Advances in Cognitive Radio Networks: A Survey
Beibei Wang and K. J. Ray Liu

Abstract—With the rapid deployment of new wireless devices and applications, the last decade has witnessed a growing demand for wireless radio spectrum. However, the fixed spectrum assignment policy becomes a bottleneck for more efficient spectrum utilization, under which a great portion of the licensed spectrum is severely under-utilized. The inefficient usage of the limited spectrum resources urges the spectrum regulatory bodies to review their policy and start to seek for innovative communication technology that can exploit the wireless spectrum in a more intelligent and flexible way. The concept of cognitive radio is proposed to address the issue of spectrum efficiency and has been receiving an increasing attention in recent years, since it equips wireless users the capability to optimally adapt their operating parameters according to the interactions with the surrounding radio environment. There have been many significant developments in the past few years on cognitive radios. This article surveys recent advances in research related to cognitive radios. The fundamentals of cognitive radio technology, architecture of a cognitive radio network and its applications are first introduced. The existing works in spectrum sensing are reviewed, and important issues in dynamic spectrum allocation and sharing are investigated in detail.

I. INTRODUCTION

The usage of radio spectrum resources and the regulation of radio emissions are coordinated by national regulatory bodies like the Federal Communications Commission (FCC). The FCC assigns spectrum to licensed holders, also known as primary users, on a long-term basis for large geographical regions. However, a large portion of the assigned spectrum remains under utilized as illustrated in Figure 1. The inefficient usage of the limited spectrum necessitates the development of dynamic spectrum access techniques, where users who have no spectrum licenses, also known as secondary users, are allowed to use the temporarily unused licensed spectrum. In recent years, the FCC has been considering more flexible and comprehensive uses of the available spectrum [1], through the use of cognitive radio technology [2].

Cognitive radio is the key enabling technology that enables next generation communication networks, also known as dynamic spectrum access (DSA) networks, to utilize the spectrum more efficiently in an opportunistic fashion without interfering with the primary users. It is defined as a radio that can change its transmitter parameters according to the interactions with the environment in which it operates [4]. It differs from conventional radio devices in that a cognitive radio can equip users with cognitive capability and reconfigurability [5] [3]. Cognitive capability refers to the ability to sense and gather information from the surrounding environment, such as information about transmission frequency, bandwidth, power, modulation, etc. With this capability, secondary users can identify the best available spectrum. Reconfigurability refers to the ability to rapidly adapt the operational parameters according to the sensed information in order to achieve the optimal performance. By exploiting the spectrum in an opportunistic fashion, cognitive radio enables secondary users to sense which portion of the spectrum are available, select the best available channel, coordinate spectrum access with other users, and vacate the channel when a primary user reclaims the spectrum usage right.

Considering the more flexible and comprehensive use of the spectrum resources, especially when secondary users coexist with primary users, traditional spectrum allocation schemes [6] and spectrum access protocols may no longer be applicable. New spectrum management approaches need to be developed to solve new challenges in research related to cognitive radio, specifically, in spectrum sensing and dynamic spectrum sharing.

As primary users have priority in using the spectrum, when secondary users coexist with primary users, they have to perform real-time wideband monitoring of the licensed spectrum to be used. When secondary users are allowed to transmit data simultaneously with a primary user, interference temperature limit should not be violated [7]. If secondary users are only allowed to transmit when the primary users are not using the spectrum, they need to be aware of the primary users’ reappearance through various detection techniques, such as energy detection, feature detection, matched filtering and coherent detection. Due to noise uncertainty, shadowing, and multipath effect, detection performance of single user sensing is pretty limited. Cooperative sensing has been considered

Copyright©2008 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.
effective in improving detection accuracy by taking advantage of the spatial and multi-user diversity. In cooperative spectrum sensing, how to select proper users for sensing, how to fuse individual user’s decision and exchange information, and how to perform distributed spectrum sensing are issues worth studying.

In order to fully utilize the spectrum resources, efficient dynamic spectrum allocation and sharing schemes are very important. Novel spectrum access control protocols and control channel management should be designed to accommodate the dynamic spectrum environment while avoid collision with a primary user. When a primary user re-appears in a licensed band, a good spectrum handoff mechanism is required to provide secondary users with smooth frequency transition with low latency. In multi-hop cognitive wireless networks, intermediate cognitive nodes should intelligently support relaying information and routing through using a set of dynamically changing channels. In order to manage the interference to the primary users and the mutual interference among themselves, secondary users’ transmission power should be carefully controlled, and their competition for the spectrum resources should also be addressed.

There have been many significant developments in the past few years on cognitive radios. This article surveys recent advances in research related to cognitive radios. In Section II, we overview the fundamentals of cognitive radio technology, architecture of a cognitive radio network and its applications. In Section III, we review existing works in spectrum sensing, including interference temperature, different types of detection techniques, and cooperative spectrum sensing. In Section IV we discuss several important issues in dynamic spectrum allocation and sharing.

