A Test Methodology for Evaluating Cognitive Radio Systems

Jared J. Thompson, *Student Member, IEEE*, Kenneth M. Hopkinson, *Senior Member, IEEE*, and Mark D. Silvius, *Member, IEEE*

***Abstract*—The cognitive radio (CR) field currently lacks a standardized end-user test methodology that is repeatable, flexible, and effective across multiple CR architectures. Furthermore, the CR field lacks a suitable device-agnostic framework that allows testing of an integrated CR system and not solely specific components. This paper presents a CR test methodology, known as the Cognitive RAdio Test Methodology (CRATM), to address these issues. CRATM proposes to use behavior-based testing in which cognition may be measured by evaluating both primary user (PU) and secondary user (SU) (i.e., the CR under test) performance. Data on behavior-based testing are collected and evaluated. An SU pair and a PU radio pair are implemented using the Wireless Open-Access Research Platform (WARP) and the WARPLab software running in MATLAB. The PU is used to create five distinct radio-frequency environments utilizing narrow-band, wideband, and noncontiguous waveforms. The SU response to the PU created environments is measured. The SU implements a simple cognitive engine (CE) that incorporates energy-detection spectrum sensing. The effect of the CE on both SU and PU performance is measured and evaluated.**

***Index Terms*—Cognitive radio, behavior-based testing, WARP.**

I. Introduction

FIRST introduced at the turn of the century, CR refers to the emergence of technology that combines software defined radio (SDR), dynamic spectrum access (DSA), networking, and artificial intelligence (AI) techniques. The primary benefit of using CR is that it addresses the issue of spectrum congestion and utilization through its use of DSA. CR helps mitigate spectrum congestion by intelligently utilizing unused portions of the spectrum, which in turn improves network operation and improves end-user data rates. Alternatively, CR may also help to efficiently utilize used portions of the spectrum.

Because CR spans several disciplines, no standardized test methodology has yet been developed for use in test and evaluation of CRs [1]. There is also widespread disagreement on what exactly a CR is or what it is required to be. Furthermore, because CR contains AI elements, 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. However, the performance and behavior of a CR, to include complex interactions between hardware and software, must be well understood if CRs are to be certified and used outside of the laboratory. *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.*

The objective of this research is to develop an end-user CR test methodology that is repeatable, flexible, and effective across multiple CR architectures. First, the test framework should not be tied to a specific CR architecture or technology. In other words, the test framework is device agnostic such that any device capable of transmitting via RF may be tested. Secondly, this test framework is approached from the perspective of an end-user; that is, the performance and specifications of components are secondary to overall performance. Finally, this framework should enable the identification of emergent behavior among complex CR systems.

This article presents the initial development and validation of CRATM to meet the research objective. To accomplish this, CRATM uses behavior-based testing. Behavior-based testing measures the two-way effects between the SU and PU to evaluate cognition performance of the SU. In accordance with the objective to have an end-user test methodology and to identify emergent behavior, CRATM is intended to only apply to real-world physical CR devices and not to theoretical or simulated CR devices. To control the test environment, the use of an emulated RF environment is proposed. However, this is not used in data collection for this research.

CRATM is a top-down systems approach to testing and evaluating CRs. The CR device or CRS under test, along with its sub-components, is treated as a black box. Any RF-capable device may be tested using CRATM. The device does not have to be cognitive to be tested using CRATM. However, the expectation is that devices with greater intelligence will perform better. CRATM does not form a cognitive model of the device under test; only performance data is used to evaluate the utility of the device under test.

This article goes far beyond early results, which appeared in SDR-WInnComm 2014 [2]. While the objective remains the same, the CRATM framework is better defined and more complete in this article. Extensive bench test results also go far beyond the minimal Matlab-based simulations that appeared earlier.

Manuscript received October 3, 2014; revised March 9, 2015; accepted June 16, 2015. Date of publication July 7, 2015; date of current version November 9, 2015. This work was supported in part by the Air Force Research Laboratory (AFRL/RYWE and AFRL/RITE) and in part by the Air Force Office of Scientific Research. The associate editor coordinating the review of this paper and approving it for publication was S. Mukherjee.

The authors are with the Department of Electrical and Computer Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH 45433 USA (e-mail: jared.thompson.6@us.af.mil; kenneth.hopkinson@afit.edu; mark.silvius@us.af.mil).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2015.2452268

U.S. Government work not protected by U.S. copyright.

II. Related Work

The CR field is diverse and has few formal systemic categorizations and classifications. However, the boundaries of a system must be well understood if it is to be effectively evaluated. This section presents an assessment of the CR field and how it is classified in terms of definitions and architectures. After that, testbeds, performance metrics, and test methodologies for CRs are presented. The information presented in this section is intended to provide context for the developed test methodology.

A. Cognitive Radio Definitions and Architectures

There are several levels of CR devices. For this research, a CR is a single device, which may be used in a cognitive radio network (CRN) [3]. A CRN is a collection of cognitive and non-cognitive devices that may also feature cognition on the network layers. By contrast, a cognitive network (CN) requires cognition at the network level, but does not require the use of CRs [4]. A CRS is used in this research as an all-encompassing term to describe either a CR or a cognitive radio network (CRN), similar to the definition of a CRS in [5]. This research does not directly address CNs.

Definitions answer "what does a CR do?" as opposed to architectures, which answers "how does a CR work?". In practice, definitions are best applied only when trying to determine if a device is actually cognitive per a definition. This research does not seek to apply specific CR definitions to evaluate cognition. By contrast, architectures generally enable a continuum of cognition where the actual cognition of a device becomes secondary to overall system performance. Thus, this research does not seek to define or clarify a definition of a cognitive radio. Instead, this research seeks to encompass a wide range of architectures, which are derived from definitions. 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 the depth of cognitive functionality, situation awareness, and where cognition takes place [6]. A summary of cognitive radio definitions and analysis of differences is also found in [6].

