Qing Zhao and Brian M. Sadler

A Survey of Dynamic Spectrum Access

Signal processing, networking, and regulatory policy

There is a common belief that we are running out of usable radio frequencies. The overly crowded U.S. frequency allocation chart and the multibillion-dollar price for a 20 MHz frequency band at the European 3G spectrum auction have certainly strengthened this belief. Are we truly approaching the capacity of the radio spectrum? Actual spectrum usage measurements obtained by the FCC's Spectrum Policy Task Force [1] tell a different story: At any given time and location, much of the prized spectrum lies idle. This paradox indicates that spectrum shortage results from the spectrum management policy rather than the physical scarcity of usable frequencies. Analogous to idle slots in a static time division multiple access (TDMA) system with bursty traffic, idle frequency bands are inevitable under the current static spectrum allotment policy that grants exclusive use to licensees.

The underutilization of spectrum has stimulated a flurry of exciting activities in engineering, economics, and regulation communities in searching for better spectrum management policies and techniques. The diversity of the envisioned spectrum reform ideas is manifested in the number of technical terms coined so far: dynamic spectrum access versus dynamic spectrum allocation, spectrum property rights versus spectrum commons, opportunistic spectrum access versus spectrum pooling, spectrum underlay versus spectrum overlay. Compounding the confusion is the use of the broad term *cognitive radio* as a synonym for dynamic spectrum access. As an initial attempt at unifying the terminology, we provide the following taxonomy.

DYNAMIC SPECTRUM ACCESS

Standing for the opposite of the current static spectrum management policy, the term *dynamic spectrum access* has broad connotations that encompass various approaches to spectrum reform. The diverse ideas presented at the first IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN) suggest the extent of this term. As illustrated in Figure 1, dynamic spectrum access strategies can be broadly categorized under three models.

DYNAMIC EXCLUSIVE USE MODEL

This model maintains the basic structure of the current spectrum regulation policy: Spectrum bands are licensed to services for exclusive use. The main idea is to introduce flexibility to improve spectrum efficiency. Two approaches have been proposed under this model: Spectrum property rights [2], [3] and dynamic spectrum allocation [4]. The former approach allows licensees to sell and trade spectrum and to freely choose technology. Economy and market will thus play a more important role in driving toward the most profitable use of this limited resource. Note that even though licensees have the right to lease or share the spectrum for profit, such sharing is not mandated by the regulation policy.

The second approach, dynamic spectrum allocation, was brought forth by the European DRiVE project [4]. It aims to improve spectrum efficiency through dynamic spectrum assignment by exploiting the spatial and temporal traffic statistics of different services. In other words, in a given region and at a given time, spectrum is allocated to services for exclusive use. This allocation, however, varies at a much faster scale than the current policy.

Based on an exclusive-use model, these approaches cannot eliminate white space in spectrum resulting from the bursty nature of wireless traffic.

OPEN SHARING MODEL

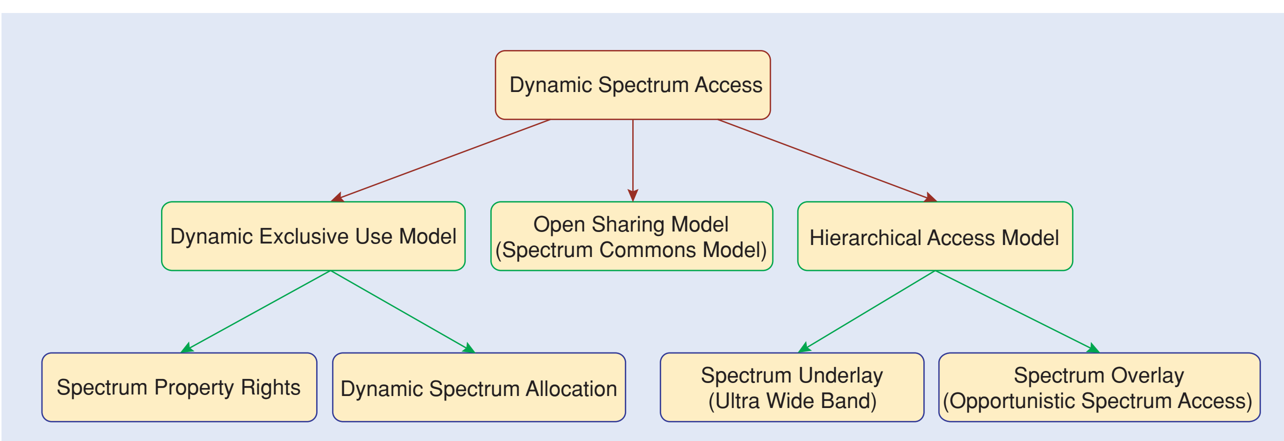
Also referred to as *spectrum commons* [5], [6], this model employs open sharing among peer users as the basis for managing a spectral region. Advocates of this model draw support from the phenomenal success of wireless services operating in the unlicensed industrial, scientific, and medical (ISM) radio band (e.g., WiFi). Centralized [7], [8] and distributed [9]–[11] spectrum sharing strategies have been initially investigated to address technological challenges under this spectrum management model.

HIERARCHICAL ACCESS MODEL

This model adopts a hierarchical access structure with primary and secondary users. The basic idea is to open licensed spectrum to secondary users while limiting the interference perceived by primary users (licensees). Two approaches to spectrum sharing between primary and secondary users have been considered: Spectrum underlay and spectrum overlay.

The underlay approach imposes severe constraints on the transmission power of secondary users so that they operate below the noise floor of primary users. By spreading transmitted signals over a wide frequency band (UWB), secondary users can potentially achieve short-range high data rate with extremely low transmission power. Based on a worst-case assumption that primary users transmit all the time, this approach does not rely on detection and exploitation of spectrum white space.

Spectrum overlay was first envisioned by Mitola [12] under the term *spectrum pooling* and then investigated by the DARPA Next Generation (XG) program under the term *opportunistic spectrum access*. Differing from spectrum underlay, this approach does not necessarily impose severe restrictions on the transmission power of secondary users, but rather on when and where they may transmit. It directly targets at spatial and temporal spectrum white space by allowing secondary users to identify and exploit local and instantaneous spectrum availability in a nonintrusive manner.



[FIG1] A taxonomy of dynamic spectrum access.

Compared to the dynamic exclusive use and open sharing models, this hierarchical model is perhaps the most compatible with the current spectrum management policies and legacy wireless systems. Furthermore, the underlay and overlay approaches can be employed simultaneously to further improve spectrum efficiency.

> DYNAMIC SPECTRUM ACCESS STRATEGIES CAN BE BROADLY CATEGORIZED UNDER THREE MODELS: DYNAMIC EXCLUSIVE USE MODEL, OPEN SHARING MODEL, AND HIERARCHICAL ACCESS MODEL.