II. FUNDAMENTALS

A. Cognitive Radio Characteristics

The dramatic increase of service quality and channel capacity in wireless networks is severely limited by the scarcity of energy and bandwidth, which are the two fundamental resources for communications. Therefore, researchers are currently focusing their attention on new communications and networking paradigms that can intelligently and efficiently utilize these scarce resources. Cognitive radio (CR) is one critical enabling technology for future communications and networking that can utilize the limited network resources in a more efficient and flexible way. It differs from traditional communication paradigms in that the radios/devices can adapt their operating parameters, such as transmission power, frequency, modulation type, etc., to the variations of the surrounding radio environment [4]. Before CRs adjust their operating mode to environment variations, they must first gain necessary information from the radio environment. This kind of characteristics is referred to as cognitive capability [5], which enables CR devices to be aware of the transmitted waveform, radio frequency (RF) spectrum, communication network type/protocol, geographical information, locally available resources and services, user needs, security policy, and so on. After CR devices gather their needed information from the radio environment, they can dynamically change their transmission parameters according to the sensed environment variations and achieve optimal performance, which is referred to as reconfigurability [5].

B. Cognitive Radio Functions

A typical duty cycle of CR, as illustrated in Figure 2, includes detecting spectrum white space, selecting the best frequency bands, coordinating spectrum access with other users and vacating the frequency when a primary user appears. Such a cognitive cycle is supported by the following functions:

- spectrum sensing and analysis;
- spectrum management and handoff;
- spectrum allocation and sharing.

Through spectrum sensing and analysis, CR can detect the spectrum white space (see Figure 3), i.e., a portion of frequency band that is not being used by the primary users, and utilize the spectrum. On the other hand, when primary users start using the licensed spectrum again, CR can detect their activity through sensing, so that no harmful interference is generated due to secondary users’ transmission.

After recognizing the spectrum white space by sensing, spectrum management and handoff function of CR enables secondary users to choose the best frequency band and hop among multiple bands according to the time varying channel characteristics to meet various Quality of Service (QoS) requirements [3]. For instance, when a primary user reclaim his/her frequency band, the secondary user that is using the licensed band can direct his/her transmission to other available frequencies, according to the channel capacity determined by the noise and interference levels, path loss, channel error rate, holding time, and etc.

In dynamic spectrum access, a secondary user may share the spectrum resources with primary users, other secondary users, or both. Hence, a good spectrum allocation and sharing mechanism is critical to achieve high spectrum efficiency. Since primary users own the spectrum rights, when secondary users co-exist in a licensed band with primary users, the interference level due to secondary spectrum usage should be limited by a certain threshold. When multiple secondary users
A secondary network refers to a network composed of a set of secondary users with/without a secondary base station. Secondary users can only access the licensed spectrum when it is not occupied by a primary user. The opportunistic spectrum access of secondary users is usually coordinated by a secondary base station, which is a fixed infrastructure component serving as a hub of the secondary network. Both secondary users and secondary base stations are equipped with CR functions. If several secondary networks share one common spectrum band, their spectrum usage may be coordinated by a central network entity, called spectrum broker [8]. The spectrum broker collects operation information from each secondary network, and allocates the network resources to achieve efficient and fair spectrum sharing.

A primary network is composed of a set of primary users and one or more primary base stations. Primary users are authorized to use certain licensed spectrum bands under the coordination of primary base stations. Their transmission should not be interfered by secondary networks. Primary users and primary base stations are in general not equipped with CR functions. Therefore, if a secondary network share a licensed spectrum band with a primary network, besides detecting the spectrum white space and utilizing the best spectrum band, the secondary network is required to immediately detect the presence of a primary user and direct the secondary transmission to another available band so as to avoid interfering with primary transmission.

Because CRs are able to sense, detect and monitor the surrounding RF environment such as interference and access availability, and reconfigure their own operating characteristics to best match outside situations, cognitive communications can increase spectrum efficiency and support higher bandwidth service. Moreover, the capability of real-time autonomous decisions for efficient spectrum sharing also reduces the burdens of centralized spectrum management. As a result, CRs can be employed in many applications.

First, the capacity of military communications is limited by radio spectrum scarcity because static frequency assignments freeze bandwidth into unproductive applications, where a large amount of spectrum is idle. CR using dynamic spectrum access can alleviate the spectrum congestion through efficient allocation of bandwidth and flexible spectrum access [2]. Therefore, CR can provide military with adaptive, seamless, and secure communications.

Moreover, a CR network can also be implemented to enhance public safety and homeland security. A natural disaster or terrorist attack can destroy existing communication infrastructure, so an emergency network becomes indispensable to aid the search and rescue. As a CR can recognize spectrum availability and reconfigure itself for much more efficient communication, this provides public safety personnel with dynamic spectrum selectivity and reliable broadband communication to minimize information delay. Moreover, CR can facilitate interoperability between various communication systems. Through adapting to the requirements and conditions of another network, the CR devices can support multiple service types, such as voice, data, video, and etc.