The architecture of a CR directly affects how it is implemented, and therefore tested. All CR architectures are governed by the Observe-Orient-Decide-Act (OODA) loop [7], [8]. The OODA loop, as a cognition cycle, forms the simplest possible framework for understanding the reaction of the CR to the environment. 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 the *act* stage corresponds to the implementation of the decision and performing normal operations. Real world CR architectural features can be traced back to how CRSs implement the OODA loop. From a top-level testing perspective, a test methodology should enable a CRS to freely execute all stages of the OODA loop. In practice, a CRS will employ far more complex cognition cycles than a simple OODA loop. This research seeks to find a top-down, architecture agnostic approach to evaluate CRSs. Therefore, the OODA loop will be used as a high-level representation of cognition in this research as the goal is not to build a cognitive model of CRSs but to evaluate the performance of CRSs.

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, which is usually either centralized or distributed. For example, the IEEE 802.22 architecture utilizes a central base station to distribute spectrum knowledge to nodes, which in turn feed localized spectrum sensing information back to the base station. The reader is referred to [9] or [10] for more information on the IEEE 802.22 architecture.

CRSs may use spectrum in one of two ways [11]. The first is to utilize spectrum overlay. Spectrum overlay is the basis for Opportunistic Spectrum Access (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. Both spectrum overlay and underlay will likely be used by CRSs.

The test methodology developed in this research seeks to avoid tying itself to any particular CR definition or architecture and instead aims to be flexible enough to accommodate all definitions and architectures.

B. 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 [8]. Of interest for this research are testbeds that enable flexible testing of real-world devices and offer control over the RF environment; that is, the testbed should not be tied to specific physical system under test (SUT) hardware. Testbeds 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 [12]. To offer control over the RF environment, testbeds are largely either over-the-air ranges (such as in an anechoic chamber) or emulated environments, though at least one group has taken a hybrid approach combining over the air transmission with a virtual emulated environment within an anechoic chamber [13].

To emulate the RF environment, there are three main approaches [14]. 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 [15], WHYNET [14], BEE2 [12], and the DYSE [16].

C. 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 [6] provide an exhaustive treatment of performance metrics for CRs.

1) Spectrum Sensing: Research on spectrum sensing has focused on the ability to detect a PU. The established performance metrics for directly evaluating spectrum sensing are Probability of Primary User Detection [8], [12] and Time to Detect a Primary User [12].

2) Cognitive Radio Artificial Intelligence: For CRSs, Zhao *et al.* [17] 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 [18].

3) 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 for the CRS. These metrics are well established and not discussed further here.

4) Integrated Metrics: Integrated metrics combine both the ability of a SU to sense a PU and to tailor an appropriate response. These metrics include Interference Temperature [19]–[21], Time to Evacuate a Channel [22], and Probability of Collision [11].

D. Cognitive Radio Test Methodologies

A test methodology describes the overall procedure in how to evaluate a CR or CRS. Two test methodologies that can be used to test a mostly complete CRS are presented here. Most existing test methodologies only focus on one sub-component of the CRS and are not suitable for evaluating a complete CRS.

1) Radio Environment Map Scenario Driven Testing Approach: Zhao *et al.* propose the use of radio environment map (REM)s as the means of testing CR systems [23]. These (REM)s 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 [9], [10] and to general case DSA environments, such that a benchmark suite is used to test CRs over a full range of situations [6].

Also in [6], 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 approach presented in [6] and [23] is well suited for a centralized topology similar to IEEE 802.22, but is not necessarily useful for other non-centralized CR architectures. This is primarily because the test methodology revolves around evaluating the CE and how it processes REMs. Actual integrated and RF performance is not measured.

2) Psychometric Approach: Since radio cognition may be viewed as an analogue to human cognition, it is proposed by Dietrich *et al.* to evaluate CR using approaches developed for testing human cognition [24]. Psychometric testing, as described by Dietrich *et al.*, utilizes item response models (IRMs) to assess CR performance. 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 latent 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 integrated system. As with REM-SDT testing, the psychometric approach as presented by [24] is geared towards an IEEE 802.22-style approach as it focuses on evaluating the CE.

III. Proposed Test Methodology

The proposed test methodology, CRATM, developed by this research is presented in this section. A methodology for evaluating CRATM is presented in the next section, followed by the results in the subsequent section.

A. Overview

CRATM is designed to be an over-arching framework that applies to evaluating a CRS. CRATM applies benchmarks to stimulate behaviors in the SUT (e.g. a CRS). The CR response to the stimuli is then measured using performance metrics. 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 to computer benchmarks such as the Standard Performance Evaluation Corporation (SPEC) suite [25]. 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 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 as proposed by [12], [17].

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. Virtual PUs are created within the emulated environment, negating the need for physical PU devices. This enables rapid changeover between benchmarks. The use of an emulated RF environment or PU is not required for CRATM, though the flexibility in running different benchmarks is lost.

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. Flexibility is provided to the end-user while enabling consistent results.

CRATM is extensible to a wide range of radios. The chief limitation is the ability of the emulated environment (if using the emulated environment) or the ability of the PU devices to create relevant environments. In an emulated environment, the only required modification to the CR under test is the substitution of an RF-cable from the CR RF-front end to the emulation machine. Then, the ability to test the CR is the limitations of the emulation machine. With current technology, these limitations include processing power and signal latency which limit the PU created environment, the number of nodes, and overall bandwidth. In a physical environment, ideally any CR or CRS can be used with no modifications. In practice, constraints such as frequency limitations, range, and RF-environment control may limit effectiveness. Furthermore, it may be expensive to set up multiple PU networks across a broad spectrum in order to fully test a CRS. In the course of this research, using CRATM with both emulated and physical techniques were investigated and found to be feasible.