COGNITIVE RADIO

The terms *software-defined radio* and *cognitive radio* were promoted by Mitola in 1991 and 1998, respectively. Software-defined radio, sometimes shortened to software radio, is generally a multiband radio that supports multiple air interfaces and protocols and is reconfigurable through software run on DSP or general-purpose microprocessors [13]. Cognitive radio, built on a software radio platform, is a context-aware intelligent radio potentially capable of autonomous reconfiguration by learning from and adapting to the communication environment [14]. While dynamic spectrum access is certainly an important application of cognitive radio, cognitive radio represents a much broader paradigm where many aspects of communication systems can be improved via cognition.

OPPORTUNISTIC SPECTRUM ACCESS: BASIC COMPONENTS

In this article, we focus on the overlay approach under the hierarchical access model (see Figure 1). The term *Opportunistic Spectrum Access (OSA)* will be adopted throughout.

Basic components of OSA include spectrum opportunity identification, spectrum opportunity exploitation, and regulatory policy. The opportunity identification module is responsible for accurately identifying and intelligently tracking idle frequency bands that are dynamic in both time and space. The opportunity exploitation module takes input from the opportunity identification module and decides whether and how a transmission should take place. The regulatory policy defines the basic etiquette for secondary users to ensure compatibility with legacy systems.

The overall design objective of OSA is to provide sufficient benefit to secondary users while protecting spectrum licensees from interference. The tension between the secondary users' desire for performance and the primary users' need for protection dictates the interaction across opportunity identification, opportunity exploitation, and regulatory policy. The optimal design of OSA thus calls for a cross-layer approach that integrates signal processing and networking with regulatory policy making.

In this article, we provide an overview of challenges and recent developments in both technological and regulatory aspects of OSA. The three basic components of OSA will be discussed in the following sections.

AN EXAMPLE OF OSA NETWORKS

To illustrate the basic technical issues in OSA, we often resort to the following example of OSA networks. The design challenges, tradeoffs, and many existing results presented in this article, however, apply to general OSA networks.

We consider a spectrum consisting of N channels. Here we use the term *channel* broadly. A channel can be a frequency band with certain bandwidth, a collection of spreading codes in a code division multiple access (CDMA) network,or a set of tones in an orthogonal frequency division multiplexing (OFDM) system. We assume that cross-channel interference is negligible. Thus, a secondary user transmitting over an available channel does not interfere with primary users using other channels. This assumption imposes constraints on the modulation of secondary users, as will be discussed later.

These N channels are allocated to a network of primary users. For ease of presentation, we assume that the primary system uses a synchronous slot structure, although the basic ideas apply more generally. The traffic statistics of the primary system are such that the occupancy of these N channels follows a Markov process with 2^N states, where the state is defined as the availability (idle or busy) of each channel. Overlaid with this primary network is an ad hoc secondary network where users seek spectrum opportunities in these N channels independently. The transition probabilities of the underlying Markov process are known or have been learned by secondary users. In each slot, a secondary user chooses a channel to sense and decides whether to access based on imperfect sensing outcomes. Accessing an idle channel leads to bit delivery, and accessing a busy channel results in a collision with primary users.

SPECTRUM OPPORTUNITY IDENTIFICATION

Spectrum opportunity identification is crucial to OSA in order to achieve nonintrusive communication. In this section, we identify basic functions of the opportunity identification module.

SPECTRUM OPPORTUNITY AND INTERFERENCE CONSTRAINT: DEFINITIONS AND IMPLICATIONS

SPECTRUM OPPORTUNITY

Before discussing spectrum opportunity identification, a rigorous definition of spectrum opportunity is necessary. Intuitively, a channel can be considered as an opportunity if it is not currently used by primary users. In a network with geographically distributed primary transmitters and receivers, however, the concept of spectrum opportunity is more involved than it at first may appear [15].

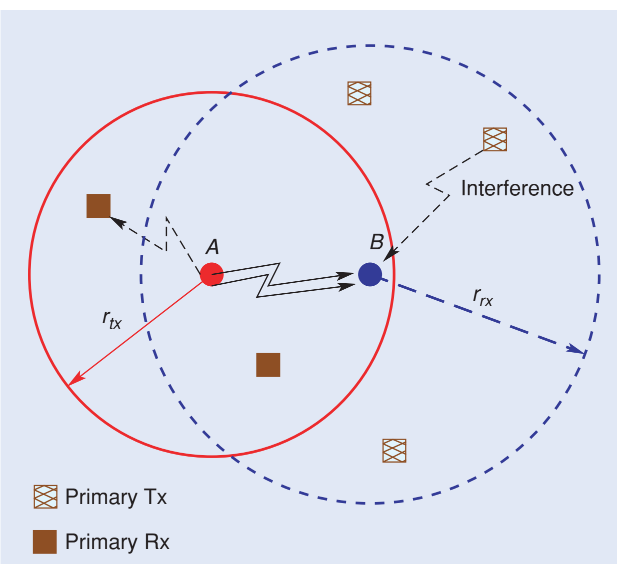
With the help of Figure 2, we identify conditions for a channel to be considered as an opportunity. Consider a pair of secondary users where A is the transmitter and B its intended receiver. A channel is an opportunity to A and B if they can communicate successfully over this channel while limiting the interference to primary users below a prescribed level determined by the regulatory policy. This means that receiver B will not be affected by primary transmitters, and transmitter A will not interfere with primary receivers.

To illustrate the above conditions, we consider monotonic and uniform signal attenuation and omnidirectional antennas. In this case, a channel is an opportunity to A and B if no primary users within a distance of r_{tx} from A are receiving and no primary users within a distance of r_{rx} from B are transmitting over this channel (see Figure 2). Clearly, r_{tx} is determined by the secondary users' transmission power and the maximum allowable interference to primary users, while r_{rx} is determined by the primary users' transmission power and the secondary users' interference tolerance. They are generally different.

THE TENSION BETWEEN THE SECONDARY USERS' DESIRE FOR PERFORMANCE AND THE PRIMARY USERS' NEED FOR PROTECTION DICTATES THE INTERACTION ACROSS OPPORTUNITY IDENTIFICATION, OPPORTUNITY EXPLOITATION, AND REGULATORY POLICY.

We make the following remarks regarding the above definition of spectrum opportunity.

- Spectrum opportunity is a local concept defined with respect to a particular pair of secondary users. It depends on the location of not only the secondary transmitter but also the secondary receiver. For multicast and broadcast, spectrum opportunity is open for interpretation, and results in networking tradeoffs.
- Spectrum opportunity is determined by the communication activities of primary users rather than that of secondary users. Failed communications caused by collisions among secondary users do not disqualify a channel from being an opportunity.



[FIG2] Illustration of spectrum opportunity (secondary user *A* wishes to transmit to secondary user *B*, where *A* should watch for nearby primary receivers and *B* nearby primary transmitters.