Another very promising application of CR is in the commercial markets for wireless technologies. Since CR can intelligently determine which communication channels are in use and automatically switches to an unoccupied channel, it provides additional bandwidth and versatility for rapidly growing data applications. Moreover, the adaptive and dynamic channel switching can help avoid spectrum conflict and expensive redeployment. As CR can utilize a wide range of frequencies, some of which has excellent propagation characteristics, CR devices are less susceptible to fading related to growing foliage, buildings, terrain and weather. When frequency changes are needed due to conflict or interference, the CR frequency management software will change the operating frequency automatically even without human intervention. Additionally, the radio software can change the service bandwidth remotely to accommodate new applications. As long as no end-user hardware needs to be updated, product upgrades or configuration changes can be completed simply by downloading newly released radio management software. Thus, CR is viewed as the key enabling technology for future mobile wireless services anywhere, anytime and with any device.

III. SPECTRUM SENSING AND ANALYSIS

Through spectrum sensing, CR can obtain necessary observations about its surrounding radio environment, such as the presence of primary users and appearance of spectrum holes. Only with this information can CR adapt its transmitting and receiving parameters, like transmission power, frequency, modulation schemes, and etc., in order to achieve efficient spectrum utilization. Therefore, spectrum sensing and analysis is the first critical step towards dynamic spectrum management. In this section, we will discuss three different aspects of spectrum sensing. First is the interference temperature model, which measures the interference level observed at a receiver
and is used to protect licensed primary users from harmful interference due to unlicensed secondary users. Then we will talk about the spectrum hole detection to determine additional available spectrum resources and compare several detection techniques. Finally we will discuss cooperative sensing with multiple users’ help.

A. Interference Temperature

In opportunistic spectrum access, secondary users need to detect primary users’ appearance and decide which portion of the spectrum is available. Such a decision can be made according to different metrics. Traditional approach is to limit the transmitter power of interfering devices, i.e., the transmitted power should be no more than a prescribed noise floor at a certain distance from the transmitter. However, due to the increased mobility and variability of radio frequency (RF) emitters, constraining the transmitter power becomes more problematic, since unpredictable new source of interference may appear. To address this issue, FCC Spectrum Policy Task Force [9] has proposed a new metric on interference assessment, the interference temperature, to enforce an interference limit perceived by receivers. The interference temperature is a measure of the RF power available at a receiving antenna to be delivered to a receiver, reflecting the power generated by other emitters and noise sources [10]. More specifically, it is defined as the temperature equivalent to the RF power available at a receiving antenna per unit bandwidth [11], i.e.,

$$T_I(f_c, B) = \frac{P_I(f_c, B)}{kB},$$  \hspace{1cm} (1)$$

where $P_I(f_c, B)$ is the average interference power in Watts centered at $f_c$, covering bandwidth $B$ measured in Hertz, and Boltzmann’s constant $k$ is $1.38 \times 10^{-23}$ Joules per degree Kelvin.

With the concept of interference temperature, FCC further established an interference temperature limit, which provides a maximum amount of tolerable interference for a given frequency band at a particular location. Any unlicensed secondary transmitter using this band must guarantee that their transmission plus the existing noise and interference must not exceed the interference temperature limit at a licensed receiver.

Since any transmission in the licensed band is viewed to be harmful if it would increase the noise floor above the interference temperature limit, it is necessary that the receiver have a reliable spectral estimate of the interference temperature. This requirement can be fulfilled by using the multitaper method to estimate the power spectrum of the
interference temperature with a large number of sensors [5]. The multitaper method can solve the tradeoff between bias and variance of an estimator and provide a near-optimal estimation performance. The large number of sensors can account for the spatial variation of the RF energy from one location to another. Subspace-based method has also been proposed to gain knowledge of the quality and usage of a spectrum band [12], where information about the interference temperature is obtained by eigenvalue decomposition.

If a regulatory body sets an interference temperature limit $T_L$ for a particular frequency band with bandwidth $B$, then the secondary transmitters has to keep the average interference below $kBT_L$. Therefore, the interference temperature serves as a cap placed on potential RF energy that could appear on that band, and there are some previous efforts about how to implement efficient spectrum allocation with the interference temperature limit.

Spectrum shaping has been proposed to improve spectrum efficiency [13] in CR networks. More specifically, using interference fitting, a CR senses the shape of the interference power spectrum and create spectra inversely shaped to the current interference environment to take advantage of gaps between the noise floor and the cap of the interference temperature limit.

Interference temperature dynamics in a CR network were investigated in [14] using a hidden Markov model (HMM), where the trained HMM can be used as a sequence generator for secondary nodes to predict the interference temperature of the channel in the future and aid their channel selection for transmission.

A comprehensive analysis has been presented in [15], which quantifies how interference temperature limits should be selected and how those choices affect the range of licensed signals. It is shown that the capacity achieved is a simple function of the number of nodes, the average bandwidth, and the fractional impact to the primary signal’s coverage area. However, as observed by [15], the achievable capacity from the interference temperature model is low, compared to the amount of interference it can cause to primary users.