B. Cognitive Radio Behavior-Based Evaluation

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

The two key elements of this definition are that the device improves performance 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 improving performance and avoiding causing interference. Note that this definition only applies to DSA environments in which spectrum adaptation is essential.

CRATM treats cognition as occurring on a continuum as opposed to occurring in discrete steps. Testers can then 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 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*. 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 CRSs can evolve from current systems without requiring a discrete jump to CRS capability.

Behavior-based testing builds upon the techniques of psychometric testing presented in [24]. 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. Whereas psychometric testing seeks to build a cognitive model to understand the CR, CRATM seeks to evaluate the performance of the CR without knowledge or direct evaluation of the CE. CRATM evaluates total system performance, including emergent behavior, rather than solely evaluating CE cognition. This distinction reflects the goal of this research to test actual physical systems and not software models of theoretical ideas. Furthermore, CRATM takes a systems engineering approach in that the system is only as strong as its weakest sub-system; for instance, a top-notch CE fielded with a sub-par transceiver will yield worse real-world performance than an lesser CE fielded with a more capable transceiver.

C. Performance Metrics

Based on the previous section, the behavioral based performance metrics available for use fall into two categories:

improving performance and avoiding causing interference. The baseline performance metrics for use in measuring both improving performance and avoiding causing interference are to measure throughput and/or BER. For instance, the SU throughput may be used as a metric of improving performance while PU BER may be used as a metric of how well the SU avoids causing interference.

Specific testers may have need for a specific performance metric besides the throughput and BER of the PU and SU. Just as with the SPEC 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.

D. Benchmarks

Benchmarks allow CRS testing to be flexible, meaningful, 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 benchmark signal of interest to detect lies outside of the capabilities of the CRS receiver test device.

Benchmark scenarios should be programmed to be time-driven as opposed to task-driven to allow system adaptation and learning to occur. 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.

Specific classes of benchmarks enable a wide variety of CRS architectures to be evaluated. 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 proposed to be split into eleven characteristics, which combined, form a unique benchmark. The first 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*.

1) Cognitive Radio Type: Cognitive radio type describes whether the SUT is a specific CR device or a CRN.

2) Number of Nodes: If the SUT is a CRN, the number of nodes must be specified.

3) Topology: The topology drives the type of scenario implemented. The topologies are distributed, centralized, and combined.

4) Radio Frequency Environment: The RF environment describes the emulated RF background environment. The options may include urban, rural, suburban, home, and battlefield environments. The RF environment may also be a combination of these environments.

TABLE I
EXAMPLE CRATM BENCHMARK

	Characteristic	Description
1	CR Type	CRN
1a	Number of Nodes	16
2	Topology	Distributed
3	RF Environment	Rural
4	Test Band	2.4 - 2.5 GHz
5	Motion	Static
6	Performance Metrics	SU - Throughput, PU - BER
7a	Timeframe	15 minutes
7b	Sampling Interval	1, 5, 15 minutes
8	Geographical Information	4x4 node centered at 39.5N, 84W, 0 m
9	Power Constraints	-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.

5) Test Band: The test band lays out the overall frequency range eligible to be reached by the SUT.

6) 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.

7) Performance Metrics: The performance metrics of interest are defined under this characteristic.

8) Timeframe and Sampling Intervals: The overall test duration and sampling intervals are defined in this characteristic.

9) Geographical Information: The locations of each node plus virtualized SU units are described.

10) Power Constraints: The maximum power level provides an upper bound on the CR power levels to reflect real-world regulatory and environmental limitations.

11) 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.

12) 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 I. Benchmarks may be constructed for various CRS taxonomies or paradigms, such as those that utilize spectrum underlay, spectrum overlay, or a combination of both with the PU. Also, benchmarks may be constructed to test relevant situations such as the hidden-node problem.

E. 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.

IV. Methodology

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 key component of CRATM, that SU cognition may be determined by measuring overall system 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 is tested by comparing CR prototypes with non-CR prototypes with the expectation that all other factors being equal, prototypes with cognitive capabilities (i.e. the ability to implement the OODA cycle and sense the spectrum, react to it, and change waveforms as required) will perform better than devices without cognitive capabilities.

A. Approach

Given a basic CRS to test, a PU network is implemented alongside a SU network. The PU is used to create the environment sensed by the SU. The SUT (the CR) is 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 flexibility are added to the SU. Cognitive flexibility for this experiment is increased waveform choices and bandwidth for the SU. It is assumed for this experiment that greater flexibility in waveform and bandwidth is correlated with greater cognitive ability.

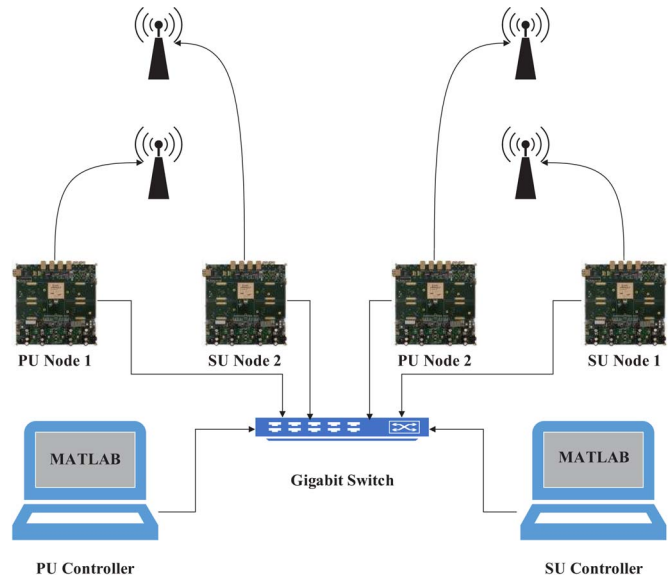

Fig. 1. Experimental setup: Two radios are designated for use as the PU pair, while the others are designated for use as the SU pair.

