of decreasing the BER when the CE is turned on. These cases are for Environment 4/FSK-1,

Table 5.2: Relative Comparison of Secondary User Throughput with CE On

Environment	Waveform 1	Waveform 2	Waveform 1 Throughput Greater than Waveform 2?	p-value	Waveform 1 Throughput Less than Waveform 2?	p-value
1	FSK 8	FSK 1	Yes	1.2E-142	No	1
1	OFDM 16	FSK 1	Yes	3.55E-56	No	1
1	OFDM 16	FSK 8	Yes	1.01E-49	No	1
1	OFDM 48	FSK 1	Yes	1.02E-65	No	1
1	OFDM 48	FSK 8	Yes	8.27E-64	No	1
1	OFDM 48	OFDM 16	Yes	7.82E-51	No	1
2	FSK 8	FSK 1	Yes	8.9E-142	No	1
2	OFDM 16	FSK 1	Yes	8.3E-112	No	1
2	OFDM 16	FSK 8	Yes	1.79E-96	No	1
2	OFDM 48	FSK 1	Yes	2.7E-111	No	1
2	OFDM 48	FSK 8	Yes	6.8E-101	No	1
2	OFDM 48	OFDM 16	Yes	1.4E-99	No	1
3	FSK 8	FSK 1	Yes	1.1E-127	No	1
3	OFDM 16	FSK 1	Yes	1.93E-35	No	1
3	OFDM 16	FSK 8	Yes	3.2E-30	No	1
3	OFDM 48	FSK 1	Yes	3.86E-67	No	1
3	OFDM 48	FSK 8	Yes	9.95E-64	No	1
3	OFDM 48	OFDM 16	Yes	4.1E-25	No	1
4	FSK 8	FSK 1	Yes	3.28E-18	No	1
4	OFDM 16	FSK 1	Yes	1.09E-18	No	1
4	OFDM 16	FSK 8	No	0.165192	No	0.834808
4	OFDM 48	FSK 1	Yes	1.43E-16	No	1
4	OFDM 48	FSK 8	Yes	0.002125	No	0.997875
4	OFDM 48	OFDM 16	Yes	0.015524	No	0.984476
5	FSK 8	FSK 1	Yes	1.66E-72	No	1
5	OFDM 16	FSK 1	Yes	2.61E-55	No	1
5	OFDM 16	FSK 8	Yes	7.01E-50	No	1
5	OFDM 48	FSK 1	Yes	1.59E-54	No	1
5	OFDM 48	FSK 8	Yes	3.19E-50	No	1
5	OFDM 48	OFDM 16	Yes	1.13E-14	No	1

Table 5.3: Secondary User BER

Environment	Waveform	Mean BER		t-test: CE On	p-value
		CE Off	CE On	less than CE Off?	
1	FSK 1	0.006614	0	Yes	6.43811E-13
1	FSK 8	0.004244	0.000106	Yes	1.74591E-08
1	OFDM 16	0.071575	0.007902	Yes	2.35172E-15
1	OFDM 48	0.101613	0.04094	Yes	2.78773E-10
2	FSK 1	0.007182	0	Yes	7.41619E-13
2	FSK 8	0.0039	0.000507	Yes	3.4321E-14
2	OFDM 16	0.074598	0.017047	Yes	1.1527E-41
2	OFDM 48	0.106953	0.049613	Yes	1.10588E-47
3	FSK 1	0.005432	0	Yes	8.0045E-11
3	FSK 8	0.000003	0	No	0.160970674
3	OFDM 16	0.014988	0.001915	Yes	1.15547E-17
3	OFDM 48	0.124882	0.017339	Yes	6.08817E-37
4	FSK 1	0.008076	0.018283	No	0.989102325
4	FSK 8	0	0.004765	No	0.999699766
4	OFDM 16	0.010543	0.005217	Yes	0.00028507
4	OFDM 48	0.070659	0.013046	Yes	1.69936E-29
5	FSK 1	0.006912	0.000327	Yes	6.18991E-10
5	FSK 8	0	0.000023	No	0.911885749
5	OFDM 16	0.019015	0.037046	No	1
5	OFDM 48	0.097331	0.052722	Yes	9.06373E-41

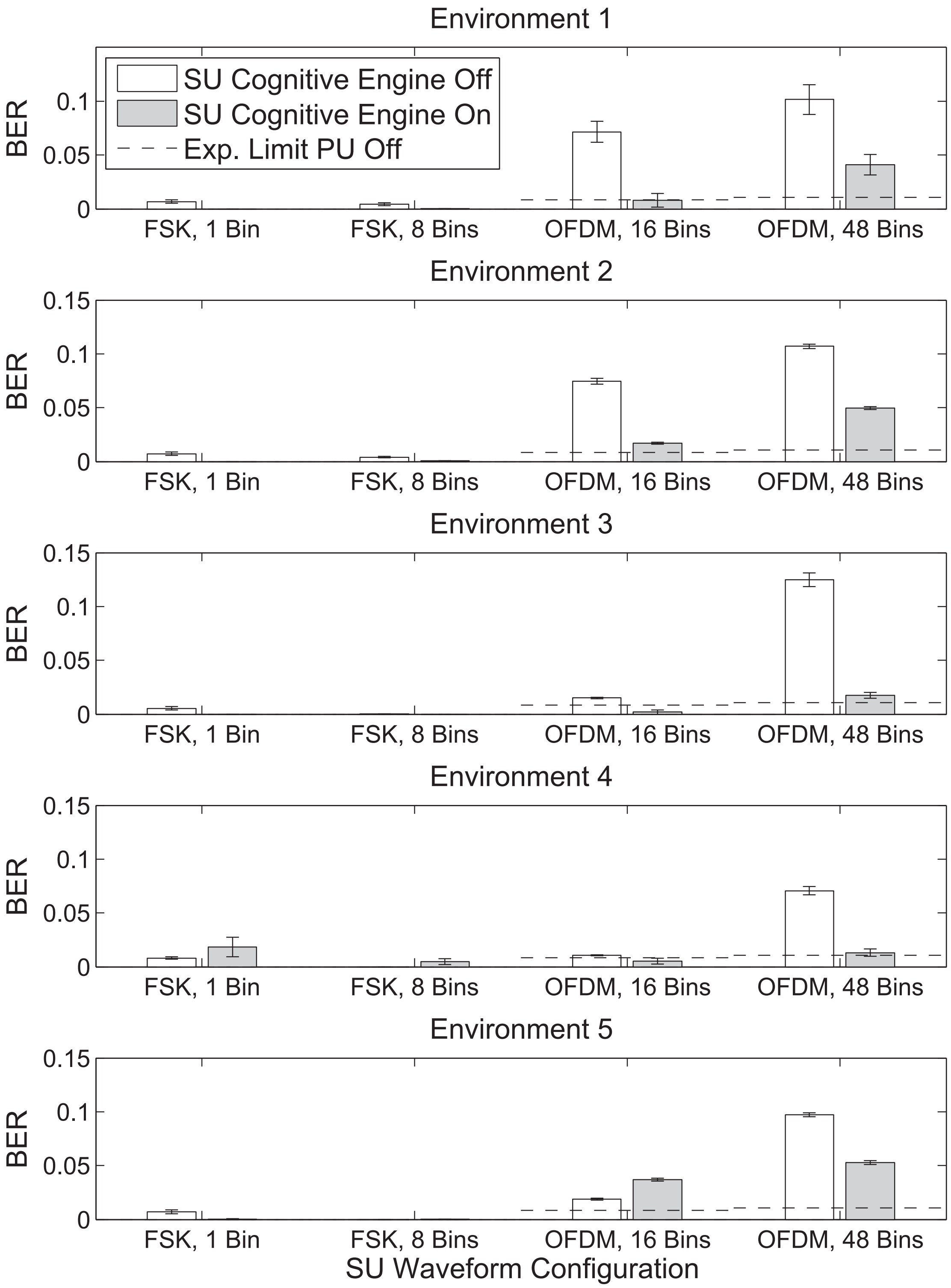


Figure 5.4: Secondary User BER
(Lines and Bars not visible are near or equal to 0)

Environment 4/FSK-8, and Environment 5/OFDM-16. The latter case is not consequential as the OFDM waveforms do not necessarily perform reliably. That is, pilot experiments found that non-contiguous waveforms have a significantly higher BER than contiguous waveforms. Since all cases using OFDM with the CE off are contiguous, this means that it is both possible and likely for the CE off to have a lower BER as the CE on condition will likely be non-contiguous due to its adapting to the non-contiguous environment. This leaves two anomalous data points: Environment 4/FSK-1 and Environment 4/FSK-8. However, the effect of the CE on PU throughput must also be considered.

5.3 Effect of Cognitive Engine on Primary User Throughput and BER

A t-test comparison of the PU throughput values is shown in Table 5.4. As with the SU throughput comparison, the t-test utilized is a single sided t-test to evaluate whether or not the PU throughput with the CE on is higher than when the CE is off. The t-test significance level used is α=0.05.

The CE improves PU throughput in every case except for one, which is Environment 2/OFDM-48. In this case, there is statistically no difference between the performance of the two CE conditions. The PU throughput never decreases with the CE on as compared to the CE off.

As with the SU throughput, the performance of the CE is evaluated with respect to relative CE capability. Again, these results only apply to the condition when the SU CE is on. For this analysis, each CE configuration is compared to all previous configurations to evaluate whether it is statistically greater or statistically less (using t-test, α=0.05). If the difference is neither greater nor less, then the configurations are considered to have equal performance. The results of this analysis are presented in Table 5.5. Waveform 1 is the waveform with greater adaptation, waveform 2 is the waveform with relatively lower adaptation.

Table 5.4: Primary User Throughput

Environment	Waveform	Mean Throughput (bps)		t-test: CE Off less than CE On?	p-value
		CE Off	CE On		
1	FSK 1	552154.1	589581.2	Yes	9.04E-34
1	FSK 8	547076.2	589512.3	Yes	2.59E-34
1	OFDM 16	582573.1	590011.1	Yes	1.42E-13
1	OFDM 48	589703.8	589871.7	Yes	0.024796
2	FSK 1	555733.9	582969.4	Yes	1.67E-32
2	FSK 8	545570.3	581846.5	Yes	4.41E-34
2	OFDM 16	581621.5	588271.4	Yes	2.07E-24
2	OFDM 48	589780.6	589836.6	No	0.252609
3	FSK 1	3494620	4176330	Yes	1.63E-30
3	FSK 8	3230874	4052200	Yes	7.39E-35
3	OFDM 16	3446377	4148722	Yes	2.05E-35
3	OFDM 48	2561634	4145628	Yes	2.64E-43
4	FSK 1	4427361	4748057	Yes	9.43E-15
4	FSK 8	4092883	4855452	Yes	2.96E-32
4	OFDM 16	4632167	4887000	Yes	4.94E-09
4	OFDM 48	3581395	4867963	Yes	9.69E-47
5	FSK 1	1110428	1192591	Yes	3.99E-22
5	FSK 8	1062312	1153178	Yes	7.29E-26
5	OFDM 16	1096236	1215137	Yes	1.63E-26
5	OFDM 48	900189.2	1209777	Yes	1.04E-63