B. Spectrum Sensing

Spectrum sensing enables the capability of a CR to measure, learn, and be aware of the radio’s operating environment, such as the spectrum availability and interference status. When a certain frequency band is detected as not being used by the primary licensed user of the band at a particular time in a particular position, secondary users can utilize the spectrum, i.e., there exists a spectrum opportunity. Therefore, spectrum sensing can be performed in the time, frequency, and spatial domains. With the recent development of beamforming technology, multiple users can utilize the same channel/frequency at the same time in the same geographical location. Thus, if a primary user does not transmit in all the directions, extra spectrum opportunities can be created for secondary users in the directions where the primary user is not operating, and spectrum sensing needs also to take the angle of arrivals into account [16]. Primary users can also use their assigned bands by means of spread spectrum or frequency hopping, and then secondary users can transmit in the same band simultaneously without severely interfering with primary users as long as they adopt an orthogonal code with respect to the codes adopted by primary users [17]. This creates spectrum opportunities in code domain, but meanwhile requires detection of the codes used by primary users as well as multipath parameters.

A wealth of literature on spectrum sensing focuses on primary transmitter detection based on the local measurements of secondary users, since detecting the primary users that are receiving data is in general very difficult. According to the a priori information they require and the resulting complexity and accuracy, spectrum sensing techniques can be categorized in the following types, which are summarized in Figure 5.

1) Energy Detector: Energy detection is the most common type of spectrum sensing because it is easy to implement and requires no prior knowledge about the primary signal.

Assume the hypothesis model of the received signal is

$$H_0 : y(t) = n(t),$$

$$H_1 : y(t) = hx(t) + n(t),$$

where $x(t)$ is the primary user’s signal to be detected at the local receiver of a secondary user, $n(t)$ is the additive white Gaussian noise, and $h$ is the channel gain from the primary user’s transmitter to the secondary user’s receiver. $H_0$ is a null hypothesis, meaning there is no primary user present in the band, while $H_1$ means the primary user’s presence. The detection statistics of the energy detector can be defined as the average (or total) energy of $N$ observed samples

$$T = \frac{1}{N} \sum_{t=1}^{N} |y(t)|^2. \quad (3)$$

The decision on whether the spectrum is being occupied by the primary user is made by comparing the detection statistics $T$ with a predetermined threshold $\lambda$. The performance of the detector is characterized by two probabilities: the probability of false alarm $P_F$ and the probability of detection $P_D$. $P_F$ denotes the probability that the hypothesis test decides $H_1$ while it is actually $H_0$, i.e.,

$$P_F = \Pr(T > \lambda | H_0). \quad (4)$$

$P_D$ denotes the probability that the test correctly decides $H_1$, i.e.,

$$P_D = \Pr(T > \lambda | H_1). \quad (5)$$

A good detector should ensure a high detection probability $P_D$ and a low false alarm $P_F$, or it should optimize the spectrum usage efficiency (e.g. QoS of a secondary user network) while guaranteeing a certain level of primary user protection. To this end, various approaches have been proposed to improve the efficiency of energy detector based spectrum sensing.

Since the detection performance is very sensitive to the noise power estimate error [18], an adaptive noise level
A novel wideband throughput for the secondary users under the constraint that the duration in a time slot is optimized to maximize the achievable opportunistic throughput of CRs and reduce their interference to the primary systems. The analysis in [26] shows that detection of narrowband transmission using energy detection over multi-band orthogonal frequency-division multiplexing (OFDM) is feasible, and can be further extended to cover more complex systems.

Besides its low computational and implementation complexity and short detection time, there also exist some challenges in designing a good energy detector. First, the detection threshold depends on the noise power, which may change over time and hence is difficult to measure precisely in real time. In low signal to noise ratio (SNR) regimes where the noise power is very high, reliable identification of a primary user is even impossible [27]. Moreover, an energy detector can only decide the primary user’s presence by comparing the received signal energy with a threshold; thus, it cannot differentiate the primary user from other unknown signal sources. As such, it can trigger false alarm frequently.

2) Feature Detector: There are specific features associated with the information transmission of a primary user. For instance, the statistics of the transmitted signals in many communication paradigms are periodic because of the inherent periodicities such as the modulation rate, carrier frequency, etc. Such features are usually viewed as the cyclostationary features, based on which a detector can distinguish cyclo-