The PU implements several different waveforms in order to provide diversity in spectrum opportunities. 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 to serve as control experiments.

For the baseline non-cognitive system, the SUT has no observe, orient, or decision part of the OODA cognition cycle. For the cognitive system, the 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. The SU CE algorithm is presented further in Section IV-E.

B. 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 as shown in Fig. 1. 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 laptops and the radios are connected via Ethernet to a Cisco Gb switch. The RF channel is approximately white Gaussian

noise. 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 laptop connected via Ethernet. There is no emulation of the radios or channel in this experiment—all communication occurs over the air.

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 undesired 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. Then the next experiment is loaded and the process begins again.

Each experiment is programmed to be 60 seconds in length. Pilot experiments showed that 60 seconds was sufficient to minimize variance of the results. The PU and SU are limited to transmissions 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.

Spectrum analyzers are used to validate the behavior and performance of the WARP radios and to verify that there were no extraneous signal sources during the experiments.

C. Performance Metrics

The SUT offers two services: communication and non-interference. Communication applies to the SU, while non-interference applies to the impact of the SU on the PU. The performance metrics derives from these two services.

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 relevant performance metric is the throughput and BER of the SU.

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. The relevant performance metric is the throughput and BER of the PU.

The throughput of the SU and PU are measured using (1) and in units of bits per second.

$$Throughput = \frac{NAR \cdot BCD \cdot WBS}{WARP\ Clock\ Frequency} \tag{1}$$

where NAR is the number of attempts to receive, BCD is the bits correctly demodulated, and WBS is the WARP buffer size.

BER, as used in this research, is calculated by simply dividing the number of bits in error by the number of bits transmitted. 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 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.

D. Workload

The workload for the SUT is PU activity, otherwise known as the environment. The PU activity creates the radio environment that enables the SUT to react and adapt to in accordance with its CE. The PU transmitter is used to create a total of five environments. In each environment, the PU creates spectrum holes that the SU may utilize. The 10 MHz of spectrum bandwidth is broken up into 64 frequency bins that the PU may transmit on.

The PU waveform is either a wideband, narrowband, or non-contiguous signal to create a range of RF environments. These specific waveforms are created using FSK, OFDM, or NC-OFDM. The overall waveform type stays constant within an experiment, though the frequencies it transmits on may change. 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 $16 \cdot \frac{F_s}{256} = 2.5$ MHz, where F_s is the sampling frequency of 40 MHz.

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 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.

Five distinct environments are created for this experiment: a constant FSK tone with 8 bins bandwidth (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 (NC-OFDM) transmission with 48 bins BW that changes waveform every 5 transmission cycles, and a randomized NC-OFDM waveform with 16 bins of bandwidth. These environments are depicted graphically in Fig. 2. The frequency bins utilized for transmission are randomly generated for each experiment.