Table 5.5: Relative Comparison of Primary User Throughput with CE On

Environment	Waveform 1	Waveform 2	Waveform 1 Throughput Greater than Waveform 2?	p-value	Waveform 1 Throughput Less than Waveform 2?	p-value
1	FSK 8	FSK 1	No	0.609159	No	0.390841
1	OFDM 16	FSK 1	Yes	0.015775	No	0.984225
1	OFDM 16	FSK 8	Yes	0.012376	No	0.987624
1	OFDM 48	FSK 1	No	0.052527	No	0.947473
1	OFDM 48	FSK 8	Yes	0.036262	No	0.963738
1	OFDM 48	OFDM 16	No	0.900236	No	0.099764
2	FSK 8	FSK 1	No	0.934301	No	0.065699
2	OFDM 16	FSK 1	Yes	1.78E-13	No	1
2	OFDM 16	FSK 8	Yes	6.35E-19	No	1
2	OFDM 48	FSK 1	Yes	7.95E-18	No	1
2	OFDM 48	FSK 8	Yes	2.78E-23	No	1
2	OFDM 48	OFDM 16	Yes	3.03E-15	No	1
3	FSK 8	FSK 1	No	0.999832	Yes	0.000168
3	OFDM 16	FSK 1	No	0.878119	No	0.121881
3	OFDM 16	FSK 8	Yes	0.000317	No	0.999683
3	OFDM 48	FSK 1	No	0.912021	No	0.087979
3	OFDM 48	FSK 8	Yes	0.000238	No	0.999762
3	OFDM 48	OFDM 16	No	0.659105	No	0.340895
4	FSK 8	FSK 1	Yes	0.004328	No	0.995672
4	OFDM 16	FSK 1	Yes	0.001667	No	0.998333
4	OFDM 16	FSK 8	No	0.244949	No	0.755051
4	OFDM 48	FSK 1	Yes	0.000683	No	0.999317
4	OFDM 48	FSK 8	No	0.363533	No	0.636467
4	OFDM 48	OFDM 16	No	0.672912	No	0.327088
5	FSK 8	FSK 1	No	1	Yes	1.41E-10
5	OFDM 16	FSK 1	Yes	6.57E-05	No	0.999934
5	OFDM 16	FSK 8	Yes	1.98E-17	No	1
5	OFDM 48	FSK 1	Yes	0.000187	No	0.999813
5	OFDM 48	FSK 8	Yes	5.42E-20	No	1
5	OFDM 48	OFDM 16	No	0.881703	No	0.118297

The PU throughput does not always increase for greater cognition levels. However, the vast majority of these cases are not statistically significant; that is, the PU throughput does not decrease for greater cognition levels. However, there are two cases in which the PU throughput does decrease for greater cognition levels: Environment 3/FSK-8 & FSK-1 and Environment 5/FSK-8 & FSK-1. In both of these cases, the FSK-8 waveform negatively impacts the PU compared to the FSK-1 waveform. Interestingly, the FSK-8 does not improve PU throughput for any environment except for Environment 4. Possible factors could be a failure of the CE to properly sense the environment, the selection of a sub-optimal channel by the CE, or additional interference caused by the FSK-8 waveform apart from channel selection. The first two factors were investigated and not found to be factors in and of themselves. Analysis of the problem revealed that the side-lobe power was not considered in development of the CE due to power not being considered a relevant factor. In reality, the FSK waveforms do not transmit solely on the desired frequencies and have non-negligible sidelobes. The additional power coming from the sidelobes is more likely to cause interference to the PU with wider bandwidth SU waveforms. This is consistent with the results found.

Also, it is important to note that the actual waveform used by the SU may be different than the desired waveform. For example, if the largest spectrum hole created by the PU is less than 8 bins wide, the CE FSK waveform will downgrade from 8 to 4, 2, or 1 bins of bandwidth. This may account for the Environment 4 being different than the other environments with respect to the FSK 8 / FSK 1 relationship.

The BER and associated statistics are shown in Table 5.6. The BER of the PU is shown graphically in Figure 5.5.

In every case, the PU has a lower BER when the SU CE is on as compared to when it is off. The two anomalous data points in which the SU BER was worse with the CE on still exhibit the desired effect towards the PU.

Table 5.6: Primary User BER

Environment	Waveform	Mean BER		t-test: CE On less than CE Off?	p-value
		CE Off	CE On		
1	FSK 1	0.063888	0.000509	Yes	7.81E-34
1	FSK 8	0.072066	0.000668	Yes	1.02E-34
1	OFDM 16	0.012551	0.00002	Yes	1.32E-13
1	OFDM 48	0.000283	0.000002	Yes	4.04E-11
2	FSK 1	0.057951	0.011575	Yes	2.47E-32
2	FSK 8	0.074967	0.013682	Yes	3.42E-34
2	OFDM 16	0.014407	0.002826	Yes	6.57E-29
2	OFDM 48	0.000273	0.00006	Yes	1.75E-12
3	FSK 1	0.180992	0.023933	Yes	3.78E-46
3	FSK 8	0.24082	0.049386	Yes	1.03E-42
3	OFDM 16	0.191427	0.023141	Yes	9.29E-42
3	OFDM 48	0.404031	0.02501	Yes	5.89E-51
4	FSK 1	0.238317	0.18244	Yes	1.08E-18
4	FSK 8	0.29457	0.166341	Yes	1.81E-35
4	OFDM 16	0.206524	0.160484	Yes	1.9E-12
4	OFDM 48	0.385684	0.162651	Yes	4.67E-50
5	FSK 1	0.178626	0.117131	Yes	1.93E-37
5	FSK 8	0.21625	0.148447	Yes	9.51E-37
5	OFDM 16	0.189113	0.102721	Yes	3.59E-45
5	OFDM 48	0.337774	0.111745	Yes	1.28E-68

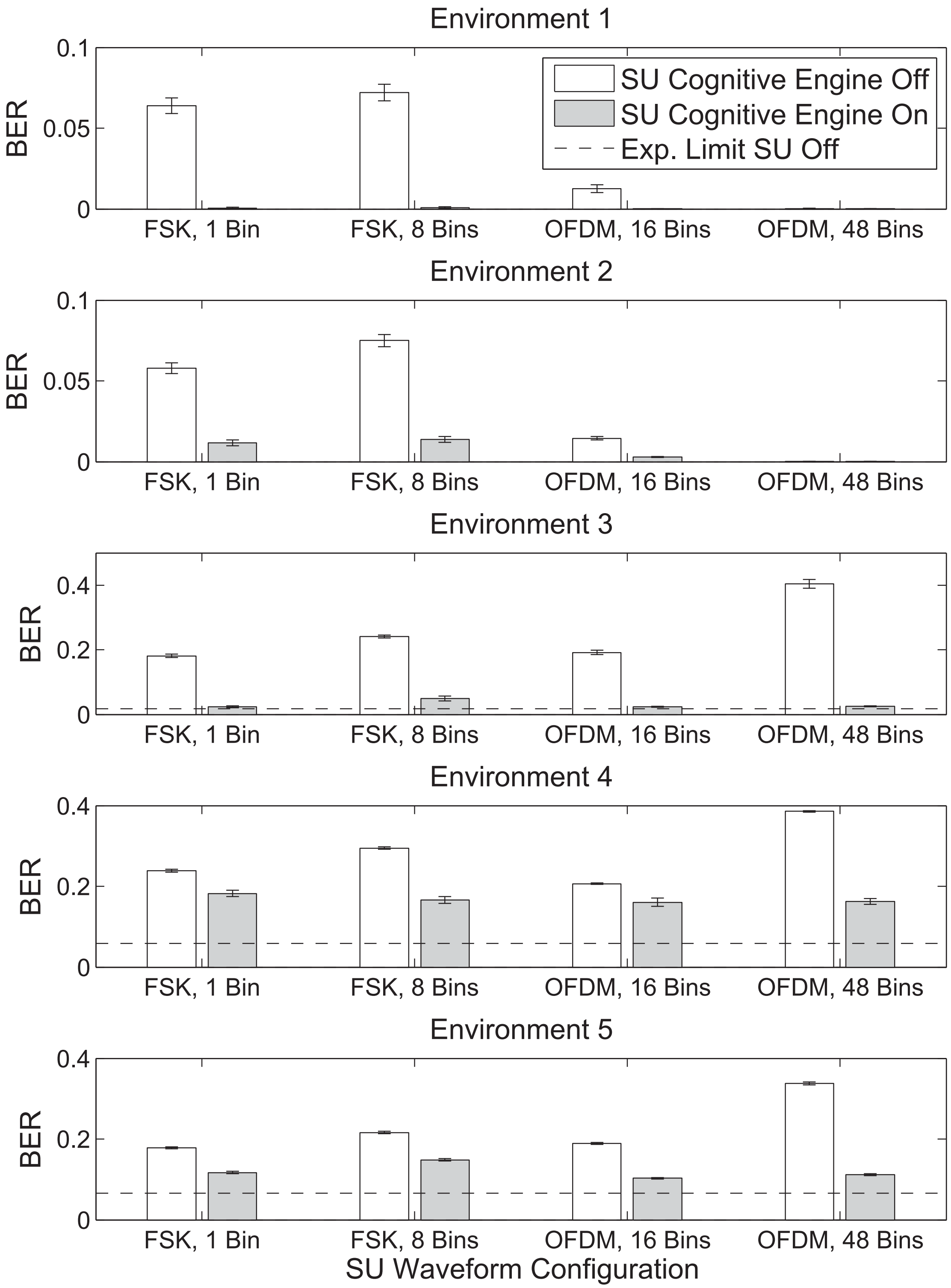


Figure 5.5: Primary User BER
(Lines and Bars not visible are near or equal to 0)

5.4 Scoring

CRATM proposes to use benchmarks and relative scoring to evaluate CRSs. A preliminary means of scoring was applied against the experimental results. The scoring presented here reflects the data analysis already presented and is not intended for drawing conclusions on the hypothesis. Rather, the scoring system is presented to show how CRATM captures the experimental results. Scores are based on data means.

For each environment, the highest performing SU configuration (e.g. SU CE on, OFDM 48 waveform) is assigned a score of 10, as measured by throughput. The lowest performing SU configuration is assigned a 0. All other data points are assigned a value between 0 and 10 based on their relative value to the minimum and maximum. This ranking is applied to both the SU throughput and the PU throughput. Then, the SU throughput score and the PU throughput score are averaged together by a specified weight to produce the final score.

The final score reflects the performance of both the SU and PU combined into a single metric. Using the CRATM CR definition, the most cognitive configuration should have the highest score while the least cognitive configuration should have the lowest score. This is not a rigorous evaluation of cognition due to the limitations of the SUT; however, it provides a rough means to correlate the experimental results with CRATM.

Table 5.7: Performance Scoring, SU Weight=50% PU Weight=50%

Environment	FSK 1 CE Off	FSK 8 CE Off	OFDM 16 CE Off	OFDM 48 CE Off	FSK 1 CE On	FSK 8 CE On	OFDM 16 CE On	OFDM 48 CE On
1	0.2	0	5.4	9.3	4.8	5.1	6.4	10
2	0.8	0	5.3	9.2	4	4.3	6.2	10
3	0.7	0	3.5	4.4	4.4	4.6	7.8	10
4	2.6	0	10	7.9	6.3	7.7	8.1	7.9
5	0.8	0	4.6	5.1	4.2	3.6	9.2	10

Table 5.7 shows the scoring using a 50/50% weighting. Table 5.8 shows the results from changing the weighting to 25/75%, where 75% of the weight is accorded to PU throughput.