<table>
<thead>
<tr>
<th>Type</th>
<th>Test statistics</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy detector</td>
<td>Energy of the received signal samples</td>
<td>• Easy to implement</td>
<td>• High false alarm due to noise uncertainty</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Not require prior knowledge about primary signals</td>
<td>• Very unreliable in low SNR regimes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Cannot differentiate a primary user from other signal sources</td>
</tr>
<tr>
<td>Feature detector</td>
<td>Cyclic spectrum density function of the received signal, or by matching general features of the received signal to the already known primary signal characteristics</td>
<td>• More robust against noise uncertainty and better detection in low SNR regimes than energy detection</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Can distinguish among different types of transmissions and primary systems</td>
<td>• Specific features, e.g., cyclostationary features, must be associated with primary signals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Particular features may need to be introduced, e.g., to OFDM-based communications</td>
</tr>
<tr>
<td>Matched filtering and coherent detection</td>
<td>Projected received signal in the direction of the already known primary signal or a certain waveform pattern</td>
<td>• More robust to noise uncertainty and better detection in low SNR regimes than feature detector</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Require less signal samples to achieve good detection</td>
<td>• Require precise prior information about certain waveform patterns of primary signals</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• High complexity</td>
</tr>
</tbody>
</table>

Fig. 5: Summary of main spectrum sensing techniques.
stationary signals from stationary noise. In a more general sense, features can refer to any intrinsic characteristics associated with a primary user’s transmission, as well as the cyclostationary features. For example, center frequencies and bandwidths [28] extracted from energy detection can also be used as reference features for classification and determining a primary user’s presence. In this section, we will introduce the cyclostationary feature detection followed by a generalized feature detection.

Cyclostationary feature detection was first introduced in [29]. As in most communication systems, the transmitted signals are modulated signals coupled with sine wave carriers, pulse trains, hopping sequences, or cyclic prefixes, while the additive noise is generally wide-sense stationary (WSS) with no correlation, cyclostationary feature detectors can be utilized to differentiate noise from primary users’ signal [30] [31] [32] and distinguish among different types of transmissions and primary systems [33].

Different from an energy detector which uses time-domain signal energy as test statistics, a cyclostationary feature detector performs a transformation from the time-domain into the frequency feature domain and then conducts a hypothesis test in the new domain. Specifically, define the cyclic autocorrelation function (CAF) of the received signal $y(t)$ by

$$ R_{y}^{\alpha}(\tau) = E[y(t + \tau)y^{\ast}(t - \tau)e^{j2\pi\alpha t}], $$

where $E[\cdot]$ is the expectation operation, $\ast$ denotes complex conjugation, and $\alpha$ is the cyclic frequency. Since periodicity is a common property of wireless modulated signals, while noise is WSS, the CAF of the received signal also demonstrates periodicity when the primary signal is present. Thus, we can represent the CAF using its Fourier series expansion, called the cyclic spectrum density (CSD) function, expressed as [29]

$$ S(f, \alpha) = \sum_{\tau = -\infty}^{\infty} R_{y}^{\alpha}(\tau)e^{-j2\pi f \tau}.$$ 

The CSD function have peaks when the cyclic frequency $\alpha$ equals to the fundamental frequencies of the transmitted signal $x(t)$, i.e., $\alpha = \frac{k}{T_{x}}$ with $T_{x}$ being the period of $x(t)$. Under hypothesis $H_{0}$, the CSD function does not have any peaks since the noise is non-cyclostationary signals. A peak detector [34] or a generalized likelihood ratio test [31] [33] can be further used to distinguish among the two hypothesis. Different primary communication systems using different air interfaces (modulation, multiplexing, coding, etc.) can also be differentiated by their different properties of cyclostationarity.

However, when OFDM becomes the air interface as suggested by several wireless communication standards, identification of different systems may become problematic, since the features due to the nature of OFDM signaling are likely to be close or even identical. To address this issue, particular features need to be introduced to OFDM-based communications. In [35], methods that induce different properties of cyclostationarity to different systems are considered. The OFDM signal is configured before transmission so that its CAF outputs peaks at certain pre-chosen cycle frequencies, and the difference in these frequencies is used to distinguish among several systems under the same OFDM air interface. A similar approach is considered in [36].

Compared to energy detectors that are prone to high false alarm probability due to noise uncertainty and cannot detect weak signals in noise, cyclostationary detectors become good alternatives because they can differentiate noise from primary users’ signal and have better detection robustness in low SNR regime. A spectrum sensing method based on maximum cyclic autocorrelation selection has been proposed in [37], where the peak and non-peak values of the cyclic autocorrelation function are compared to determine whether the primary signal is present or not. This method does not require noise variance estimation, and is robust against noise uncertainty and interference signals. Frequency-selective fading and uncertain noise impair the robustness of cyclostationary signal detection in low SNR environments. Run time noise calibration has been considered in [27] and [38], in order to improve detector robustness. The method exploits the in-band measurements at frequencies where a pilot is absent to calibrate the noise statistics at the pilot frequencies. By combining neural network for signal classification with cyclic spectral analysis, a more efficient and reliable classifier is developed in [39]. Since a large amount of processing is performed offline using neural networks, the online computation for signal classification is greatly reduced.

Generalized feature detection refers to detection and classification that extracts more feature information other than the cyclostationarity due to the modulated primary signals, such as the transmission technologies used by a primary user, the amount of energy and its distribution across different frequencies [40] [41], channel bandwidth and its shape [42] [28], power spectrum density [43], center frequency [28], idle guard interval of OFDM [44], FFT-type feature [45], etc. By matching the features extracted from the received signal to the a priori information about primary users’ transmission characteristics, primary users can be identified.