INTERFERENCE CONSTRAINT

How to impose interference constraints is a complex regulatory issue. Restrictive constraints may marginalize the potential gain of OSA, while loose constraints may affect the compatibility with legacy systems.

Generally speaking, an interference constraint should implicitly or explicitly specify at least two parameters: The maximum interference power level η perceived by an active primary receiver and the maximum probability ζ that the interference level at an active primary receiver may exceed η [15]. The first parameter, η, can be considered as specifying the noise floor of primary users; interference below η does not affect primary users, while interference above η results in a collision. It is thus inherent to the definition of spectrum opportunity (through r_{tx} in Figure 2) and determines the transmission power of secondary users as discussed later [see (1)].

The second parameter, ζ, specifies the maximum allowable collision probability. Given that errors in spectrum opportunity detection are inevitable, a positive value of ζ is necessary for secondary users to ever be able to exploit an opportunity. As discussed later, ζ determines a secondary transmitter's access decision based on imperfect spectrum opportunity detection. A cautionary aspect is that different definitions of collision probability offer different levels of protection to primary users [15]. For example, the collision constraint can be imposed on the joint probability that both primary and secondary users access the same channel or on the conditional probability that a secondary user transmits given that the channel is occupied by primary users. The protection offered by the former constraint varies with the traffic load of primary users; primary users with a light traffic load may not be as well protected as those with a heavy traffic load. Another issue is whether the constraint should be imposed in each slot over each channel or on the collision probability averaged over channels and a long period of time. The former offers a specific level of protection to primary users no matter when and over which channel they transmit, while the protection given by the latter can be unpredictable when primary users have bursty arrivals of short messages.

While an interference constraint specified by $\{\eta, \zeta\}$ should be imposed on the aggregated transmission activities of all secondary users, each secondary user needs to know the node-level constraint in order to choose transmission power and make access decisions. The translation from a network-level interference constraint to a node-level one depends on the geolocation and traffic of secondary users as well as the signal attenuation model in the communication environment with shadowing and fading.

SPECTRUM OPPORTUNITY DETECTION

SIGNAL PROCESSING AND NETWORKING TECHNIQUES FOR OPPORTUNITY DETECTION

From the definition of spectrum opportunity illustrated in Figure 2, it is clear that, in a general network setting, spectrum opportunity detection needs to be performed jointly by the secondary transmitter and receiver. It thus has both signal processing and networking aspects.

WHICH DESIGN CRITERION SHOULD BE ADOPTED AND WHAT OPERATING CHARACTERISTICS OF THE SPECTRUM DETECTOR ARE DESIRABLE SHOULD BE ADDRESSED BY CONSIDERING MAC LAYER PERFORMANCE: THE THROUGHPUT OF SECONDARY USERS AND THE PROBABILITY OF COLLIDING WITH PRIMARY USERS.

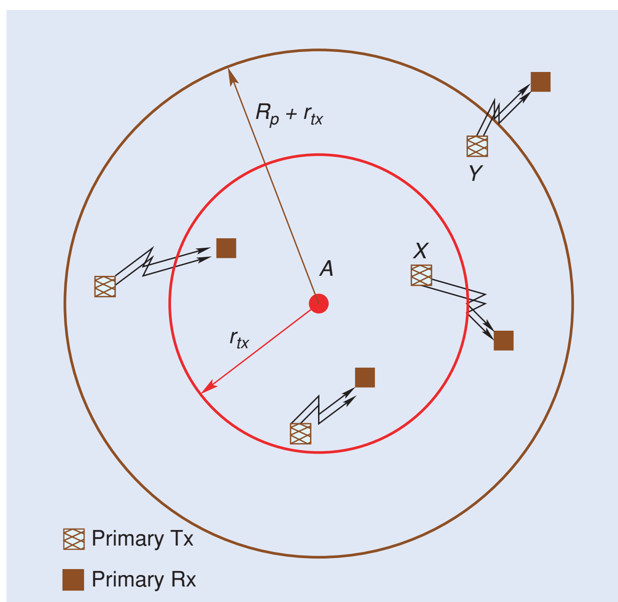
Consider the OSA network example given earlier. At the beginning of each slot, a pair of communicating secondary users need to determine whether a chosen channel is an opportunity in this slot. Ignore for now the contention among secondary users. One approach to opportunity detection is as follows [16]: The transmitter first detects the receiving activities of primary users in its neighborhood (see Figure 2). If the channel is available (no primary receivers nearby), it transmits a short request-to-send (RTS) message to the receiver. The receiver, upon successfully receiving the RTS, knows that the channel is also available at the receiver side and replies with a clear-to-send (CTS) message. A successful exchange of RTS-CTS completes opportunity detection and is followed by data transmission. As detailed in [16], when RTS and CTS are transmitted using carrier sensing, this RTS-CTS exchange has dual functions. Besides facilitating opportunity detection, it also addresses contention among secondary users and mitigates the hidden and exposed terminal problem as in a conventional communication network.

What remains to be solved is the detection of the receiving activities of primary users by the secondary transmitter. Without assuming cooperation from primary users, primary receivers are much harder to detect than primary transmitters. For the application of secondary wireless services operating in the TV bands, Wild and Ramchandran [17] proposed to exploit the local oscillator leakage power emitted by the RF front end of TV receivers to detect the presence of primary receivers. The difficulty of this approach lies in its short detection range and long detection time to achieve accuracy. It is proposed in [17] that low-cost sensors be deployed close to primary receivers for spectrum opportunity detection.

Another approach is to transform the problem of detecting primary receivers to detecting primary transmitters. Let R_p denote the transmission range of primary users, i.e., primary receivers are within R_p distance to their transmitters. A secondary transmitter can thus determine that a channel is available if no primary transmitters are detected within a distance of $R_p + r_{tx}$ as illustrated in Figure 3. This approach, however, is conservative, potentially leading to overlooked opportunities. As shown in Figure 3, the transmission activities of primary nodes X and Y may prevent A from accessing an opportunity even though the intended receivers of X and Y are outside the interfering range r_{tx} of A. Note that by adjusting the detection range (with $R_p + r_{tx}$ being the most conservative), we reach tradeoffs between the throughput of secondary users and interference to primary users.

The above approach reduces spectrum opportunity detection to a classic signal processing problem. As discussed in [18], based on the secondary user's knowledge of the signal characteristics of primary users, three traditional signal detection techniques can be employed: Matched filter, energy detector (radiometer), and cyclostationary feature detector. A matched filter performs coherent detection. It requires only $\mathcal{O}(1/SNR)$ samples to achieve a given detection power but relies on synchronization and a priori knowledge of primary users' signaling (the detector might also use known flags or training symbols in the primary users' signal). On the other hand, the noncoherent energy detector requires only basic information of primary users' signal characteristics but suffers from long detection time: $\mathcal{O}(1/SNR^2)$ samples are needed for a given detection power. A cyclostationary feature detector can improve the performance over an energy detector by exploiting an inherent periodicity in the primary users' signal. Details of this detector can be found in [19].