Table 5.8: Performance Scoring, SU Weight=25% PU Weight=75%

Environment	FSK 1 CE Off	FSK 8 CE Off	OFDM 16 CE Off	OFDM 48 CE Off	FSK 1 CE On	FSK 8 CE On	OFDM 16 CE On	OFDM 48 CE On
1	0.7	0	6.9	9.6	7.4	7.6	8.2	10
2	1.6	0	6.7	9.6	6.3	6.3	7.9	10
3	2.8	1.3	3.9	0	7.7	7.2	9	10
4	4.7	1.3	8.7	0	8.4	9.6	10	9.8
5	4.3	2.7	5.1	0	7.6	6.3	9.9	10

In the 50/50 case, there is an upward trend in scores for greater levels of cognition when the CE is on. The one exception to this trend is from a case where the SU performs better with increased cognition, but the PU performance difference is statistically insignificant. Furthermore, the CE on condition scores higher than the CE off condition except for Environment 4. As discussed previously, in this environment the SU with the CE on effectively constrains itself to not cause interference to the PU, meaning that the SU throughput drops by a large amount relative to the CE off condition. In the 25/75 case, the same trends are evident except that the CE on condition always performs better than the corresponding CE off condition.

For future scoring, actual weights would be set to desired benchmark conditions. For example, if PU non-interference is a key performance parameter, then the weight set for that metric would be set accordingly higher.

5.5 DYSE

Data collection was attempted using the DYSE. A virtual PU transmitter was created in the DYSE that transmitted a recurring bit-sequence to a physical PU receiver. A physical SU transmitter and receiver communicated through the DYSE RF interface. Unfortunately, the RF signals in and out of the DYSE were not useable and yielded bit error probabilities of 0.5. The likely problem was a lack of compatible RF interface between the DYSE and the SUT which necessitated use of mixers and signal generators. However, this generated additional noise and distorted the waveforms. Thus, no data was collected on the efficacy of using the proposed test framework in an emulated environment.

VI. Conclusion

This chapter first summarizes the conclusions drawn from the results of research into developing a test methodology to test CRSs. Then, these conclusions are used to form a basis for future work.

6.1 Summary

The proposed test methodology, or CRATM, was developed in response to a perceived need in the CR field to have a unified means of testing CR devices. The CR field is diverse with a wide range of views on what it means to be a CR, what cognition is required, how performance is measured, and differing opinions on how to test. CRATM was developed in response to this lack of unification. CRATM is designed to be an overarching framework that allows CRs developed under different viewpoints to be tested using the same test methodology.

CRATM uses the idea of benchmarks to allow repeatable, measurable, and comparative experiments. CRATM proposes to use an emulated radio environment, but is not required to do so. With CRATM, the impact of the SU on the PU can be measured directly by measuring the performance of the PU. The PU creates the radio environment that the SU operates in. CRATM assumes that cognition may be measured by evaluating the ability of the SU to improve throughput while minimizing interference to the PU in accordance with user goals and policy. CR performance may be measured by evaluating the performance of both the SU and the performance of the PU. This is known as behavior-based testing.

CRATM was found to be in general agreement with the literature except that it does not directly test cognitive or spectrum sensing abilities. In addition, CRATM does not use psychometric or REM-SDT testing, which are two proposed test methodologies in the literature. Behavior-based testing was proposed as a solution to these discrepancies. An experiment was setup to gather data on the hypothesis that CR cognition may be measured by evaluating the performance of both the SU and PU without explicitly evaluating the cognition of the SU.

The experiment did not disprove the hypothesis that SU cognition may be measured via SU and PU performance. PU throughput was found to increase, or stay the same, for every case where the SU used the CE and/or a greater level of cognitive waveform adaptability. The SU throughput increased for PU narrow-band environments but decreased for PU wide-band environments. However, this response is to be expected as the SU adapts to the environment. BER of both the PU and SU improved when the SU CE was turned on. Both PU and SU performance improved whenever the SU CE was turned on. Additionally, both PU and SU performance improved or stayed the same for greater levels of SU cognitive waveform adaptability. Finally, the experiment showed that it is possible to collect data on SU and PU performance by operating a SU and PU system simultaneously.

Based on the results of these experiments, behavior-based testing was not found to be invalid. Because of this, CRATM was not shown to invalid. However, further research is required in order to fully justify behavior-based testing and CRATM. A wider range of PU workloads and SU CRSs need to be evaluated to determine if behavior-based testing is effective across broad CR architectural features and DSA environments. For the limited scope of this research, behavior-based testing and CRATM was shown to be effective.

6.2 Future Work

Future research should continue to validate CRATM as well as to continue investigation on the efficacy of evaluating SU and PU performance by directly measuring their performance. Practical benchmarks and benchmark classes should be formalized and tested in accordance with the CRATM framework. More CR prototypes with a wider variety of cognition should be tested under the framework, especially to determine if measured performance does in fact correlate with underlying cognition. Future research should transition the test methodology over to the DYSE in order to fully justify using an emulated environment in CRATM. In the short term, ACRO researchers should continue to put cognitive functions onto hardware in order to build and test a stand-alone CR prototype. There are certainly other suitable areas for future research, but based on the results of this research, these are the most viable.

Appendix A: Data Validation

This chapter analyzes the collected experimental data for validity. Besides the throughput and BER data presented in Section 5.1, additional experimental data was collected. This additional data collected for each experiment included the elapsed time, total number of transmissions, number of symbols counted, and spectrum sensing calibration data. The goal of this data validation analysis is to determine if appropriate experiment controls were in place and that the collected throughput/BER data is valid.

For the results to be valid, there should not be any inconsistencies in the data. Inconsistencies are flagged if output residuals are non-normal or if there are excessive and/or influential outliers. The measurable outputs included the total number of transmissions, the number of symbols counted, the time spent in the experiment, the mean received signal strength (from calibration) as seen by the spectrum sensor, and the computed threshold values for the spectrum sensor. The total number of transmissions and the experiment elapsed time are impacted primarily by the the hardware, software and experiment setup. Inconsistencies here indicate a problem in the SUT. The number of symbols counted, as well as the mean received signal strength and thresholds, are impacted primarily by the external RF environment. Inconsistencies here indicate a problem in the experimental test setup or the SUT. Throughput and BER data is also examined for problems.

The data is analyzed for two trends. First, each output is examined for global behavior. Second, the residuals for each output are examined for each specific environment and CE configuration. Both the global behavior and specific configuration residuals should show a normal distribution (when applicable). If not, there is a confounding factor that needs to be examined.

For the following data analysis, the SU waveform configurations will be referred to as a number. The legend for the SU waveform configuration is shown in Table A.1.

Table A.1: SU Waveform Configuration Legend

SU Waveform Configuration	Identifier
FSK, 1 Bins BW, CE Off	1
FSK, 8 Bins BW, CE Off	2
OFDM, 16 Bins BW, CE Off	3
OFDM, 48 Bins BW, CE Off	4
FSK, 1 Bins BW, CE On	5
FSK, 8 Bins BW, CE On	6
OFDM, 16 Bins BW, CE On	7
OFDM, 48 Bins BW, CE On	8

A.1 Elapsed Time

The first output to be examined is the time spent in the experiment, or the elapsed time. This time is measured from the receipt/acknowledgement of the experiment start trigger signal until the total experiment target time has been reached. Additionally, a time window was used to keep the PU and SU transmitting to a semi-synchronous level. For instance, if both PU and SU are programmed for 60 seconds, it is not desired to have the SU finish at 60.01 seconds while the PU starts a new one second transmission at 59.99 seconds. For the PU, the total experiment target time is 60 seconds +/- 0.5 seconds. The programmed duration for each transmit cycle is one second. For the SU, the total experiment target time is 60 seconds +/- 0.125 seconds. The programmed duration for each transmit cycle (which includes sensing the environment) is 0.25 seconds. Of interest in this analysis are any outliers, and if so, if they come from a particular input. If any experiment configurations show a consistent bias in elapsed time, then the experiment results will be impacted as well.

The histogram of the elapsed times for both the SU and PU are shown in Figure A.1. The elapsed times follow a normal distribution, as expected. Next, the elapsed times are investigated for non-normal responses to inputs.

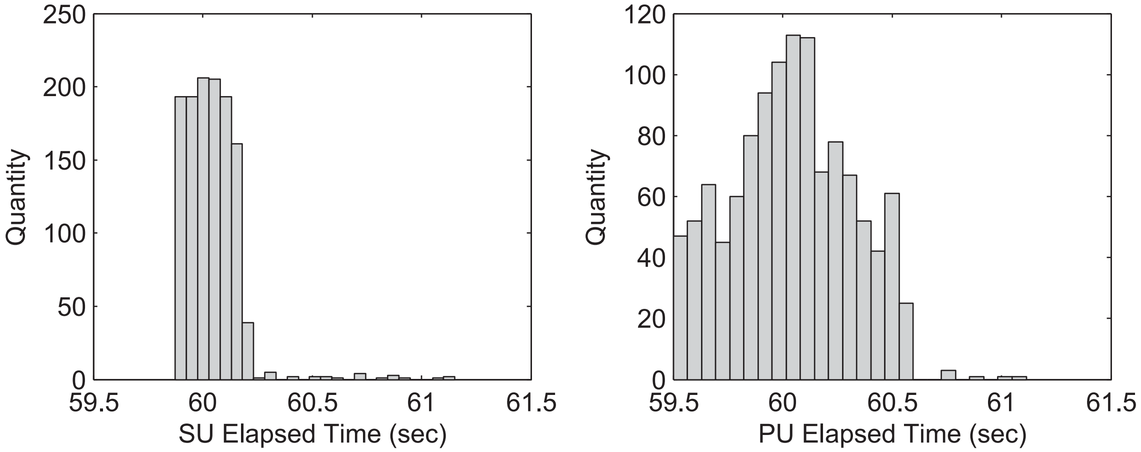


Figure A.1: Histograms of Total Experiment Elapsed Time

For the SU, the input driver for elapsed time is the choice of the CE. For the PU, the controlled input factor for elapsed time is the Environment. In the experiment code, the PU environment is converted to a CE configuration with a pre-generated transmit waveform, as opposed to the SU which creates the transmit waveform based on the sensed environment. The residuals for the elapsed time are shown in Figure A.2. The residuals are evaluated with respect to the global mean of elapsed time.

In neither case is the controlled input affecting the elapsed time at a statistically significant level. The residual outliers are likely due to delays in communicating with the WARP boards when sending/receiving data or commands. Regardless of the cause, the data shows that the SU and PU elapsed time follows the expected normal distribution.

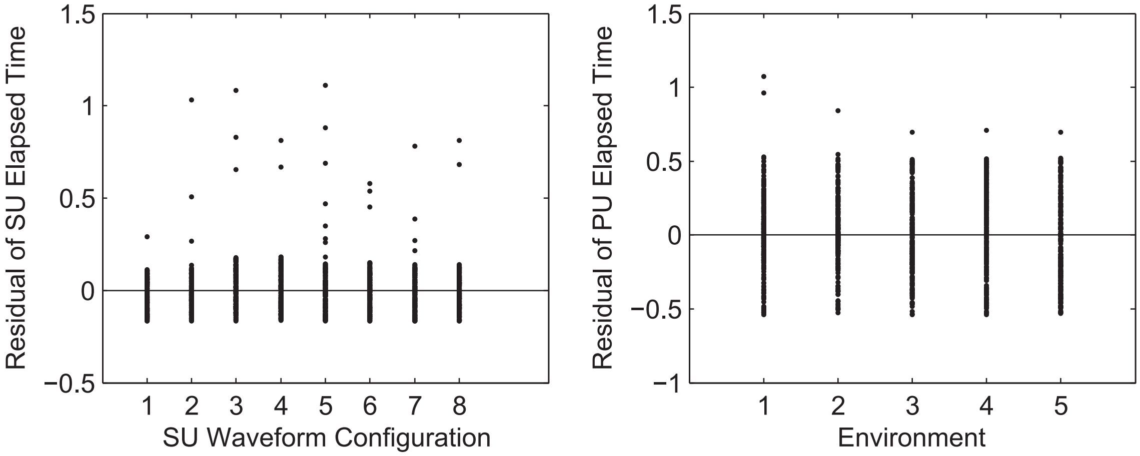


Figure A.2: Total Experiment Elapsed Time Residuals

A.2 Number of Transmissions Received

The number of transmissions received is the total number of times the receiver node (either SU or PU) polled its RX buffer to capture the transmitted signal. For the PU, this number is affected by MATLAB processing time, which is in turn affected by the choice of waveform modulation. For the SU, the number of transmissions received is affected by the use of spectrum sensing or not, the PU environment, and the waveform modulation. Figure A.3 shows the relationship of the number of transmissions received to the input factor (SU waveform configuration or PU Environment). As with elapsed time, the data is examined for outliers or data inconsistencies, which may indicate problems either with the SUT.