Location information of the primary signal is also an important feature that can be used to distinguish a primary user from other signal sources. Under primary user emulation attack, a malicious secondary user transmits signals whose characteristics emulate those of the primary signals. A transmitter verification scheme is proposed in [46] to secure trustworthy spectrum sensing based on the location verification of the primary user.

3) Matched Filtering and Coherent Detection: If secondary users know information about a primary user’s signal a priori, then the optimal detection method is the matched filtering [47], since a matched filter can correlate the already known primary signal with the received signal to detect the presence of the primary user and thus maximize the SNR in the presence of additive stochastic noise. The merit of matched filtering is the short time it requires to achieve a certain detection performance such as a low probability of missed detection and false alarm [48], since a matched filter needs less received signal samples. However, the required number of signal samples also grows as the received SNR decreases, so there exists a SNR wall [27] for a matched filter. In addition, its implementation complexity and power consumption is too
high [49], because the matched filter needs receivers for all types of signals and corresponding receiver algorithms to be executed.

Matched filtering requires perfect knowledge of the primary user’s signal, such as the operating frequency, bandwidth, modulation type and order, pulse shape, packet format, etc. If wrong information is used for matched filtering, the detection performance will be degraded a lot. On the other hand, most wireless communication systems exhibit certain patterns, such as pilot tones, preambles, midambles, spreading codes, and etc., which are used to assist control, equalization, synchronization, continuity, or reference purposes. Even though perfect information of a primary user’s signal may not be attainable, if a certain pattern is known from the received signals, coherent detection (a.k.a. waveform-based sensing) can be used to decide whether a primary user is transmitting or not [50]. As an example, the procedure of coherent detection using pilot pattern is explained as follows [50].

There are two hypothesis in the coherent detection:

\[ H_0 : y(t) = n(t), \]
\[ H_1 : y(t) = \sqrt{\epsilon} x_p(t) + \sqrt{1 - \epsilon} x(t) + n(t), \] (8)

where \( x_p(t) \) is a known pilot tone, \( \epsilon \) is the fraction of energy allocated to the pilot tone, \( x(t) \) is the desired signals assumed to be orthogonal to the pilot tone, and \( n(t) \) is additive white noise. The test statistics of the coherent detection is defined as the projected received signal in the pilot direction, i.e.,

\[ T = \frac{1}{N} \sum_{i=1}^{N} y(t) \hat{x}_p(t), \] (9)

with \( \hat{x}_p \) is a normalized unit vector in the direction of the pilot tone. As \( N \) increases, the test statistics \( T \) under hypothesis \( H_1 \) is much greater than that under \( H_0 \). By comparing \( T \) with a pre-determined detection threshold, one can decide the presence of a primary user.

Coherent detection can also be performed in frequency domain [43]. One can express the binary hypothesis test using the power spectrum density of the received signal \( S_Y(\omega) \), and distinguish between \( H_0 \) and \( H_1 \) by exploiting the unique spectral signature exhibited in \( S_X(\omega) \).

Coherent detection is shown to be robust to noise uncertainty, and not limited by the SNR wall [50] as \( N \) is large enough. Moreover, coherent detection outperforms energy detection in the sensing convergence time [51] [52], because the sensing time of energy detection increases quadratically with the SNR reduction, while that of coherent detection only increases linearly [52]. However, information about the waveform patterns is a prerequisite for implementing coherent detection; the more precise information a coherent detector has, the better the sensing performance will be.

4) Other techniques: There are several other spectrum sensing techniques proposed in recent literature, and some of them are variations inspired by the above-mentioned sensing techniques.

**Statistical Covariance Based Sensing:** Since the statistical covariance matrices of the received signal and noise are generally different, the difference is used in [53] [54] to differentiate the desired signal component from background noise. The eigenvalues of the covariance matrix of the received signal can also be used for primary detection [55]. Based on random matrix theory [56], the ratio of the maximum eigenvalue to the minimum eigenvalue is quantized, and the detection threshold can be found among them. From the simulation on detecting digital TV signals, these methods based on statistical covariances are shown to be more robust to noise uncertainty while requiring no a priori information of the signal, the channel, and noise power.

**Filter Based Sensing:** Application of a specific class of filter banks is proposed in [57] for spectrum sensing in CR systems. When filter banks are used for multicarrier communications in CR networks, the spectrum sensing can be performed by only measuring the signal power at the outputs of subcarrier channels with virtually no computational cost. The multitaper method [5] can also be thought as a filter bank spectrum estimation with multiple filter banks.