While these classic signal detection techniques are well known, detecting primary transmitters in a dynamic wireless



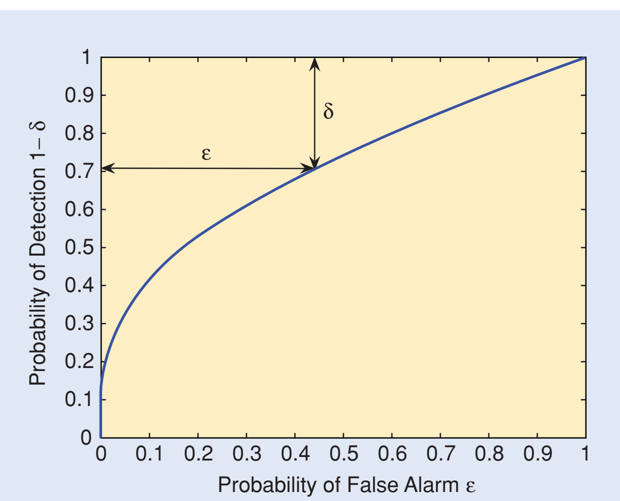
[FIG3] Spectrum opportunity detection: A conservative approach that transforms the problem of detecting primary receivers to detecting primary transmitters.

environment with noise uncertainty, shadowing, and fading is a challenging problem as articulated in [20]. To improve detection accuracy, cooperative spectrum sensing has been proposed [18], [21], [22]. The basic idea is to overcome shadowing and multipath fading by allowing neighboring secondary users to exchange sensing information through a dedicated control channel. The overhead associated with sensing information exchange, the feasibility of a control channel, and the applicability to OSA networks with fast varying spectrum usage remain significant challenges.

DESIGN CRITERIA AND PERFORMANCE CHARACTERISTICS OF SPECTRUM OPPORTUNITY DETECTOR

The spectrum opportunity detector discovers the presence of primary users in a given channel. It can be considered as performing a binary hypotheses test, where the null hypothesis $\mathcal{H}_0$ indicates the absence of primary users (an opportunity), and hypothesis $\mathcal{H}_1$ is the alternative. If the detector mistakes $\mathcal{H}_0$ for $\mathcal{H}_1$, a false alarm occurs, and a spectrum opportunity is overlooked by the detector. On the other hand, when the detector mistakes $\mathcal{H}_1$ for $\mathcal{H}_0$, we have a miss detection, which potentially leads to a collision with primary users. Let ϵ and δ denote the probabilities of false alarm and miss detection, respectively. The performance of the detector is specified by the receiver operating characteristic (ROC) curve, which gives $1-\delta$ (probability of detection or detection power) as a function of ϵ. As illustrated in Figure 4, a smaller false alarm probability ϵ implies a larger miss detection probability δ. As an example, consider the energy detector. Changing the energy detection threshold leads to different operating points $\delta \in (0, 1)$ on the ROC curve.

In general, which design criterion should be adopted and what operating characteristics (false alarm versus miss detection probabilities) of the detector are desirable should be addressed by considering MAC layer performance—the throughput of secondary users and the probability of colliding with primary users [24]. For example, should we use Bayesian or Neyman-Pearson criterion? If the former, how do we choose the risk function? If the latter, how do we set the constraint on the false alarm probability? On the other hand, spectrum access strategies at the MAC layer should take into account the operating characteristics of the opportunity detector as shown later. A joint design of opportunity identification at the physical layer and opportunity exploitation at the MAC layer is thus necessary to achieve the optimal performance [23], [24].



[FIG4] Receiver operating characteristics of spectrum opportunity detector.

SPECTRUM OPPORTUNITY TRACKING

Due to hardware limitations and energy constraints, a secondary user may not be able to sense all N channels in the spectrum simultaneously. In this case, a sensing strategy for intelligent channel selection to track the rapidly varying spectrum opportunities is necessary. The purpose of the sensing strategy is twofold: Catch a spectrum opportunity for immediate access and obtain statistical information on spectrum occupancy so that more rewarding sensing decisions can be made in the future. A tradeoff has to be reached between these two often conflicting objectives.

Consider again the OSA network example. A simple static sensing strategy would choose the channel most likely to be available (weighted by its bandwidth) based on the stationary distribution of the underlying Markov process. In this case, the secondary user simply waits on a particular channel predetermined by the spectrum occupancy statistics and the channel bandwidths. Missing in this approach is that every sensing outcome provides information on the state of the underlying Markov process. Channel selection should be based on the conditional distribution of channel availability that exploits the whole history of sensing outcomes.

The optimal sensing strategy is thus one of sequential decision making that achieves the best tradeoff between gaining immediate access in the current slot and gaining system state information for future use. In [16], [25], and [26], the design of optimal sensing strategies has been formulated and addressed within the framework of partially observable markov decision processes (POMDP). Based on these results, we illustrate the potential gain of optimally using the observation history with a simple numerical example where we have three channels with unit bandwidth. The throughput of the secondary user as a function of time is plotted in Figure 5. We see from this figure that the performance of the optimal approach based on the POMDP framework improves over time, which results from the increasingly accurate information on the system state drawn from accumulated observations. Approximately 40% improvement is achieved over the static approach.

It is therefore apparent that a simple yet sufficiently accurate statistic model of spectrum occupancy is crucial to the efficiency of spectrum opportunity tracking. Spectrum monitoring testbeds [27] and cognitive radio prototypes [28] are being developed by researchers from both academia and industry. They provide empirical data for the statistical modeling of spectrum occupancy. Results in [27] demonstrate the Markovian transition between busy and idle channel states in 802.11 b, and a continuous-time semi-Markov process model is proposed.

SPECTRUM OPPORTUNITY EXPLOITATION

Once spectrum opportunities are detected, secondary users need to decide whether and how to exploit them. Specific issues include whether to transmit given that opportunity detectors will make mistakes, what modulation and transmission power to use, and how to share opportunities among secondary users to achieve a network-level objective.

WHETHER TO ACCESS

A secondary user needs an access strategy to determine whether to transmit over a particular channel based on the detection outcome. If the spectrum detector was perfect, the design of the access strategy would have been straightforward. In the presence of detection errors, the access strategy is complicated by the need to decide how much and when to trust the detector. The tradeoff is between minimizing overlooked spectrum opportunities and avoiding collisions with primary users.

The optimal access strategy should take into account the operating characteristics of the spectrum detector. Intuitively, when the miss detection probability of the detector is large (i.e., a busy channel is often detected as idle), the access policy should be conservative to avoid excessive collisions. On the other hand, when the detector has a high false alarm probability, the access policy should be aggressive to reduce overlooked spectrum opportunities. For any given operating point δ on the ROC curve, exactly how aggressive or how conservative the optimal access policy should be is, however, not a trivial problem.