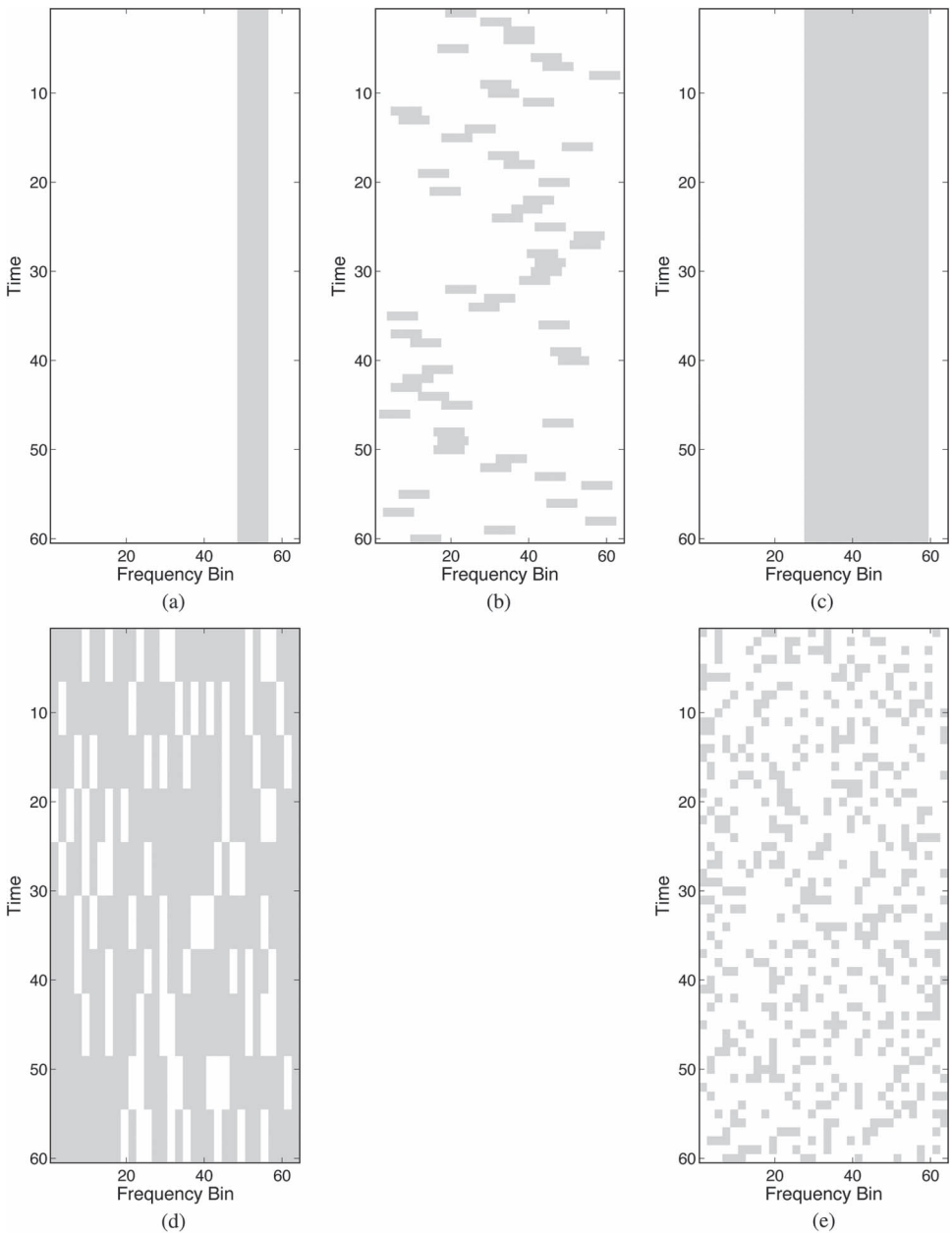

Fig. 2. PU Created Workload Environments consisting of: (a) FSK-1, (b) FSK-8, (c) OFDM-32 Contiguous, (d) OFDM-48 Non-contiguous, and (e) OFDM-16 Randomized. Note the dark areas indicate frequencies that the PU transmits on. (a) Environment 1. (b) Environment 2. (c) Environment 3. (d) Environment 4. (e) Environment 5.

E. System Under Test

Initial pilot experiments reduced the number of key system parameters down to the desired waveform and the CE algorithm. Additionally, factor levels were reduced due to experimental constraints.

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 NC-OFDM. 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 CE is on, the CE 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.

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, 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

TABLE II
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

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. NC-OFDM has lower performance than contiguous OFDM.

The CE allows a radio node to optimize communications performance and non-interference. The CE 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 order of preference for waveforms is OFDM, NC-OFDM, and then FSK for when there are less than 8 bins available. This CE algorithm is intended to represent the baseline functionality of a CE and not to be a fully functional, state of the art CE.

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 is 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.

The SUT uses a simple energy detection algorithm to determine if a channel is open or not. At the beginning of the experiment, the channel noise floor is characterized via a calibration sequence. Then, a threshold is determined by having the PU transmit a known waveform. The threshold is fixed when the SU received REM matches the PU transmitted REM. This threshold is used during the experiment to determine if a channel is occupied or not as sensed by the SU pair.

The workload and SUT factors for this experiment are summarized in Table II.

F. Experimental Design

A full factorial experimental design is used (40 trials). Additionally, baseline performance data is collected for each factor, which adds another 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.

There are two expected results. First, the results should show that for all PU workload environments, the SU configuration with the CE on improves both PU and SU throughput when compared to when the CE is off. Secondly, 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.

V. RESULTS AND ANALYSIS

The experiments were collected in the AFIT Cognitive Radio Laboratory (ACRO) lab located at the Air Force Institute of Technology (AFIT). 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. The performance and behavior of the WARP radios was validated using spectrum analyzers.

The throughput results of each PU environment are shown in Fig. 3. The charts are interpreted as follows. The top subplot shows the SU throughput for each CE configuration, while the bottom subplot shows the PU throughput. Error bars are shown for each bar in the plot. Experiment limit lines are shown for both the SU and PU. 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 results show that 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.

Next, an analysis of the throughput data is performed in order to evaluate the hypothesis and expected results. The analysis consists of two parts. The first is to evaluate if the effect of the SU CE on throughput to be statistically significant by utilizing Student's t-test. The SU and PU performance should be greater for each SU CE on condition vs SU CE off condition. Also, the SU and PU performance should be seen to increase whenever the SU CE is on for greater levels of CE flexibility. The t-test significance level used is $\alpha = 0.05$. The second part of the analysis is to examine BER data to see if it exhibits the same trends as throughput. This is performed for both SU and PU performance.

Experimental data is analyzed for inconsistencies in the data that would suggest errors in the experimental control or BER/throughput data invalidity. Besides BER and throughput, additional experimental data included the elapsed experiment