**Fast Sensing:** By utilizing the theory of quickest detection, which performs a statistical test to detect the change of distribution in spectrum usage observations as quickly as possible, an agile and robust spectrum sensing is achieved in [58]. The unknown parameters after a primary user appears can be estimated using the proposed successive refinement, which combines both generalized likelihood ratio and parallel cumulative sum tests. An efficient sensing-sequence is developed in [59] to reduce the delay due to spectrum opportunity discovery. The probability that a frequency band is available at sensing, the sensing duration and the channel capacity are three factors that determine the sensing sequence.

**Learning/Reasoning Based Sensing:** An approach based on reinforcement learning for the detection of spectral resources in a multi-band CR scenario is investigated in [60], where the optimal detection strategy is obtained by solving a Markov decision process (MDP). A medium access control layer spectrum sensing algorithm using knowledge-based reasoning is proposed in [61], where the optimal range of channels to finely sense is determined through proactive fast sensing and channel quality information.

**Measurements Based Sensing and Modeling:** By collecting data over a long period of time at many base stations, [62] provides a unique analysis of cellular primary usage. The collected data is dissected along different dimensions to characterize the primary usage. With the aid of spectrum observatory, [63] extends short-term spectrum usage measurements to study the spectrum usage trend over long periods, observes spectrum usage patterns, and detects the positions of spectrum white space in time and spatial domains. Such information can be greatly helpful in developing good dynamic access protocols and governing secondary systems.

C. Cooperative Sensing

The performance of spectrum sensing is limited by noise uncertainty, shadowing, and multi-path fading effect. When the received primary SNR is too low, there exists a SNR wall, below which reliable spectrum detection is impossible even with a very long sensing time. If secondary users can not detect
the primary transmitter, while the primary receiver is within the secondary users’ transmission range, a hidden primary user problem will occur, and the primary user’s transmission will be interfered.

By taking advantage of the independent fading channels (i.e., spatial diversity) and multiuser diversity, cooperative spectrum sensing is proposed to improve the reliability of spectrum sensing, increase the detection probability to better protect a primary user, and reduce false alarm to utilize the idle spectrum more efficiently. In centralized cooperative spectrum sensing, a central controller, e.g., a secondary base station, collects local observations from multiple secondary users, decides the available spectrum channels using some decision fusion rule, and informs the secondary users which channels to access. In distributed cooperative spectrum sensing, secondary users exchange their local detection results among themselves without requiring a backbone infrastructure with reduced cost. Relays can also be used in cooperative spectrum sensing, such as the cooperative sensing scheme proposed in [64], where the cognitive users operating in the same band help each other relay information using amplify-and-forward protocol. It is shown that the inherent network asymmetry can be exploited to increase the agility.

There also exist several challenges on cooperative spectrum sensing. For instance, secondary users can be low-cost devices only equipped with a limit amount of power, so they can not afford very complicated detection hardware and high computational complexity. In wideband cooperative sensing, multiple secondary users have to scan a wide range of spectrum channels and share their detection results. This results in a large amount of sensory data exchange, high energy consumption, and an inefficient data throughput. If the spectrum environment is highly dynamic, the sensed information may even be stale due to user mobility, channel fading, etc.

1) User Selection: Due to secondary users’ different locations and channel conditions, it is shown in [65] that cooperating all secondary users in spectrum sensing is not optimal, and the optimum detection/false alarm probability is achieved by only cooperating a group of users who have higher SNR of the received primary signal.

Since detecting a primary user costs battery power of secondary users, and shadow fading may be correlated for nearby secondary users, an optimal selection of secondary users for cooperative spectrum sensing is desirable. In [66], different algorithms based on different amount of available information are proposed to select a proper set of sensors that experience uncorrelated shadow fading. A joint spatial-temporal sensing scheme for CR networks is proposed in [67], where secondary users collaboratively estimate the location and transmit power of the primary transmitter to determine their maximum allowable transmission power, and use the location information to decide which users should participate in collaborative sensing in order to minimize correlation among the secondary users. Performance evaluation of cooperative spectrum sensing over realistic propagation environments, i.e., correlated log-normal shadowing in both sensing and reporting channels, is investigated in [68]. This work also provides guidelines to select the optimal number of users in order to guarantee a certain detecting performance in a practical radio environment.

In a CR sensor network, individual sensor nodes may experience heterogeneous false alarm and detection probability due to their different locations, making it harder to determine the optimal number of cooperative nodes. Sensor clustering is proposed in [69], where the optimal cluster size is derived so as to upper-bound the variation of the average received signal strength in a cluster of sensor nodes. Moreover, sensor density is optimized so that average distance between neighboring nodes is lower-bounded and their measurements are nearly independent without much correlation.

If a secondary user can not distinguish between the transmissions of a primary user and another secondary user, he will lose the opportunity to use the spectrum. It is shown in [50] that the presence/absence of possible interference from other secondary users is the main reason of the uncertainty in primary user detection, and coordinating nearby secondary users can greatly reduce the noise uncertainty due to shadowing, fading, and multi-path effect. A good degree of coordination should be chosen based on the channel coherent times, bandwidths, and the complexity of the detectors.