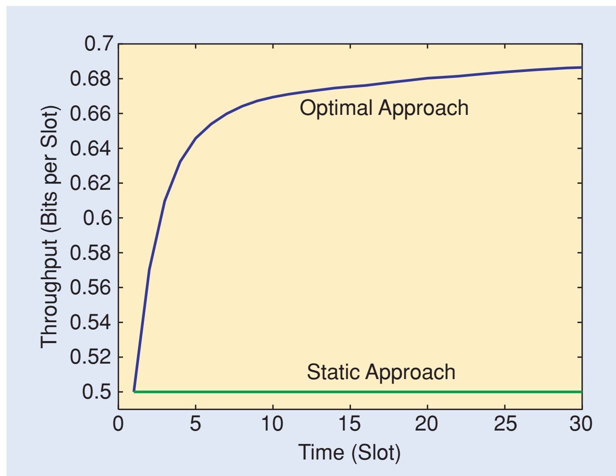
For the OSA network example, a separation principle based on a POMDP framework has been established in [23] and [24] that leads to a closed-form characterization of the optimal access strategy jointly designed with the sensing strategy for any operating point $\delta \in (0, 1)$ of the spectrum detector. As illustrated in Figure 6, the ROC curve of the detector is partitioned into two regions by the maximum allowable collision probability ζ. When the detector operates at $\delta > \zeta$, there is a high chance that a busy channel is detected as an opportunity. The optimal access policy should be conservative. Specifically, when the channel is detected as busy, the secondary user should always refrain from transmission; even when the channel is detected to be available, it should only transmit with probability $(\zeta/\delta) < 1$.

On the other hand, in the region of $\delta < \zeta$, false alarms are likely to happen. The user should adopt an aggressive access strategy: When the channel is detected to be available, always transmit; even when the channel is detected to be busy, one should still transmit with probability $(\zeta - \delta)/(1 - \delta) > 0$.

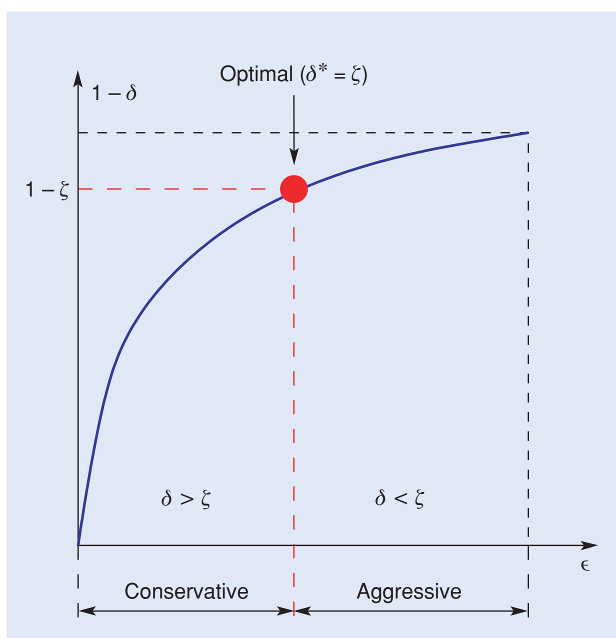
Interestingly, as highlighted in Figure 6, the optimal joint design of opportunity detector and sensing and access strategies requires that the detector be designed under the Neyman-Pearson criterion and operate at the transition point $\delta^* = \zeta$. The corresponding optimal access strategy is to simply trust the detector: access if and only if the channel is detected to be available. In other words, the access strategy does not need to be conservative or aggressive to balance the occurrence of false alarms and miss detection.

HOW TO ACCESS

Modulation and power control in OSA networks also present unique challenges not encountered in the conventional wired or wireless networks. Since secondary users may need to transmit over noncontiguous frequency bands, OFDM is an attractive candidate for modulation in OSA networks [29]–[31]. The reconfigurable subcarrier structure of OFDM allows secondary users to efficiently fill the spectral gaps left by primary users without causing unacceptable interference. The FFT component of OFDM can also be used by the energy detector of secondary users for opportunity detection.



[FIG5] Spectrum opportunity tracking: A sequential decision-making problem where the throughput of secondary users improves over time due to accumulated observations.



[FIG6] The optimal access strategy should take into account the operating characteristics of the spectrum opportunity detector (ϵ: probability of false alarm, δ: probability of miss detection, ζ: maximum allowable collision probability).

There are, however, several constraints in designing an OFDM overlay system. First, the subcarrier spacing and symbol interval need to match with the spectral and temporal duration of spectrum opportunities [30]. Second, cross-channel spectrum leakage caused by signal truncation in the time domain and nonlinearity of the transmitter's power amplifier needs to be controlled to ensure nonintrusive communication. Carefully designed pulse shaping can reduce the smearing effect in the frequency domain induced by time-domain truncation. Subcarriers adjacent to channels occupied by primary users may be nulled or allocated with low power to meet the interference requirement, giving power allocation an interesting twist. Furthermore, the impact of nulled subcarriers on the peak-to-average-power ratio of the transmitted OFDM signal requires careful study.

Transmission power control is another complex issue in OSA networks [15]. To illustrate the basic parameters that affect power control, we ignore shadowing and fading and focus on a single secondary user. Consider first that the secondary transmitter A is able to detect the presence of primary receivers within a distance of d (see Figure 2 with r_{tx} replaced by d). The transmission power P_{tx} of A should ensure that the signal strength at d away from A is below the maximum allowable interference level η:

$$P_{tx} \leq \eta d^{\alpha}, \quad (1)$$

where α is the path attenuation factor. The above equation indicates how the maximum transmission power of a secondary user depends on the detection range d of its spectrum detector, the prescribed maximum interference level η, and the path loss factor α.