As can be seen in Figure A.3, both the PU and SU number of transmissions received are impacted by the choice of inputs, among other factors. The impact on the PU is discussed first.

For the PU, a linear model was created to examine the relationship between the number of transmissions received and the PU Environment. Furthermore, an additional factor was added to the model to distinguish between OFDM and FSK waveforms, as well as factor for OFDM modulation/demodulation time processing differences. While the various OFDM

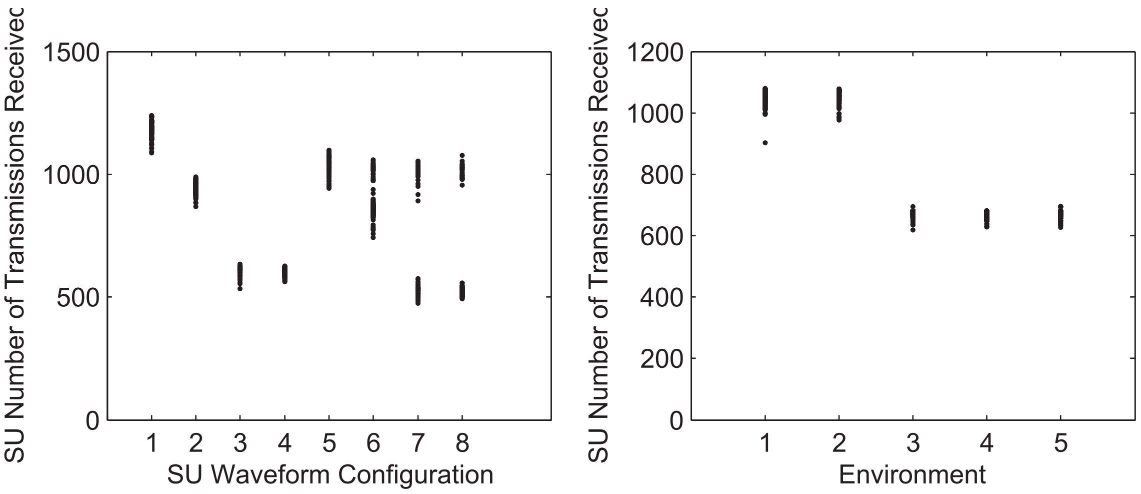


Figure A.3: Scatterplot of Total Number of Transmissions Received

environments (Environments 3, 4, and 5) execute the same code, the bandwidth changes the number of symbols created and modulated, which has an effect on overall processing time. The residuals of the fitted linear model are shown in Figure A.4 on both a probability plot and a histogram.

The residuals approximate a normal distribution; however, there is a significant leftwards skew. Additional inputs factors were investigated to determine if they impacted the PU number of transmissions received, such as the order of the trial to determine if time-based effects were present. None of the additional input factors were determined to have a discernible effect on the number of transmissions received. Underlying causes for the left skewed data points could arise from the limited amount of samples collected, the effect of the computer operating system and MATLAB software, or delays in communicating with the WARP hardware.

The SU number of transmissions show several interesting trends (see Figure A.3). To begin with, the CE on waveform configurations (5 - 8) are clearly shifted downwards from the CE off waveform configurations. This is indicative of the extra time required to

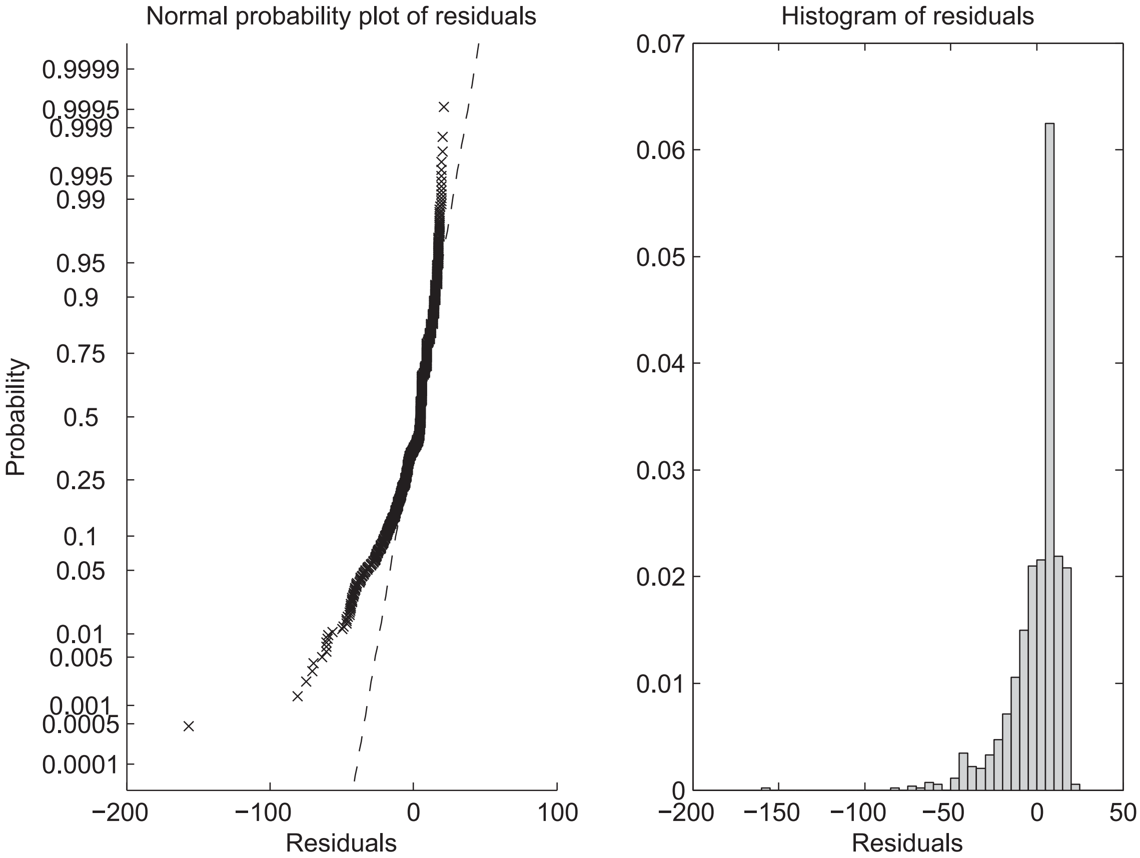


Figure A.4: Residuals of Fitted Linear Model: PU Total Number of Transmissions receive spectrum data from the WARP boards and process it. Furthermore, evidence of the CE waveform downgrading to lower performing waveforms is clearly evident in the dual clusters for waveforms 6 - 8 (note: lower performing waveforms actually have higher number of transmissions received due to less modulation/demodulation requirements). In waveform 6, the CE waveform seems to transition between FSK-8 down to FSK-1 (with possibly 2 intermediate stages in between). The OFDM waveforms show transitions from the OFDM waveform down to FSK waveforms.

As with the PU, a linear model was fitted to the data to ascertain undesired behavior. Factors used to create the linear model include the PU environment, CE state (on/off), desired waveform, and whether or not the waveform is likely to downgrade based on the PU environment. The residuals of the model are shown in Figure A.5 for two cases. In case 1, the data is left unmodified. In case 2, the outliers, or top clusters utilizing FSK modulation, for waveform configurations 7 and 8 are removed.

The SU residuals prior to removing outliers show a cluster of data not lying on the main regression line. Once the outliers are removed, the resulting residuals do show this cluster. The data is not completely normal; however, it is likely that with an increased sample size this effect would go away.

A.3 Number of Symbols Counted

An output collected that is similar to the number of transmissions received is the number of symbols counted. The number of symbols counted is a reflection of the modulation used for the experiment. For the SU, the CE off waveform configuration should show tight groupings whereas the CE on waveform configuration should show a spectrum of discrete clusters indicative of adaption to the environment by changing the waveform used. For the PU, the symbols counted should show tight groupings. Lower performing

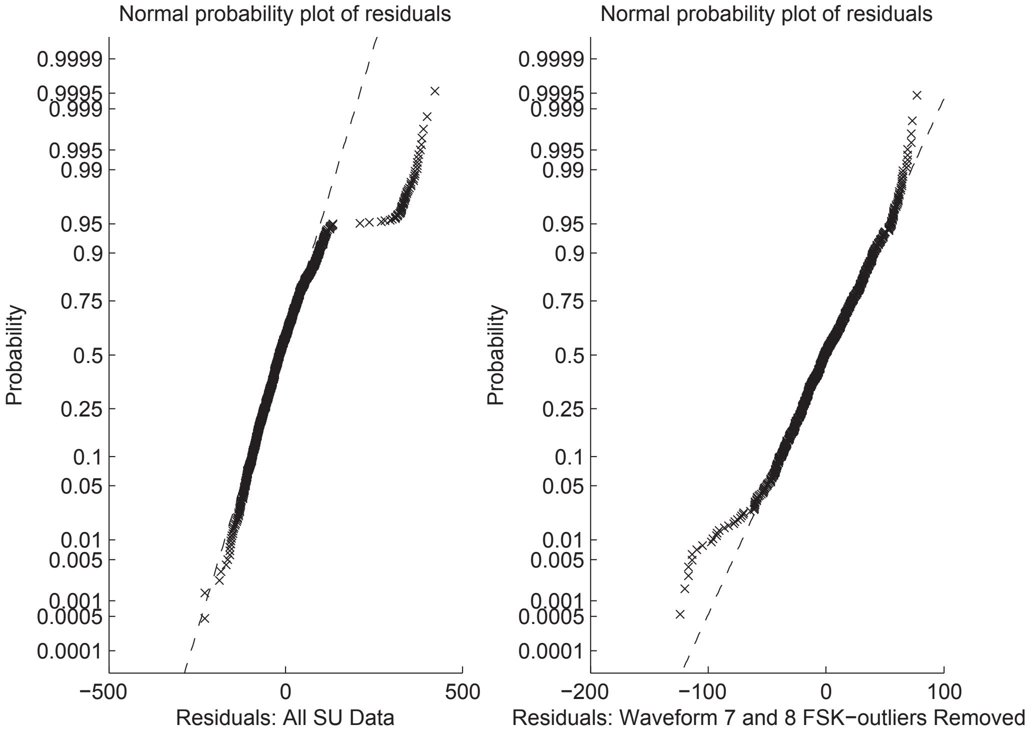


Figure A.5: Residuals of Fitted Linear Model: SU Total Number of Transmissions

waveforms will have a lower symbol count than higher performing waveforms. These expected groupings are in fact found in the data, as shown in Figure A.6.