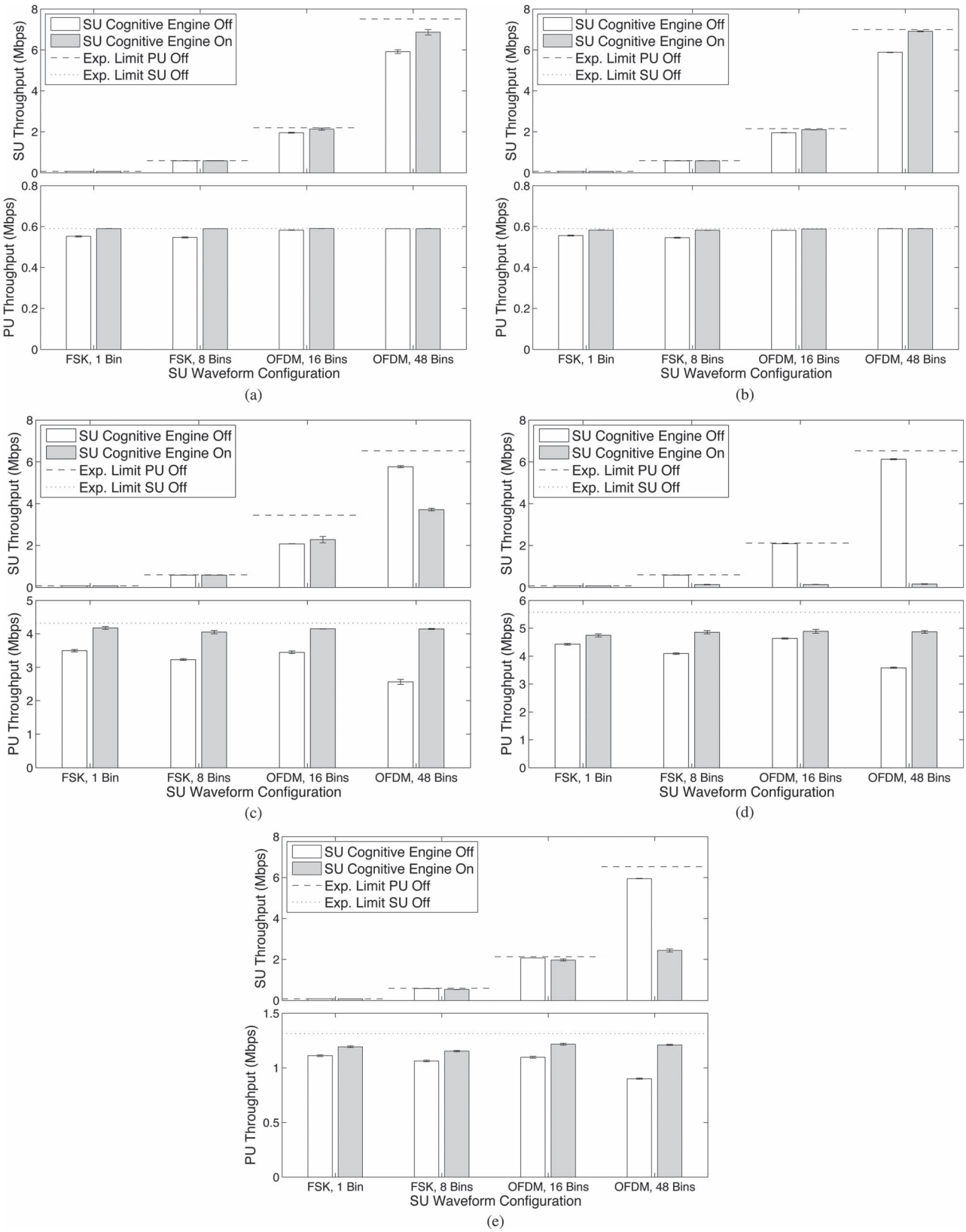

Fig. 3. PU and SU throughput across the five tested environments. The SU throughput is primarily affected by the desired waveform and CE on/off state. The PU throughput magnitude is primarily affected by the choice of the PU environment. (a) Environment 1. (b) Environment 2. (c) Environment 3. (d) Environment 4. (e) Environment 5.

time, total number of transmits and receives, mean received signal strengths, and SU calibrated energy-detection thresholds. None of the data suggested abnormalities that would affect the results suggested below. All data residuals are normal (or as expected) with little skew and no statistically significant outliers.

A. Effect of SU CE on SU Throughput and BER

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 primarily for one reason. 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. 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, $\alpha = 0.05$). If the difference is neither greater nor less, then the configurations are considered to have equal performance.

In this case, 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 is shown graphically in Fig. 4 for the SU. Relative to the SU waveform with the CE off, the BER decreases at a statistically significant level ($\alpha = 0.05$) 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, Environment 4/FSK-8, and Environment 5/OFDM-16. The latter case is not consequential as the OFDM waveforms do not necessarily perform consistently in terms of BER. 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.

B. Effect of SU CE on PU Throughput and BER

The PU throughput and BER data is analyzed in the same manner as the SU data. The CE (i.e. SU) 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 SU CE on as compared to the SU CE off.

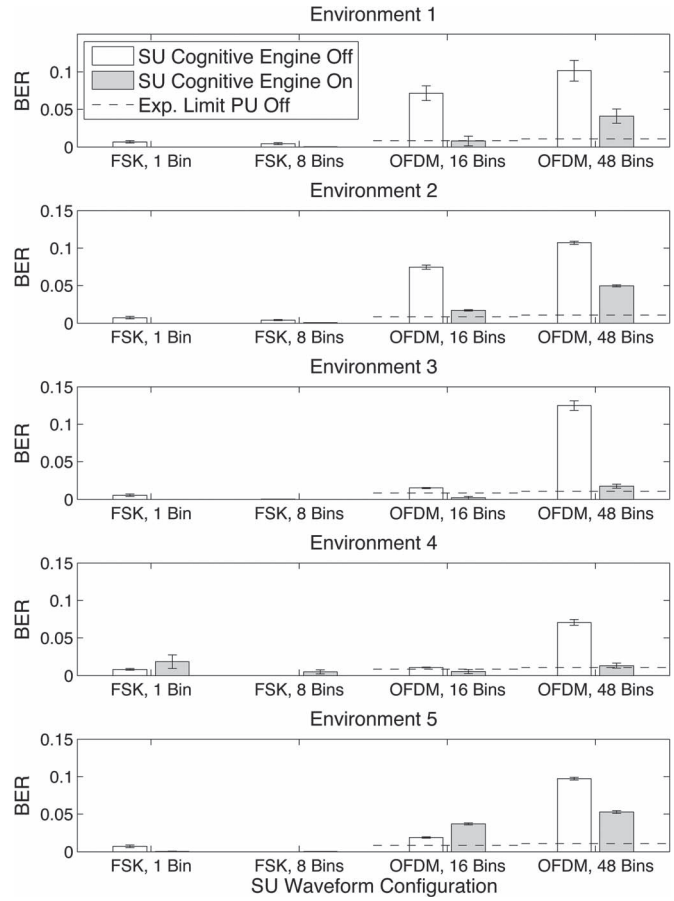

Fig. 4. Secondary user BER (lines and bars not visible are near or equal to 0).