When the secondary user can only detect the presence of primary transmitters within a distance of d (see Figure 3 with $R_p + r_{tx}$ replaced by d), we have

$$P_{tx} \leq \eta (d - R_p)^{\alpha},$$

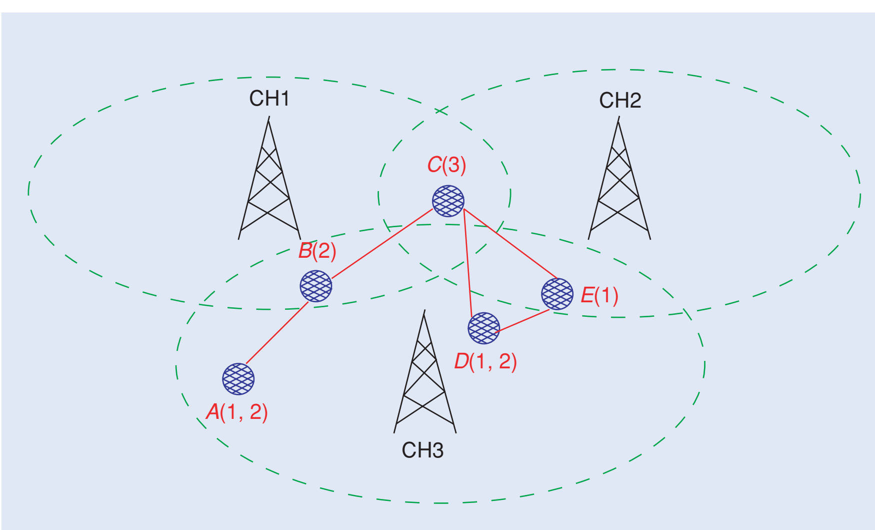
where R_p is the transmission range of primary users. In other words, power control for secondary users should also take into account the transmission power of primary users. When we consider shadowing, fading, and interference aggregation due to simultaneous transmissions from multiple secondary users, a probabilistic model may be necessary to address power control in OSA networks.

OPPORTUNITY SHARING AMONG SECONDARY USERS

So far we have been focusing on individual noncooperative secondary users. In the context of exploiting locally unused TV broadcast bands, spatial spectrum opportunity sharing among secondary users has been investigated (see [32]–[34] and references therein). For this type of application, spectrum opportunities are considered static or slowly varying in time. Real-time opportunity identification is not as critical a component as in applications that exploit temporal spectrum opportunities (as in the example OSA network given previously). It is often assumed that spectrum opportunities at any location over the entire spectrum are known.

We illustrate the problem of spatial spectrum opportunity sharing with the help of Figure 7. Assume that there are three primary users, each occupying one of the three channels. A secondary user within the coverage area of a primary user cannot use the channel occupied by that primary user. For example, channels available to secondary user A are $(1, 2)$. Furthermore, neighboring secondary users (indicated by a line connecting two secondary users in Figure 7) interfere with each other if they access the same channel. The problem is how to allocate available channels to secondary users to optimize certain network utility such as sum capacity under fairness constraints.

It has been shown in [32] and [33] that spatial opportunity allocation is equivalent to graph coloring. Specifically, secondary users form vertices in a graph, and an edge between two vertices indicates two interfering users. Treating each channel as a color, we arrive at a graph coloring problem: Color each vertex using a number of colors from its color list under the constraint that two vertices linked by an edge cannot share the same color. The objective is to obtain a color assignment that maximizes a given utility function.



[FIG7] Spatial spectrum opportunity sharing among secondary users formulated as a graph coloring problem.

Obtaining the optimal coloring is known to be NP-hard. Centralized and distributed suboptimal approaches have been proposed [32], [33]. Game theory provides another approach to spatial opportunity allocation [35]. An interesting connection between the resultant colored graph and the Nash equilibria of the corresponding game is noted in [35].

REGULATORY POLICIES

ASPECTS OF OSA POLICY

Policy is obviously an important piece of OSA, establishing rules of cooperation and joint usage between primary and secondary users. In the United States, the FCC is studying ways to advance secondary markets, such as via interruptible leasing, a logical first step for commercial and user mutual benefit. A supporting policy could be fixed or open to dynamic negotiation and bidding; it could be centralized or decentralized. Basic policy questions such as these are affected by a variety of factors, many noted above, depending on the application and legacy systems. It can be expected that intramilitary systems, as well as intracommercial systems, can benefit greatly from policies allowing spectrum sharing. Should military and commercial systems interact and coexist? What form should such a policy take, perhaps allowing for different modes of operation, such as in times of national emergency? Spectrum regulatory policies vary over countries and regions as well as across spectral sections. How can policies be defined across international boundaries and regions? While it is generally agreed that OSA can potentially bring numerous benefits, there are many technical as well as cost and business issues to address before widespread deployment can occur, and all these issues are intertwined with policy.

IN THE PRESENCE OF DETECTION ERRORS, THE ACCESS STRATEGY NEEDS TO DECIDE HOW MUCH AND WHEN TO TRUST THE SPECTRUM DETECTOR.

Policies must be implemented on radio devices. A logical argument for separation of the radio and the policy software includes the option to add OSA capability to legacy systems and the ability to update or drop in new policies. However, implementation of a separate software section raises security and software verifiability issues. Modification by users could result in policy violations [36]. Furthermore, device testing and verification for policy compliance will be greatly complicated by dynamic policies and the complex interaction of networked devices sensing and reacting to the environment.

A wide range of policies are easily envisioned, spanning nonaggressive to aggressive or restrictive to permissive. An obvious extreme is a do-no-harm policy, e.g., maintain complete orthogonality between systems at all times. Less restrictive policies may allow limited harm. On the other extreme, e.g., in times of national emergency, a secondary system might have complete freedom to operate without restriction in an otherwise occupied band. In addition, while individual policies should be unambiguous, it is easy to envision that multiple conflicting policies may arise.

An early example of a sense-and-respond OSA policy is DFS [37]–[39]. DFS allows unlicensed 802.11 communications devices in the 5 GHz band to coexist with legacy radar systems. The policy specifies the sensor detection threshold as well as timeline for radar sensing, usage, abandoning the channel, and a nonoccupancy time after detection. This policy allows limited but minimal harm to legacy radar systems by accounting for the specific form of sensor for detection and prescribing the timeline for channel use and departure. Another early example has been developed in the DARPA NeXt Generation (XG) program, as a general listen-before-talk strategy, analyzed by Leu et al. [40].

WHERE POLICY MEETS SIGNAL PROCESSING AND NETWORKING

A policy and compliance with a policy is a function of specific parameters available in a node. What should be sensed and what additional parameters should be fed to the policy software to determine policy compliance? Parameters may be raw or processed sensor outputs or environmental parameters and might include security codes or keys. Environmental parameters might include node identity, node location (e.g., from a GPS sensor), or time of day as well as the location of nearby broadcasters, e.g., television. Node location can be heavily leveraged, and policy could be both time and location dependent, e.g., perhaps it would be desirable that elevated nodes have more restrictive transmission power levels during the day. Location might also be used along with a propagation model to estimate signal levels and their compliance with policy such as a prescribed spectral level or mask.

The most fundamental sensor parameters come from a power spectral estimate, providing a means to estimate spectral occupancy. Interesting extensions include the number and/or locations of other nearby nodes, locations of cooperating nodes, types of message traffic, priorities, delay constraints, and observations about the environment other than just the power spectrum. This is obviously not an exhaustive list of possibilities. In addition to the sensing and environmental parameters, the radio may also have a proposed action, whose compliance with policy may require verification. For example, the frequency and transmit power of a possible transmission might be suggested for permission to transmit. Spectral masks may be employed to determine acceptable levels of power both in-band and in adjacent bands due to spectral leakage, as discussed previously.