A.4 Mean Received Signal Strength and Threshold

A key variable that will affect SU and PU performance is the presence of unwanted RF interference. One way to detect RF interference and to evaluate the stability of the RF environment is to chart the mean received signal strength detected and thresholds set by the SU during the spectrum sensor calibration. The spectrum sensor thresholds were calibrated prior to each new experiment. This was done to ensure the spectrum sensor threshold levels were current and accurate. After initial radio configuration, the PU transmitted a known waveform. The SU, with foreknowledge of this known waveform, lowered the threshold in 0.5 dB increments until the PU waveform matched the expected REM. If there are as many

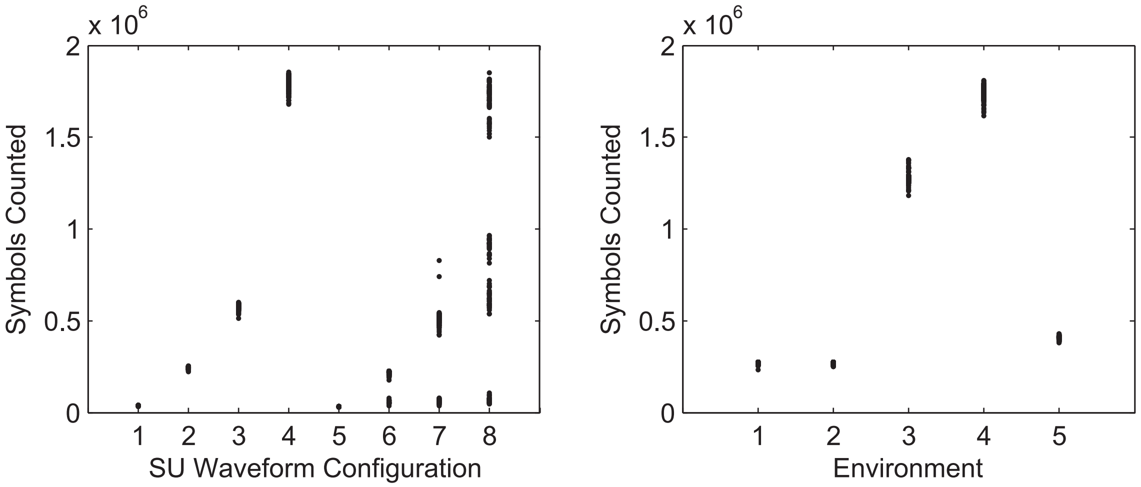


Figure A.6: Scatterplot of Symbols Counted

false REM values as correct REM values, the current threshold value was thrown out and the process repeated. Once a valid threshold was obtained, the process is repeated with a second known signal. Due to received power fluctuations, the SU was calibrated to the type of PU waveform signal (FSK or OFDM). The mean received signal strength is the mean of all power spectral density bins computed using the Fast Fourier Transform (FFT).

The spectrum sensor thresholds for both SU radios are shown in Figure A.7. Note that the threshold for Radio B is higher than the threshold for Radio A. For both, the higher threshold grouping is the FSK threshold, the lower dB threshold grouping is the OFDM threshold. The threshold values show expected behavior, though there is one outlier at -73.5 dB for Radio B.

The mean received signal strength is shown plotted in Figure A.8. The variance is smaller than the threshold data. The mean received signal strength shows expected behavior.

Both the SU and PU data for elapsed time, number of transmissions received, and spectrum sensor thresholds do not show substantial undesired effects. Therefore, there is

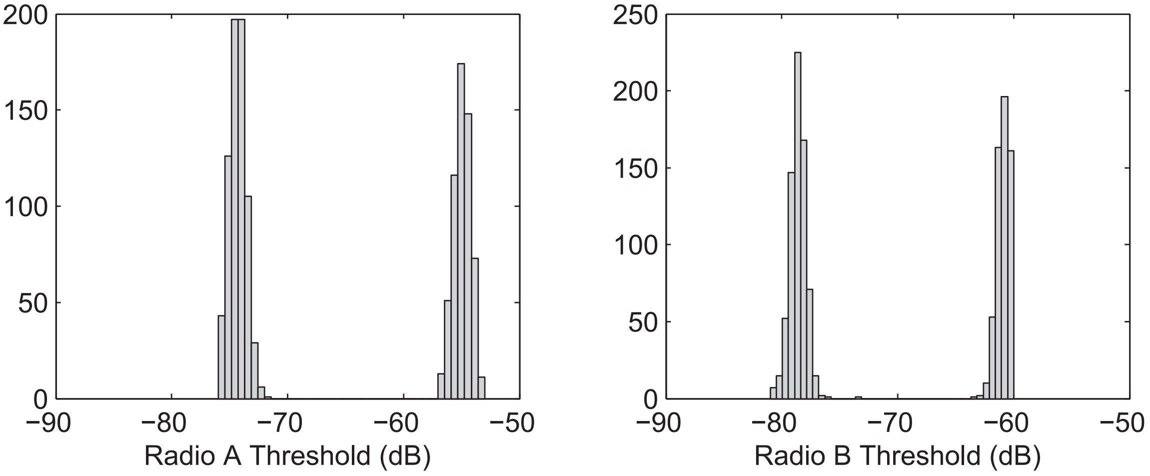


Figure A.7: Histogram of Spectrum Sensor Thresholds

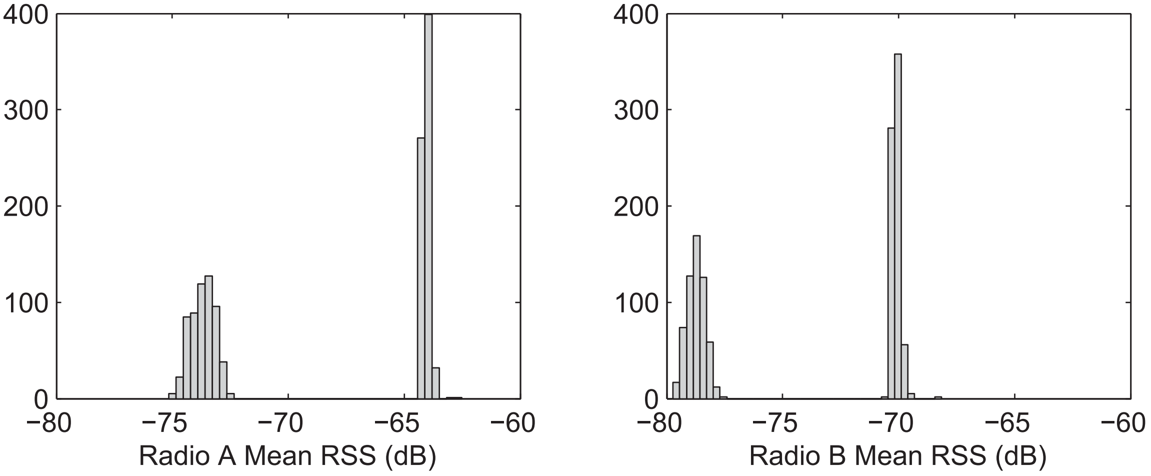


Figure A.8: Histogram of Mean Received Signal Strength (RSS) no evidence that the SUT is not operating as intended, nor that the external RF environment is causing excessive variance in the results.

A.5 Throughput and BER

Next, the primary outputs, throughput and BER, are examined. Boxplots of the raw data are shown in Figure A.9. A linear model was fitted to the data. The predictors used in the linear model included the PU Environment, SU desired waveform, CE status (on/off), experimental order, spectrum sensor thresholds, and elapsed times. The linear model predictors and statistics are shown in Table A.2. Figure A.10 shows the linear model fitted residuals for throughput and BER for both the SU and PU.

Table A.2: Fitted Linear Model Parameters

Predictor	Estimate	Standard Error	t-statistic	p-value
(Intercept)	-1.1804	0.90953	-1.2979	0.19461
ExpNum	-1.6098e-05	7.5581e-06	-2.1299	0.033406
ExpConfig	-0.0043267	0.00014747	-29.339	1.0238e-139
ElapsedTimeP	0.0094172	0.0065644	1.4346	0.15169
ElapsedTimeS	0.0088554	0.013314	0.66512	0.50612
WaveformP	-0.013832	0.062051	-0.22292	0.82364
Environment	0.032127	0.002802	11.466	8.3421e-29
CogEngine	0	0	NaN	NaN
ThreshA	0.00018669	0.0028943	0.064504	0.94858
ThreshB	-0.0020201	0.0029318	-0.68902	0.49096
EnvFactor	0.037239	0.0031466	11.835	1.7814e-30
WF	0.034152	0.001789	19.089	2.9834e-70