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. 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 data.

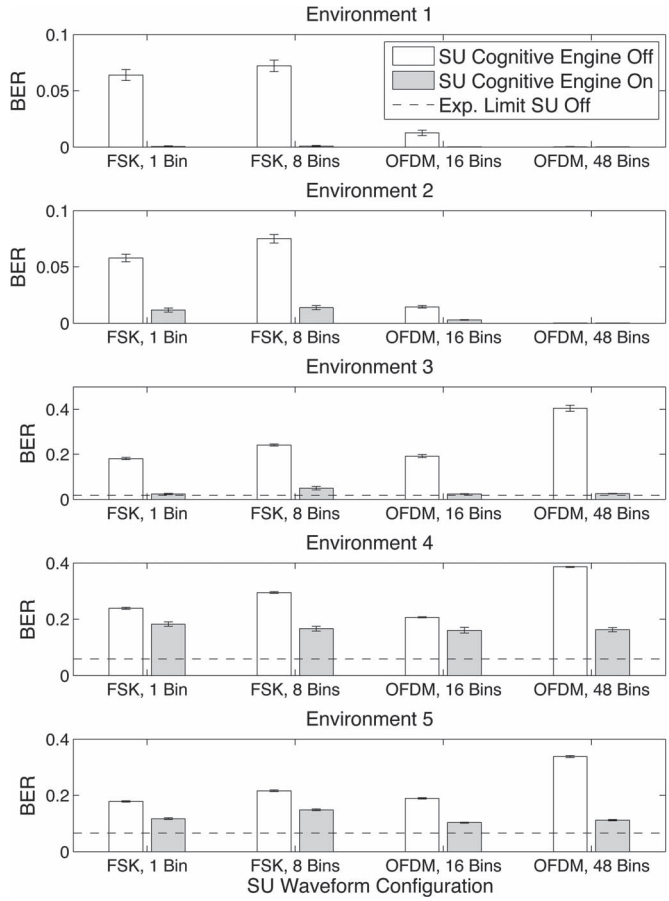

Fig. 5. Primary user BER (lines and bars not visible are near or equal to 0).

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 of the PU is shown graphically in Fig. 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 show that the PU BER decreases.

C. Scoring

CRATM proposes to use relative scoring to evaluate CRSs. A preliminary means of scoring was applied against the experimental results and is shown in Table III. 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 is assigned a score of 10, as measured by throughput. The lowest performing SU configuration is assigned a 1. All other data points are assigned a value between 1 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 combined using a geometric mean. The scores in this analysis are unweighted; however, there is no reason scores cannot be weighted dependent upon specific user desires in future analysis.

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.

There is an upward trend in scores for greater levels of cognition when the CE is on. Furthermore, the CE on condition scores higher than the CE off condition except for one instance in Environment 4. As discussed previously, in this environment the SU with the CE on throttles its performance so as to not cause interference to the PU, meaning that the SU throughput drops by a large amount relative to the CE off condition.

VI. Conclusion

A. Summary

The proposed test methodology, or CRATM, was developed to serve as an end-user, end-to-end, test methodology that tests integrated CR systems. CRATM is also developed to be architecture agnostic and to identify emergent behavior. CRATM uses benchmarks to enable 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 PU. This is known as behavior-based testing.

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 overall 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 flexibility. 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.

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

TABLE III
PERFORMANCE SCORING, SU WEIGHT = 50% PU WEIGHT = 50%

	SU Waveform Config							
Environment	FSK 1 CE Off	FSK 1 CE On	FSK 8 CE Off	FSK 8 CE On	OFDM 16 CE Off	OFDM 16 CE On	OFDM 48 CE Off	OFDM 48 CE On
1	1.4	3.1	1.3	4.1	5.4	6.1	9.3	10.0
2	1.8	2.9	1.3	3.8	5.4	6.0	9.3	10.0
3	2.5	3.2	2.9	4.1	5.0	6.7	3.2	8.2
4	2.6	3.0	2.8	3.3	5.7	3.3	3.2	3.3
5	2.7	3.1	3.2	3.8	5.2	6.3	3.2	6.8

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. Future work should repeat this experiment but with a broader range of CR architectures and features, particular with prototypes that are emerging into the field. Practical benchmarks and benchmark classes should be formalized and tested in accordance with the CRATM framework. The end goal of developing CRATM is that when a new CRS is ready to be tested, it is plugged into the testing framework and produces measurable results that are directly comparable to similar CRS designs for evaluation with minimal system modifications to the CRS under test.

The CRATM framework is geared towards measuring CR system performance as a whole. This is a good way to measure CR effectiveness when, for example, evaluating competing CR systems as a consumer making price/performance trade-offs. Engineers developing new CR systems may desire more fine-grained profiling tools to optimize various metrics and evaluate their impact on system components, such as the sensing subsystem, with the goal of tuning such components in order to improve overall CR effectiveness. The development of such profiling tools and corresponding metrics is beyond the scope of this article. We believe that it is another fruitful area for future work.

ACKNOWLEDGMENT

The views expressed in this document are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

REFERENCES

[1] P. Steenkiste, D. Sicker, G. Minden, and R. Dipankar, "Future directions in cognitive radio network research," National Science Foundation, NSF Workshop Rep., Mar. 9/10, 2009, p. 3.