The OSA process requires sensing, and a simple digital or analog radiometer is a natural starting place (see “Spectrum Opportunity Detection”). Some have coined the phrase *interference temperature* to refer to spectrum power levels and related masks. Along with temperature is the clear implication of a continuously variable level of interference. A power spectral density (PSD) estimator is easily implemented and not too costly, and its statistics are well known. Regulatory bodies, such as the FCC, will seek simple solutions based on mature technology that do not drive up device or systems costs and are easily understood. However, even with this simplest of detectors, there are a variety of subtleties in the sensing and interaction with policy. While the size of a frequency bin might be obvious in a given legacy case, wireless brings the usual variety of complicating factors such as fading, widely varying local propagation environments, hidden nodes, and indoor/outdoor applications. There are typically many nonstationarities present in the wireless world, raising such questions as what is the desired sensing (look-through) rate to accommodate mobility, and what is the appropriate averaging time? As described above, sensing is inherently a probabilistic process, with the implication for harm in the OSA setting given an incorrect hypothesis test outcome. It is therefore interesting to contemplate policies that include probabilities or confidence levels.

[THE OPTIMAL ACCESS STRATEGY SHOULD TAKE INTO ACCOUNT THE OPERATING CHARACTERISTICS OF THE DETECTOR.]

As discussed in "Spectrum Opportunity Detection," more sensitive and sophisticated detectors are available, such as those based on cyclic statistics. These are particularly effective for cyclostationary signals such as digital broadcast with long duration emission and fixed known signal parameters.

The emergence of both sensor networks and MIMO technologies brings a variety of possible extensions to the generic single sensor case. A single node may incorporate array processing and thus spatial detection. This will facilitate significant MAC improvements in wireless networks, and can also facilitate OSA for array-equipped nodes. In analogy to sensor networks, nodes may cooperate to perform distributed detection and cooperative transmission, with many potential benefits. Distributed detection approaches may overcome adverse local fading effects to a large extent, but this approach adds communications overhead. Multisensor techniques appear to complicate the policy definition, e.g., consider that the location and number of cooperating nodes may be highly variable. In addition, going beyond detection, signal parameters or features may be estimated and signal classifiers employed, such as modulation classification [41].

Another approach to facilitating wireless networking is to take advantage of the broadcast medium by employing beacons or control channels. Beacons facilitate medium access and so facilitate channel sensing, and thus they may ultimately prove to be integral to the success of OSA schemes. Beacons can be deployed to define permit-use or deny-use areas, and control channels are integral to centralized systems such as cellular. Thus, beacons may be a simple adjunct to facilitate OSA with legacy systems. A drawback is that, for military systems at least, the use of beacons is problematic from a security and vulnerability standpoint.

Thus, the type and capability of the sensor can play a fundamental role in policy, suggesting a device-based policy. This could be in a hierarchy, e.g., a more sophisticated detector would enable a more aggressive policy, because the more capable detector would presumably lead to better decisions as discussed previously.

POLICY REASONING

Given a set of numerical parameters, it is straightforward to determine policy compliance. However, it may be highly desirable to have a policy reasoner (PR). The need for reasoning arises when a request is not posed in a yes/no answerable form. What frequencies are available at power P_0 in frequency band B_0? What constraints must be met to allow a certain transmission? This provides options for the radio, and requires an interaction between the radio and the PR. Interaction between the sensor/radio and the PR are highly desirable, and so defining and standardizing this interface and possible interactive behaviors is important. Wilkins et al. [42] have defined a policy reasoning language specifically for OSA, and an interface with three possible responses; "yes," "no," and "yes with constraints." In the last case, the PR provides additional constraints that the radio would need to satisfy in order to enable the requested transmission. Example constraints are transmit power limit, transmit duration limit, and so on. This opens the question as to what is a rich enough set of constraints, and this is subject to the particular device capabilities, e.g., whether it supports power control and over what range.

Now is the time to carefully explore the interaction of policy, signal processing, and networking to study systems tradeoffs (complexity versus benefit, etc.), and to provide the tools to the systems designers and the governmental policy makers. Many specifics will be determined by the nature of legacy systems. Technologies will evolve to enable more and more sophisticated signal processing, at cheaper cost, so policies will also need to evolve. This clearly motivates the need for a drop-in policy reasoner approach. Extensions must accommodate security aspects, and further incorporate networking and network constraints.

CONCLUDING REMARKS

Opportunistic spectrum access is still in its infancy. Many complex issues in technical, economical, and regulatory aspects need to be addressed before its potential can be assessed and realized. Research efforts in the signal processing community are particularly important in providing technical data for the crafting of spectrum regulatory policies.

In this article, we have provided an overview of major technical and regulatory issues in OSA. Given the complexity of the topic and the diversity of existing technical approaches, our presentation is by no means exhaustive. We hope that this article provides a glimpse of the technical and regulatory challenges of OSA and serves as an initial documentation of exciting research activities in the signal processing, networking, and regulatory communities.

ACKNOWLEDGMENTS

This work was supported by the Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011 and by the National Science Foundation under Grants CNS-0627090 and ECS-0622200.

AUTHORS

Qing Zhao (qzhao@ece.ucdavis.edu) received the Ph.D. degree in electrical engineering in 2001 from Cornell University. In 2004, she joined the ECE department at University of California at Davis, where she is currently an assistant professor. Her research interests are in the general area of signal processing, communications, and wireless networking. She received the 2000 IEEE Signal Processing Society Young Author Best Paper Award. She is an associate editor of IEEE Transactions on Signal Processing and an elected member of the Signal Processing for Communications committee of the IEEE Signal Processing Society.

Brian M. Sadler is a senior research scientist at the Army Research Laboratory in Adelphi, Maryland. He has been an editor for several journals in communications and signal processing, and he received a Best Paper Award from the Signal Processing Society in 2006. His research interests include signal processing for mobile wireless and ultra-wideband systems, and sensor signal processing and networking, and he supports DARPA programs in these areas. He is a Fellow of the IEEE.

REFERENCES

[1] FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group," Nov. 2002. [Online]. Available: http://www.fcc.gov/sptf/reports.html

[2] R. Coase, "The federal communications commission," *J. Law and Economics*, pp. 1–40, 1959.

[3] D. Hatfield and P. Weiser, "Property rights in spectrum: Taking the next step," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, Nov. 2005, pp. 43-55.

[4] L. Xu, R. Tonjes, T. Paila, W. Hansmann, M. Frank, and M. Albrecht, "DRiVE-ing to the Internet: Dynamic radio for IP services in vehicular environments," in *Proc. 25th Annual IEEE Conf. Local Computer Networks*, Nov. 2000, pp. 281–289.