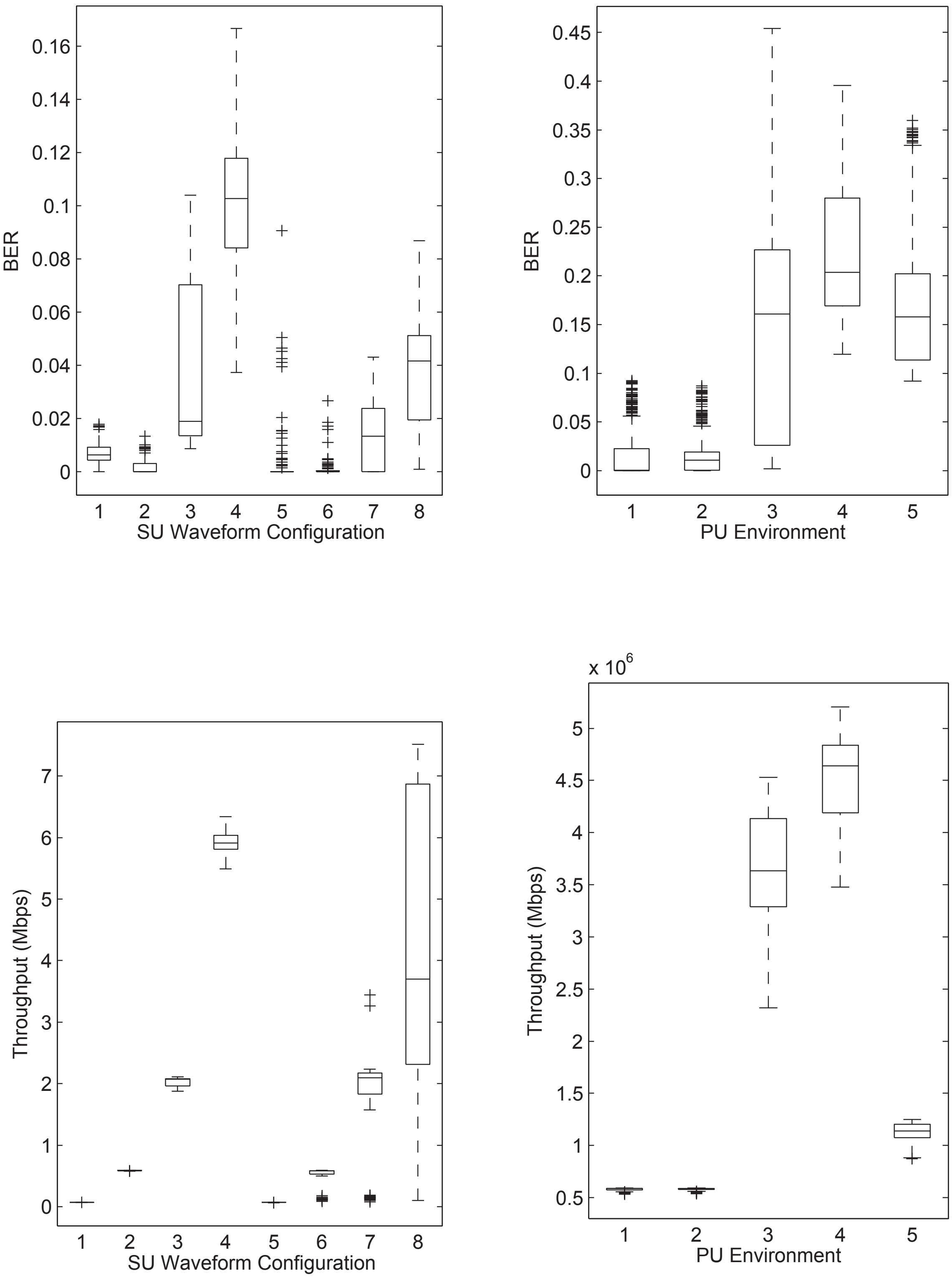


Figure A.9: Experimental Data for Throughput and BER

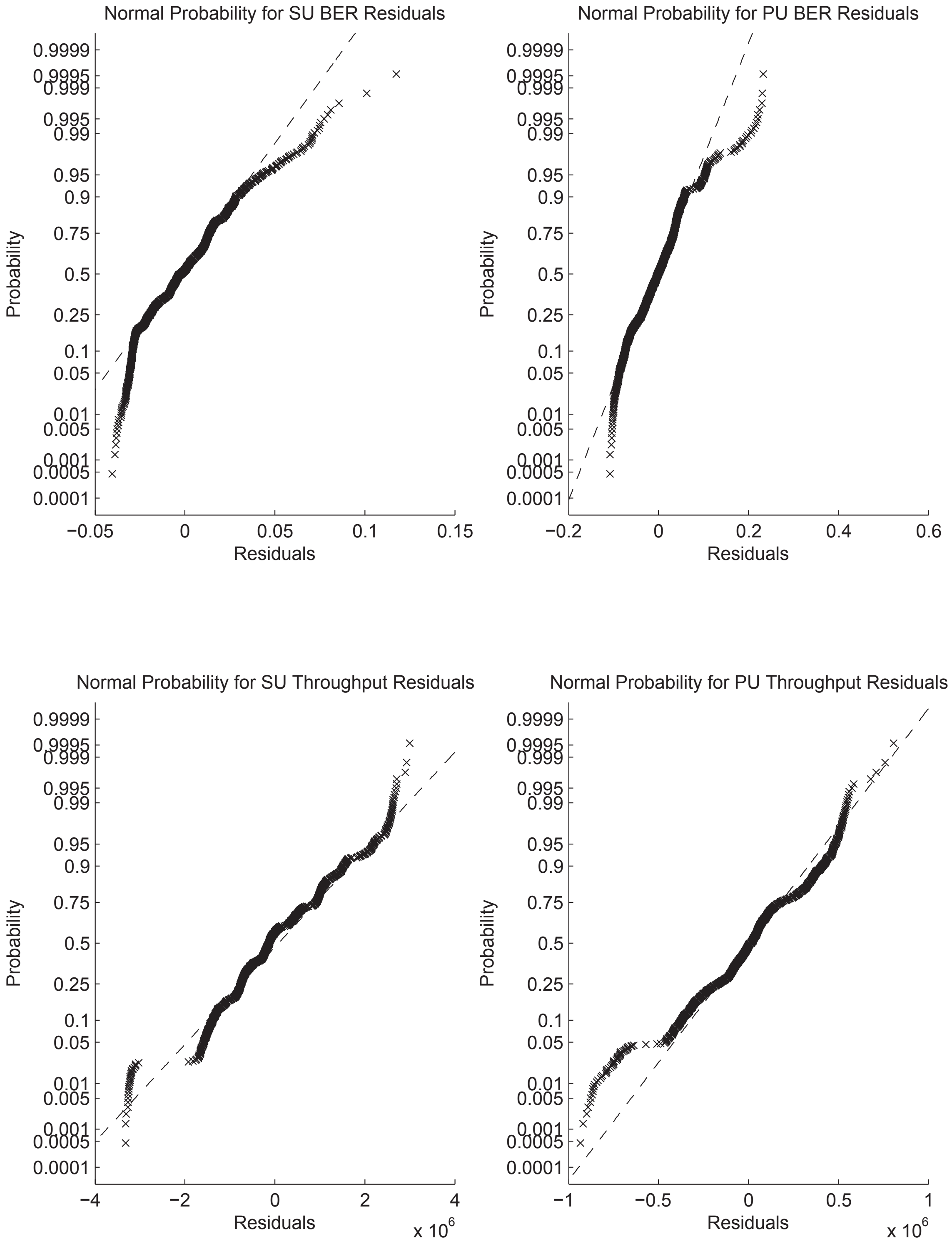


Figure A.10: Experimental Data for Throughput and BER

The residuals generally follow the normal distribution. All residuals show non-ideal distributions at the tails. This could reflect either a unknown factor, a shortage of sample data, or the results of attempting to use a linear model with non-linear predictors. Predictors were added to best capture the non-linear relationship in the SU CE and PU Environments. For example, the SU OFDM-48 behavior with the CE on may vary considerably as shown in Figure A.9. The relationship between the output and the SU waveform configuration is seen to be non-linear. To remedy this, additional predictors were used to clarify the CE on behavior.

Appendix B: Spectrum Sensor Core

An energy-detection based spectrum sensor was developed for use in this research effort. The spectrum sensor was designed to be incorporated into the WARP boards as an IP core. The spectrum sensor may be used on any Xilinx FPGA with sufficient area. This section describes the spectrum sensor developed for this and future ACRO research.

B.1 Background on Energy Detection for Spectrum Sensing

The principle of energy detection is discussed further in [27, 57]. Energy detection is the optimal method when only power measurements are available [57, 58] and is easy to design and implement [27]. In its simplest form, energy detection categorizes measurements above a certain threshold as occupied spectrum as shown in Figure 4.2. When noise statistics are known, three primary means of detecting signals are the *m*-dB criterion, maximum noise level criterion, and the Probability of False Alarm (PFA) criterion [59]. In the *m*-dB criterion, the threshold is simply set at an arbitrary value above the noise floor such as 6 or 10 dB. In the maximum noise level criterion, the threshold is set at the maximum noise level. In the PFA criterion, the threshold is set at a level such that the threshold leads to a specified probability of false alarm. This is discussed further in [27, 60]. There are algorithms that do not need prior knowledge of noise properties, these are Otsu's algorithm and the Recursive One-Side Hypothesis Testing algorithm [61, 62], though these are outside the scope of the spectrum sensor for this research.

The spectrum sensor was designed to apply a threshold to the incoming data to determine if spectrum is occupied or not. Furthermore, the spectrum sensor was designed to accommodate all three energy detection schemes presented above. That is, the spectrum sensor has both a changeable threshold in addition to providing noise statistics (average noise floor and variance). Each FFT output bin corresponds to a spectrum bin. In the output

REM, each bit is a one or zero as it relates to the threshold level for the corresponding frequency. Zeros indicate an open frequency; ones indicate an occupied frequency. The functional block diagram of the spectrum sensor is shown in Figure B.1. The variable *n* corresponds to the number of frequency bins to be used.

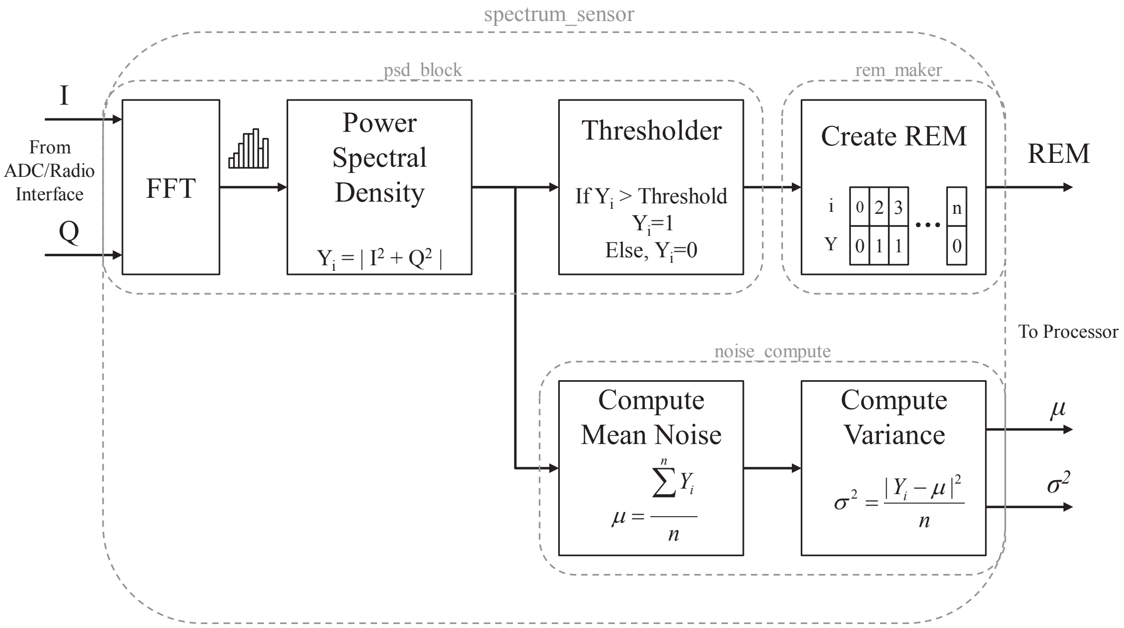


Figure B.1: Spectrum Sensor Functional Block Diagram

B.2 Spectrum Sensor Design Overview

The spectrum sensor was designed to accommodate a range of spectrum sensing needs. To do this, the underlying code was built to respond to a flexible range of inputs. The spectrum sensor has the following characteristics:

- *Reconfigurable FFT Size:* The FFT size is configurable to sizes from 2^6 to 2^{14}. All FFT sizes are powers of 2. The FFT size may be set by setting the FFT write enable (`NFFT_WE`) flag high.

- *14-Bit I/Q Inputs:* The input accepts up to 14 bit signed integers for both I and Q components.

- *Reconfigurable Threshold Level:* The threshold level may be set according to user needs anytime during operation. The threshold level is an unsigned 32-bit integer.

- Variable Size REM The REM output is of variable length from 32 bits to 1024 bits, in 32 bit-increments. That is, there are thirty-two 32-bit output registers that correspond to the output REM.

- *Variable Selection of Spectrum Bins:* The user has control over which FFT-output bins are used to create the REM. For example, a user may be interested in only 64 particular frequencies/bins, though a 1024-point FFT is used. The user may specify the frequency/bin range of interest in contiguous 32-bit chunks.

- *Spectrum Bin Masking (Hopset Mask):* The user may specify certain frequency bins to be excluded from the REM. This is useful for frequency hopping applications when spectrum sensing occurs simultaneously with user transmissions.

- *Noise Floor Average Value:* The average noise floor, or mean received signal strength, is computed using the power spectral density values. This is output as a 32-bit integer.

- *Noise Floor Variance:* The variance of the noise floor, or received signal, is computed and output as a 32-bit integer.

- *Data Valid Check:* The user has control over the spectrum sensor start/stop flag. Furthermore, the user can monitor the output data valid flags such that when a flag is set to high, the data is ready to be read. This prevents the user from using obsolete or redundant data, should the spectrum sensor start flag be set to low or if the FFT determines the output is invalid.