[2] J. Thompson, E. Hennessey, K. Hopkinson, and M. Silvius, "Evaluation of fast frequency hopping model using evaluation-based test framework for cognitive radios," in *Proc. Wireless Innov. Forum Conf. SDR-WInnComm Technol.*, Schaumburg, IL, USA, 2014, pp. 211–216.

[3] M. Masonta, M. Mzyece, and N. Ntlatlapa, "Spectrum decision in cognitive radio networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1088–1107, 3rd Quart. 2013.

[4] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, "Cognitive networks," in *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*. Dordrecht, The Netherlands: Springer-Verlag, 2007.

[5] International Telecommunication Union, Report ITU-R SM.2152: Definitions of Software Defined Radio (SDR) and Cognitive Radio System (CRS), ser. Spectrum Management Series, 2009.

[6] Y. Zhao, S. Mao, J. O. Neel, and J. H. Reed, "Performance evaluation of cognitive radios: Metrics, utility functions, and methodology," *Proc. IEEE*, vol. 97, no. 4, pp. 642–659, Apr. 2009.

[7] J. Boyd, "A discourse on winning and losing," A Collection of Unpublished Briefings and Essays, Air University Library, Montgomery, AL, USA, Document No. M-U 43947, Aug. 1987.

[8] B. Wang and K. J. R. Liu, "Advances in cognitive radio networks: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 5–23, Feb. 2011.

[9] C. Stevenson *et al.*, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 130–138, Jan. 2009.

[10] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, "IEEE 802.22: The first worldwide wireless standard based on cognitive radios," in *Proc. 1st IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, 2005, pp. 328–337.

[11] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[12] S. M. Mishra *et al.*, "A real time cognitive radio testbed for physical and link layer experiments," in *Proc. 1st IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, 2005, pp. 562–567.

[13] R. K. Sharma *et al.*, "Over-the-air testing of cognitive radio nodes in a virtual electromagnetic environment," *Int. J. Antennas Propag.*, vol. 2013, 2013, Art. ID. 945283.

[14] J. Zhou *et al.*, "WHYNET: A hybrid testbed for large-scale, heterogeneous and adaptive wireless networks," in *Proc. 1st Int. Workshop WiNTECH*, 2006, pp. 111–112.

[15] K. Borries, X. Wang, G. Judd, P. Steenkiste, and D. Stancil, "Experience with a wireless network testbed based on signal propagation emulation," in *Proc. EW Conf. Rec. IEEE IAS Annu. Meeting*, 2010, pp. 833–840.

[16] Echo Ridge LLC DYSE: DYnamic Spectrum Environment Emulator Product Overview. Accessed 25 Oct. 2013. [Online]. Available: http://www.echoridgenet.com/products/dyse

[17] Y. Zhao and L. Morales-Tirado, "Cognitive radio: Forging ahead from concept, testbed to large-scale deployment," *J. Commun.*, vol. 7, no. 7, pp. 514–523, 2012.

[18] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[19] P. J. Kolodzy, "Interference temperature: A metric for dynamic spectrum utilization," *Int. J. Netw. Manage.*, vol. 16, no. 2, pp. 103–113, Mar. 2006.

[20] T. C. Clancy, "Formalizing the interference temperature model," *Wireless Commun. Mobile Comput.*, vol. 7, no. 9, pp. 1077–1086, Nov. 2007.

[21] T. C. Clancy and W. Arbaugh, "Measuring interference temperature," in *Proc. Virginia Tech Symp. Wireless Pers. Commun.*, 2006, pp. 1–10.

[22] M. McHenry, E. Livsics, T. Nguyen, and N. Majumdar, "XG dynamic spectrum access field test results," *IEEE Commun. Mag.*, vol. 45, no. 6, pp. 51–57, Jun. 2007.

[23] Y. Zhao *et al.*, "Applying radio environment maps to cognitive wireless regional area networks," in *Proc. 2nd IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, 2007, pp. 115–118.

[24] C. Dietrich, E. Wolfe, and G. Vanhoy, "Cognitive radio testing using psychometric approaches: Applicability and proof of concept study," *Analog Integr. Circ. Signal Process.*, vol. 73, no. 2, pp. 627–636, Nov. 2012.

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

Jared J. Thompson (S'13) received the B.S. degree in astronautical engineering from the U.S. Air Force Academy, Colorado Springs, CO, USA, in 2008, the M.E. degree in aerospace engineering from the University of Colorado, Boulder, CO, in 2012, and the M.S. degree in electrical engineering from the Air Force Institute of Technology (AFIT), Wright-Patterson AFB, OH, USA, in 2014. He is currently with AFIT. He is a Graduate Student Member of the IEEE and a member of the Eta Kappa Nu honor society.

Kenneth M. Hopkinson (S'98–M'04–SM'10) received the B.S. degree from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1997 and the M.S. and Ph.D. degrees from Cornell University, Ithaca, NY, in 2002 and 2004, respectively, all in computer science. He is currently a Professor of computer science at AFIT, Wright-Patterson AFB, OH, USA. His research interests are in the areas of simulation, networking, and distributed systems.

Mark D. Silvius (M'12) received the B.S. degree from Cornell University, Ithaca, NY, USA, in 2000, the M.S. degree from Syracuse University, Syracuse, NY, in 2003, and the Ph.D. degree from Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, in 2009, all in electrical engineering. He is currently an Adjunct Assistant Professor of electrical engineering at AFIT, Wright-Patterson AFB, OH, USA. His research interests are in the areas of wireless communications, digital design, field-programmable gate arrays, and cognitive radio.