[5] Y. Benkler, "Overcoming agoraphobia: Building the commons of the digitally networked environment," *Harv. J. Law Tech*, vol. 11, no. 2, Winter, 1998.

[6] W. Lehr and J. Crowcroft, "Managing shared access to a spectrum commons," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, Nov. 2005, pp. 420-444.

[7] C. Raman, R. Yates, and N. Mandayam, "Scheduling variable rate links via a spectrum server," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, 2005, pp. 110–118.

[8] O. Ileri, D. Samardzija, and N. Mandayam, "Demand Responsive Pricing and Competitive Spectrum Allocation via a Spectrum Server," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, 2005, pp. 194-202.

[9] S. Chung, S. Kim, J. Lee, and J. Cioffi, "A game-theoretic approach to power allocation in frequency-selective Gaussian interference channels," in *Proc. IEEE Int. Symp. Information Theory*, June 2003, pp. 316–316.

[10] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, 2005, pp. 251-258.

[11] J. Huang, R. Berry, and M. Honig, "Spectrum sharing with distributed interference compensation," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, 2005, pp.88-93.

[12] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Proc. IEEE Int. Workshop Mobile Multimedia Communications*, 1999, pp. 3–10.

[13] J. Mitola, *Software Radios: Wireless Architecture for the 21st Century*. New York: Wiley, 2000.

[14] J. Mitola, "Cognitive Radio." Licentiate proposal, KTH, Stockholm, Sweden.

[15] Q. Zhao, "Spectrum opportunity and interference constraint in opportunistic spectrum access," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, April 2007.

[16] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Selected Areas in Commun.: Special Issue on Adaptive, Spectrum Agile and Cognitive Wireless Networks*, vol. 25, no. 3, Apr. 2007.

[17] B. Wild and K. Ramchandran, "Detecting primary receivers for cognitive radio applications," in *Proc. IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, Nov. 2005, pp. 124-130.

[18] D. Cabric, S.M. Mishra, and R.W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proc. 38th. Asilomar Conf. Signals Systems, Computers*, 2004, pp. 772–776.

[19] W. Gardner, "Signal interception: A unifying theoretical framework for feature detection," *IEEE Trans. Commun.*, vol. 36, pp. 897–906, Aug. 1988.

[20] A. Sahai and N. Hoven and R. Tandra, "Some fundamental limits on cognitive radio," in *Proc. Allerton Conf. Communication, Control, Computing*, Oct. 2004.

[21] A. Ghasemi and E. Sousa, "Collaborative Spectrum Sensing for Opportunistic Access in Fading Environments," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, 2005, pp. 131-136.

[22] S. Mishra, A. Sahai, and R. Brodersen, "Cooperative sensing among cognitive radios," in *Proc. ICC*, 2006, pp. 1658-1663.

[23] Y. Chen, Q. Zhao, and A. Swami, "Joint design and separation principle for opportunistic spectrum access in the presence of sensing errors," in *Proc. IEEE Asilomar Conf. Signals, Systems, Computers*, 2006.

[24] Y. Chen, Q. Zhao, and A. Swami, "Joint PHY-MAC design for opportunistic spectrum access in the presence of sensing errors," submitted for publication.

[25] Q. Zhao, L. Tong, and A. Swami, "Decentralized cognitive MAC for dynamic spectrum access," in *Proc. IEEE Symp. New Frontiers Dynamic Spectrum Access Networks (DySPAN)*, Nov. 2005, pp.224-232.

[26] Q. Zhao and A. Swami, "A decision-theoretic framework for dynamic spectrum access," to appear in *IEEE Wireless Commun. Mag.: Special Issue on Cognitive Wireless Networks*, 2007.

[27] S. Geirhofer, L. Tong, and B. Sadler, "A measurement-based model for dynamic spectrum access in WLAN channels," in *Proc. IEEE Military Communication Conf.*, 2006.

[28] J. Kleider, S. Gifford, K. Nolan, D. Hughes, and S. Chuprun, "Demonstrating robust high data rate capability on a software defined radio using anti-jam wideband OFDM waveforms," in *Proc. IEEE Military Communication Conf.*, 2005.

[29] T. Weiss and F. Jondral, "Spectrum pooling: An innovative strategy for enhancement of spectrum efficiency," *IEEE Commun. Mag.*, vol. 42, pp. 8–14, Mar. 2004.

[30] U. Berthold and F.K. Jondral, "Guidelines for designing OFDM overlay systems," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, Nov. 2005, pp. 626-629.

[31] H. Tang, "Some physical layer issues of wide-band cognitive radio systems," in *Proc. 1st IEEE Symposium New Frontiers Dynamic Spectrum Access Networks*, Nov. 2005, pp. 151-159.

[32] H. Zheng and C. Peng, "Collaboration and fairness in opportunistic spectrum access," in *Proc. IEEE Int. Conf. Communications (ICC)*, 2005, pp. 3132-3136.

[33] W. Wang and X. Liu, "List-coloring based channel allocation for open-spectrum wireless networks," in *Proc. IEEE VTC*, 2005, pp. 690-694.

[34] S. Sankaranarayanan, P. Papadimitratos, A. Mishra, and S. Hershey, "A bandwidth sharing approach to improve licensed spectrum utilization," in *Proc. 1st IEEE Symp. New Frontiers Dynamic Spectrum Access Networks*, 2005, pp. 279-288.

[35] M. Halldorsson, J. Halpern, L. Li, and V. Mirrokni, "On spectrum sharing games," in *Proc. 23rd Annual ACM Symp. Principles Distributed Computing*, 2004, pp. 107–114.

[36] J. Chapin, D.C. Sicker, "Safety and certification for new radio technologies," *IEEE Commun. Mag.*, vol.44, no. 9, pp. 30–32, Sept. 2006.

[37] "Broadband Radio Access Networks (BRAN); 5 GHz high performance RLAN; Harmonized EN covering essential requirements of article 3.2 of the R&TTE Directive," ETSI, EN 301 893 V1.2.3, Aug. 2005.

[38] ETSI EN 301 893 V1.3.1 (2005-08), Candidate Harmonized European Standard (Telecommunications series) Broadband Radio Access Networks (BRAN); 5 GHz high performance RLAN; Harmonized EN covering essential requirements of article 3.2 of the R\&TTE Directive.

[39] Federal Communications Commission Report and Order FCC 03 287 released November 18, 2003.

[40] A.E. Leu, M. McHenry, and B.L. Mark, "Modeling and analysis of interference in listen-before-talk spectrum access schemes," *Int. J. Network Manage.*, vol. 16, no. 1, pp. 131–147, 2006.

[41] A. Swami and B.M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Communications*, vol. 48, no. 3, pp. 416–429, Mar. 2000.

[42] D. Wilkins, G. Denker, M.-O. Stehr, "Policy-based cognitive radios," submitted for publication.