These characteristics may be tied together in an illustration. A user first specifies the desired FFT size, which for this example if $2^8 = 256$. The user desires to use frequency bins 2 to 66, which correspond to a frequency range of 0.3125 to 10.3125 MHz. The REM size is $n = 64$. Frequency bins outside of the range 2 to 66 are not included in the REM, noise, or variance calculations. Furthermore, the user does not seek to use frequency bin 35 due to known interference issues; therefore, the hopset mask is set to discard bin 35.

The schematic of the spectrum sensor is shown in Figure B.2. Sub-components of the spectrum sensor correspond to the dashed line blocks in Figure B.1. The FFT core used in the spectrum sensor is the Xilinx DS260 Fast Fourier Transform v7.1. Other IP cores include a RAM block (Xilinx DS512) in `noise_compute` and multipliers (Xilinx DS255) in `psd_block`.

The inputs of the spectrum sensor are as follows. `xn_re` and `xn_im` refer to the I and Q components of the input signal. The `threshold` and `fft_size` are self-explanatory. `nfft_we` is used as a signal to write the `fft_size`. `num_rems` is the number of REMs to be output, where each REM is 32 bits. Due to implementation constraints, the number of REMs must be 1, 2, 4, 8, 16, or 32. The `bw_mask_mode` is used to set the indexing scheme used to output the REMs. There are three modes. Mode 1 uses all FFT outputs (only valid for $NFFT <= 1024$). Mode 2 outputs only positive frequency components (only valid for $NFFT <= 2048$). Mode 3 allows mixed positive and negative frequency components, which are specified using the next four inputs. The `bw_mask` components correspond to the minimum positive frequency index, maximum positive frequency index, minimum negative frequency index, and maximum frequency index. The `start` input is used to trigger the operation of the spectrum sensor. `start` is tied high for continuous operation. Finally, there are thirty-two 32-bit `hopset_mask` registers which correspond to desired REM output frequency bins.

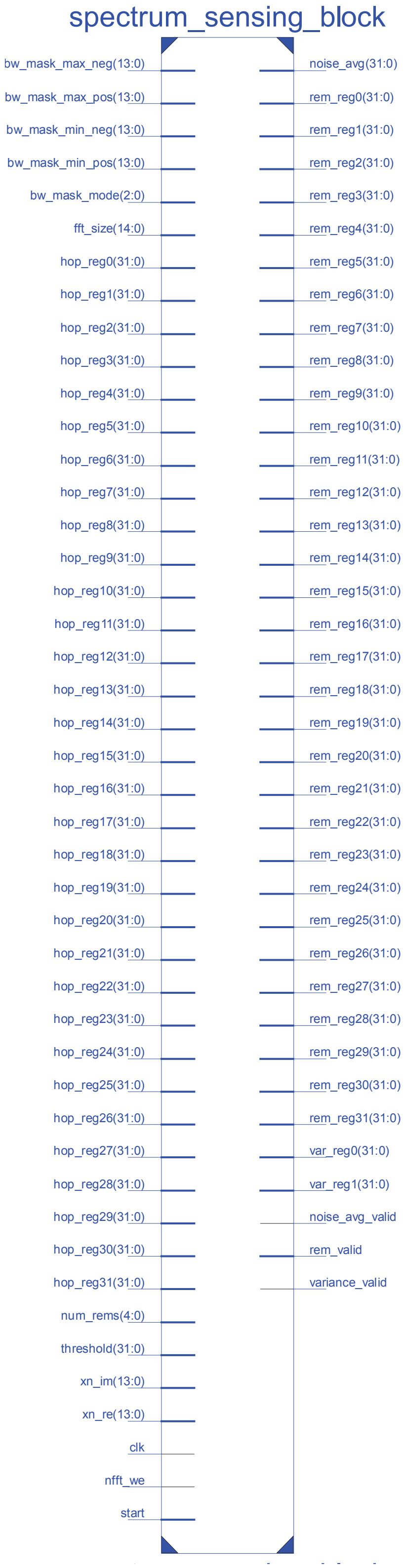


Figure B.2: Spectrum Sensor Top-Level Schematic

The outputs of the spectrum sensor are the average noise (`noise_avg`), variance (`var_reg0` and `var_reg1`) , and thirty-two 32-bit REM registers. Each output also has a valid flag which goes high one clock-cycle prior to the respective output becoming valid. For example, in streaming operation with $NFFT = 256$, the `rem_valid` flag will go high every 256 clock cycles.

B.3 Performance, Area, and Timing

System performance of the spectrum sensor was validated using a Xilinx System Generator testbench. In this testbench, a MATLAB script created the range of inputs, which were then executed in the System Generator testbench module, and the outputs were validated against predicted responses in MATLAB. Following this, system performance was validated on the WARP board for a special case of $NFFT = 256$ with frequency bins [1 64] to create two 32-bit REMs. The spectrum sensor was found to exhibit the desired outputs for the REM, noise, and variance by comparing the results with stored I/Q samples pulled from the board using WARPLab.

The spectrum sensor was designed to use minimum area while preserving performance. The Xilinx FFT core uses a relatively large amount of DSP48s, which allow fast mathematical operations. To implement the design in WARPLab, it was not necessary to change FFT core instance. However, it was necessary to change the FFT core to use logic slices instead of DSP48s for implementation on other WARP designs such as the WARP OFDM reference design. Performance and timing were not affected by this change. For future use, the FFT core may be setup for a specific FFT size in order to save space.

The area summary for the stock spectrum sensor (reconfigurable FFT up to $NFFT = 16384$) is shown in Table B.1. The synthesis was targeted towards the Virtex 4 chip (xc4vfx100-11ff1517) used on the WARP v2 board using a balanced synthesis goal. The synthesis engine was the default choice in Xilinx Integrated Software Environment (ISE) 13.2.

Table B.1: Spectrum Sensor Area Summary

Logic Utilization	Used	Available	Utilization
Number of Slices	4819	42176	11%
Number of Slice Flip Flops	6397	84352	7%
Number of 4-input LUTs	5954	84352	7%
Number of FIFO16/RAMB16s	4	376	1%
Number of DSP48s	31	160	19%

The total duration is proportional to $NFFT$. Each sub-component in the spectrum sensor also takes a time proportional to $NFFT$ to compute its output. In addition, the number of REMs also impacts the total duration. For example, for a $NFFT = 1024$ case with only one 32-bit REM, the total duration will be $1024 - 32 = 992$ clock cycles shorter than if all 32 REMs were used. Table B.2 shows the time from start going high or the first valid input until the corresponding first valid output flag for the REM, noise average, and variance.

The spectrum sensor is designed to be clocked to the I/Q sample rate. Maximum clock rate is 64.4 MHz. This relatively slow clock rate may be increased by fixing one particular logic sequence in the FFT core; this was not done for this research as the clock rate is sufficient to be used on the WARP core. As originally designed, the spectrum sensor was intended for use with small $NFFT$ sizes, which makes it feasible to operate at the I/Q sample rate. However, future revisions of the spectrum sensor should investigate buffering the I/Q inputs and then operating the spectrum sensor at a much higher clock rate in order to shorten the overall latency.

Table B.2: Spectrum Sensor Timing Summary

	REM Output Ready		Noise Avg Output Ready		Variance Output Ready	
NFFT	Clock Cycles	Time (μs)	Clock Cycles	Time (μs)	Clock Cycles	Time (μs)
64	288	7.2	286	7.15	354	8.85
128	502	12.55	500	12.5	632	15.8
256	885	22.125	883	22.075	1143	28.575
512	1675	41.875	1673	41.825	2189	54.725
1024	3210	80.25	3208	80.2	4236	105.9
2048	6304	157.6	6302	157.55	7330	183.25
4096	12447	311.175	12445	311.125	13473	336.825
8192	24756	618.9	24754	618.85	25782	644.55
16384	49331	1233.275	49329	1233.225	50357	1258.925

B.4 WARP Implementation

The spectrum sensor was successfully integrated with the WARP v2 board. The spectrum sensor I/Q inputs were tied directly to the radio bridge outputs. The remaining spectrum sensor inputs and outputs were tied to the 40 MHz PLB. The embedded PowerPC 405 chip on the Virtex 4 ran the software necessary to operate the spectrum sensor inputs and outputs. Besides the *C* software written to integrate the spectrum sensor with the board, additional software was written in MATLAB such that the WARPLab software could control the spectrum sensor from the PC. The user can configure the spectrum sensor using MATLAB, as well as retrieve the REMs, noise, and variance values.

References

[1] SDR Forum, “Cognitive Radio Definitions and Nomenclature,” June 2008. SDRF-06-P-0009-V1.0.0.

[2] Q. Zhao and B. M. Sadler, “A Survey of Dynamic Spectrum Access,” *IEEE Signal Processing Magazine*, vol. 24, no. 3, pp. 79–89, 2007.

[3] Office of the Air Force Chief Scientist, “Report on Technology Horizons: A Vision for Air Force Science & Technology During 2010-2030,” 2010.

[4] P. Steenkiste, D. Sicker, G. Minden, and R. Dipankar, “Future Directions in Cognitive Radio Network Research,” in *NSF Workshop Report, 9-10 March 2009*, p. 3, National Science Foundation, 2009.

[5] Y. Zhao and L. Morales-Tirado, “Cognitive Radio: Forging Ahead from Concept, Testbed to Large-Scale Deployment,” *Journal of Communications*, vol. 7, no. 7, pp. 514–523, 2012.

[6] J. Mitola III and G. Q. Maguire Jr, “Cognitive Radio: Making Software Radios More Personal,” *IEEE Personal Communications*, vol. 6, no. 4, pp. 1, 13–18, 1999.

[7] J. Boyd, “A Discourse on Winning and Losing.” A collection of unpublished briefings and essays. Air University Library, Document No. M-U 43947, August 1987.

[8] Wikipedia, “Cognitive Network — Wikipedia, The Free Encyclopedia.” Accessed 22 October 2013. Online: http://en.wikipedia.org/w/index.php?title=Cognitive_network&oldid=545105654, 2013.

[9] R. Thomas, L. DaSilva, and A. MacKenzie, “Cognitive Networks,” in *New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. 2005 First IEEE International Symposium on*, pp. 352–360, 2005.

[10] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, “Cognitive Networks: Adaptation And Learning To Achieve End-to-End Performance Objectives,” *IEEE Communications Magazine*, vol. 44, no. 12, pp. 51–57, 2006.

[11] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, *Cognitive Networks*. Springer, 2007.

[12] M. Masonta, M. Mzyece, and N. Ntlatlapa, “Spectrum Decision in Cognitive Radio Networks: A Survey,” *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1088–1107, 2013.

[13] S. Haykin, “Cognitive Radio: Brain-Empowered Wireless Communications,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

[14] IEEE Standards Association, "IEEE DySPAN Standards Committee (DySPAN-SC." Accessed 22 Dec 2013. Online: http://www.dyspan-sc.org/.

[15] M. Murroni, R. Prasad, P. Marques, B. Bochow, D. Noguet, C. Sun, K. Moessner, and H. Harada, "IEEE 1900.6: Spectrum Sensing Interfaces and Data Structures for Dynamic Spectrum Access and Other Advanced Radio Communication Systems Standard: Technical Aspects and Future Outlook," *IEEE Communications Magazine*, vol. 49, no. 12, pp. 118–127, 2011.

[16] R. Venkatesha Prasad, P. Pawelczak, J. Hoffmeyer, and H. Berger, "Cognitive Functionality in Next Generation Wireless Networks: Standardization Efforts," *IEEE Communications Magazine*, vol. 46, no. 4, pp. 72–78, 2008.

[17] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, "IEEE 802.22: The First Worldwide Wireless Standard Based On Cognitive Radios," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005.*, pp. 328–337, IEEE, 2005.

[18] "IEEE 802.22 Working Group on Wireless Regional Area Networks." Accessed 22 Dec 2013. Online: http://www.ieee802.org/22/.

[19] C. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. Shellhammer, and W. Caldwell, "IEEE 802.22: The First Cognitive Radio Wireless Regional Area Network Standard," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130–138, 2009.

[20] "IEEE Standard for Spectrum Sensing Interfaces and Data Structures for Dynamic Spectrum Access and other Advanced Radio Communication Systems.," *IEEE Std 1900.6-2011*, pp. 1–168, 2011.

[21] Federal Communications Commission, "FCC Report and Order 05-57." Cognitive Radio Technologies Proceeding ET Docket No. 03-108, March 2005.

[22] Federal Communications Commission, "Establishment of Interference Temperature Metric to Quantify and Manage Interference and to Expand Available Unlicensed Operation in Certain Fixed Mobile and Satellite Frequency Bands." ET Docket 03-289, Notice of Inquiry and Proposed Rulemaking, 2003.

[23] M. McHenry, E. Livsics, T. Nguyen, and N. Majumdar, "XG Dynamic Spectrum Access Field Test Results," *IEEE Communications Magazine*, vol. 45, no. 6, pp. 51–57, 2007.

[24] Y. Zhao, S. Mao, J. O. Neel, and J. H. Reed, "Performance Evaluation of Cognitive Radios: Metrics, Utility Functions, and Methodology," *Proceedings of the IEEE*, vol. 97, no. 4, pp. 642–659, 2009.

[25] R. K. McLean, "An Architecture for Coexistence with Multiple Users in Frequency Hopping Cognitive Radio Networks," Masters Thesis, Air Force Institute of Technology, March 2012.

[26] B. Wang and K. J. R. Liu, "Advances in Cognitive Radio Networks: A Survey," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 5–23, 2011.

[27] T. Yucek and H. Arslan, "A Survey of Spectrum Sensing Algorithms for Cognitive Radio Applications," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 116–130, 2009.

[28] J. Ma, G. Y. Li, and B. H. Juang, "Signal Processing in Cognitive Radio," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 805–823, 2009.

[29] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative Spectrum Sensing in Cognitive Radio Networks: A Survey," *Physical Communication*, vol. 4, no. 1, pp. 40–62, 2011.

[30] Y. Zhao, L. Morales, J. Gaeddert, K. K. Bae, J.-S. Um, and J. H. Reed, "Applying Radio Environment Maps to Cognitive Wireless Regional Area Networks," in *2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 115–118, IEEE, 2007.

[31] A. Wyglinksi, M. Nekovee, and Y. T. Hou, *Cognitive Radio Communications and Networks: Principles and Practice*. Elsevier, 2010.

[32] R. K. Mclean, M. D. Silvius, K. M. Hopkinson, B. N. Flatley, E. S. Hennessey, C. C. Medve, J. J. Thompson, M. R. Tolson, and C. V. Dalton, "An Architecture for Coexistence with Multiple Users in Frequency Hopping Cognitive Radio Networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 3, p. To appear, 2014.

[33] S. M. Mishra, D. Cabric, C. Chang, D. Willkomm, B. Van Schewick, S. Wolisz, and B. Brodersen, "A Real Time Cognitive Radio Testbed for Physical and Link Layer Experiments," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 562–567, 2005.

[34] T. R. Newman and T. Bose, "A Cognitive Radio Network Testbed for Wireless Communication and Signal Processing Education," in *IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop (DSP/SPE)*, pp. 757–761, IEEE, 2009.

[35] Y. Kim, K. Taylor, C. Dunbar, B. Walker, and P. Mundur, "Reality vs Emulation: Running Real Mobility Traces on a Mobile Wireless Testbed," in *Proceedings of the 3rd ACM international workshop on MobiArch*, pp. 23–28, ACM, 2011.

[36] G. Judd and P. Steenkiste, "Using Emulation to Understand and Improve Wireless Networks and Applications," in *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation*, vol. 2 of *NSDI'05*, pp. 203–216, 2005.

[37] A. Sanchez, I. Moerman, S. Bouckaert, D. Willkomm, J. Hauer, N. Michailow, G. Fettweis, L. Dasilva, J. Tallon, and S. Pollin, "Testbed Federation: An Approach For Experimentation-Driven Research in Cognitive Radios and Cognitive Networking," in *Future Network & Mobile Summit*, pp. 1–9, IEEE, 2011.

[38] D. Tandur, J. Duplicy, K. Arshad, D. Depierre, K. Moessner, J. Lehtomaki, K. Briggs, L. Goncalves, and A. Gameiro, "Cognitive Radio Systems Evaluation: Measurement, Modeling, and Emulation Approach," *IEEE Vehicular Technology Magazine*, vol. 7, no. 2, pp. 77–84, 2012.

[39] K. Borries, X. Wang, G. Judd, P. Steenkiste, and D. Stancil, "Experience with a Wireless Network Testbed Based on Signal Propagation Emulation," in *European Wireless Conference (EW)*, pp. 833–840, 2010.

[40] J. Zhou, Z. Ji, M. Varshney, Z. Xu, Y. Yang, M. Marina, and R. Bagrodia, "WHYNET: a Hybrid Testbed for Large-Scale, Heterogeneous and Adaptive Wireless Networks," in *Proceedings of the 1st International Workshop on Wireless network Testbeds, Experimental Evaluation & Characterization (WiNTECH)*, pp. 111–112, 2006.

[41] D. Raychaudhuri, I. Seskar, M. Ott, S. Ganu, K. Ramachandran, H. Kremo, R. Siracusa, H. Liu, and M. Singh, "Overview of the ORBIT Radio Grid Testbed for Evaluation of Next-Generation Wireless Network Protocols," in *IEEE Wireless Communications and Networking Conference*, vol. 3, pp. 1664–1669, 2005.

[42] T. Newman, S. Shajedul Hasan, D. DePoy, T. Bose, and J. Reed, "Designing and Deploying a Building-Wide Cognitive Radio Network Testbed," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 106–112, 2010.

[43] P. De, A. Raniwala, S. Sharma, and T.-c. Chiueh, "MiNT: A Miniaturized Network Testbed for Mobile Wireless Research," in *INFOCOM 2005 Proceedings. 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 4, pp. 2731–2742, IEEE, 2005.

[44] Echo Ridge, LLC, *DYnamic Spectrum Environment Emulator (DYSE) User Guide 4.1.1*, 2013.

[45] I. Broustis, J. Eriksson, S. Krishnamurthy, and M. Faloutsos, "A Blueprint for a Manageable and Affordable Wireless Testbed: Design, Pitfalls and Lessons Learned," in *3rd International Conference on Testbeds and Research Infrastructure for the Development of Networks and Communities (TridentCom)*, pp. 1–6, 2007.

[46] P. J. Kolodzy, "Interference Temperature: a Metric for Dynamic Spectrum Utilization," *International Journal of Network Management*, vol. 16, no. 2, pp. 103–113, 2006.

[47] T. C. Clancy, "Formalizing the Interference Temperature Model," *Wireless Communications and Mobile Computing*, vol. 7, no. 9, pp. 1077–1086, 2007.

[48] T. C. Clancy and W. Arbaugh, "Measuring Interference Temperature," in *Virginia Tech Symposium on Wireless Personal Communications*, pp. 1–10, 2006.

[49] "CREW Project: Cognitive Radio Experimentation World." Accessed 23 Dec 2013. Online: http://www.crew-project.eu/.

[50] C. Dietrich, E. Wolfe, and G. Vanhoy, "Cognitive Radio Testing Using Psychometric Approaches: Applicability and Proof of Concept Study," *Analog Integrated Circuits and Signal Processing*, vol. 73, no. 2, pp. 627–636, 2012.

[51] Ettus Research, "USRP Overview." Accessed 3 Feb 2014. Online: http://www.ettus.com/home.

[52] "Mango Communications: WARP FPGA Board Hardware Version 2.2." Accessed 25 Oct 2013. Online: http://mangocomm.com/products/boards/warp-fpga-board-v2.

[53] "WARP Project." http://warpproject.org.

[54] Nutaq, "Cognitive Radio Product Overview, howpubished=Accessed 3 Feb 2014. Online: http://nutaq.com/en/applications/wireless/cognitive-radio."

[55] "Standard Performance Evaluation Corporation." Accessed 23 Dec 2013. Online: http://www.spec.org/.

[56] "WARP Project: WARPLab 7." Accessed 25 Oct 2013. Online: http://warpproject.org/trac/wiki/WARPLab.

[57] H. Urkowitz, "Energy Detection of Unknown Deterministic Signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[58] R. Tandra and A. Sahai, "Fundamental Limits on Detection in Low SNR Under Noise Uncertainty," in *International Conference on Wireless Networks, Communications and Mobile Computing*, vol. 1, pp. 464–469, IEEE, 2005.

[59] M. Lopez-Benitez and F. Casadevall, "Methodological Aspects of Spectrum Occupancy Evaluation in the Context of Cognitive Radio," in *2009 European Wireless Conference*, pp. 199–204, 2009.

[60] D. Cabric, A. Tkachenko, and R. W. Brodersen, "Experimental Study of Spectrum Sensing Based on Energy Detection and Network Cooperation," in *Proceedings of the First International Workshop on Technology and Policy for Accessing Spectrum*, p. 12, ACM, 2006.

[61] D. Datla, A. M. Wyglinski, and G. J. Minden, "A Spectrum Surveying Framework for Dynamic Spectrum Access Networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4158–4168, 2009.

[62] T. Hanninen, J. Vartiainen, M. Juntti, and M. Raustia, “Implementation of Spectrum Sensing on Wireless Open-Access Research Platform,” in *3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, pp. 1–5, IEEE, 2010.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From — To)*
27–03–2014	Master's Thesis	Oct 2013–Mar 2014

4. TITLE AND SUBTITLE
A Test Methodology for Evaluating Cognitive Radio Systems

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Thompson, Jared J., Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENG-14-M-77

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Vasu Chakravarthy
2241 Avionics Circle
WPAFB, OH 45433
vasu.chakravarthy@us.af.mil
937-528-8269

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RYWE

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES
This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

The cognitive radio field currently lacks a standardized test methodology that is repeatable, flexible, and effective across multiple cognitive radio architectures. Furthermore, the cognitive radio field lacks a suitable framework that allows testing of an integrated cognitive radio system and not solely specific components. This research presents a cognitive radio test methodology, known as CRATM, to address these issues. CRATM proposes to use behavior-based testing, in which cognition may be measured by evaluating both primary user and secondary user performance. Data on behavior-based testing is collected and evaluated. Additionally, a unique means of measuring secondary user interference to the primary user is employed by direct measurement of primary user performance. A secondary user pair and primary user radio pair are implemented using the Wireless Open-Access Research platform and WARPLab software running in MATLAB. The primary user is used to create five distinct radio frequency environments utilizing narrowband, wideband, and non-contiguous waveforms. The secondary user response to the primary user created environments is measured. The secondary user implements a simple cognitive engine that incorporates energy-detection spectrum sensing. The effect of the cognitive engine on both secondary user and primary user performance is measured and evaluated.

15. SUBJECT TERMS
cognitive radio, test methodology, behavior-based, primary user, secondary user

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Kenneth Hopkinson (ENG)
					19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	123	(937) 255-3636 x4579 kenneth.hopkinson@afit.edu