2) Decision Fusion: Different decision fusion rules for cooperative spectrum sensing have been studied in the literature. Logic OR rule is used [70] for combining multiple users’ decisions for spectrum sensing in fading environments. Cooperative spectrum sensing using counting rule is studied in [71], where sensing errors are minimized by choosing the optimal settings for both matched filtering and energy detection. It is shown in [72] that half-voting rule is the optimal decision fusion rule in cooperative sensing based on energy detection. Light-weight cooperation based on hard decisions is proposed [73] for cooperative sensing to alleviate the sensitivity requirements on individual users. A liner-quadratic strategy is developed [74] to combat the detriment effects of correlation between different secondary users.

A good way to optimally combine the received primary signal samples in space and time is to maximize the SNR of local energy detectors. However, optimal combination requires information of the signal and channel. Blindly combined energy detection is proposed in [75], without requiring such information and noise power estimation, while performing much better than energy detector and more robust to noise uncertainty. Hard decision combining with the logic AND rule and soft decision using the likelihood ratio test are proposed in [76] in collaborative detection of TV transmissions. It is shown that soft decision combining for spectrum sensing yields more precise detection than hard decision combining. Soft decision combination for cooperative sensing based on energy detection is investigated in [77], and maximal ratio combination is proved to be near optimal in low SNR region and reduce the SNR wall.

In general, cooperative sensing is coordinated over a separate control channel, so a good cooperation schemes should be able to use a small bandwidth and power for exchanging local detection results while maximizing the detection reliability. An efficient linear cooperation framework for spectrum sensing is proposed in [78], where the global decision is a linear
combination of the local statistics collected from individual nodes using energy detection. Compared to the likelihood ratio test, the proposed method has lower computational complexity, closed-form expressions of detection and false alarm probabilities, and comparable detection performance.

Performance of cooperative spectrum sensing depends on the correctness of the local sensing data reported by the secondary users. If malicious users enter a legitimate secondary network and compromise the secondary users, false detection results will be reported to the fusion center, and this kind of attack is called spectrum sensing data falsification (SSDF) attack [79]. In order to guarantee satisfying detection performance under the SSDF attack, a weighted sequential probability ratio test (WSPRT) is proposed in [79], which incorporates a reputation-based mechanism to the sequential probability ratio test. If a secondary user’s local detection result is identical to the final result after decision fusion, his/her reports will carry more weight in future decision fusion. The proposed WSPRT approach is more robust against the SSDF attack than commonly adopted decision fusion rules, such as AND, OR, and majority rules [80].

3) Efficient Information Sharing: In order to coordinate the cooperation in spectrum sensing, a lot of information exchange is needed among secondary users, such as their locations, estimation of the primary user’s location and power, which users should be clustered into a group, which users should perform cooperative sensing at a particular time epoch, and so on. Such a large amount of information exchange brings a lot of overhead to the secondary users, which necessitates an efficient information sharing among the secondary users.

GUESS protocol, an incremental gossiping approach is proposed in [81], to coordinate the dissemination of spectrum sensing results with reduced overhead. In order to reduce the bandwidth required by a large number of secondary users for reporting their sensing results, a censoring method with quantization is proposed in [82]. Only users with reliable information will send their local observations, i.e. one bit decision 0 or 1, to the common receiver. A pipelined spectrum sensing framework is proposed in [83], where spectrum sensing is conducted concurrently with when secondary users are sending their detection reports. The proposed method alleviates sensing overhead by making use of the reporting time, provides more time for spectrum sensing, and thus improves the detection performance.

4) Distributed Cooperative Sensing: Cooperative spectrum sensing has been shown to be able to greatly improve the sensing performance in CR networks. However, if cognitive users belong to different service providers, they tend to contribute less in sensing in order to increase their own data throughput. A distributed cooperating spectrum sensing scheme based on evolutionary game theory is proposed in [84] to answer the question of “how to collaborate” in multiuser de-centralized CR networks. Using replicator dynamics, the evolutionary game modeling provides an excellent means to address the strategic uncertainty that a user may face by exploring different actions, adaptively learning during the strategic interactions, and approaching the best response strategy under changing conditions and environments. The behavior dynamics and the optimal cooperation strategy of the secondary users are characterized. A distributed learning algorithm is further developed so that the secondary users approach the optimal strategy solely based on their own payoff observations. The proposed game is demonstrated to achieve a higher system throughput than the fully cooperative scenario, where all users contribute to sensing in every time slot.

IV. Dynamic Spectrum Allocation and Sharing

In the previous section, we have discussed various detection techniques and how to perform efficient cooperative spectrum sensing in order to obtain an accurate estimation of the interference temperature and spectrum occupancy status. With the detection results, secondary users will have an idea on which spectrum bands he/she could use. However, the availability and quality of a spectrum band may change rapidly with time due to primary users’ activity and competition from other secondary users. In order to utilize the spectrum resources efficiently, secondary users need to be able to address issues such as when and how to use a spectrum band, how to co-exist with primary users and other secondary users, and which spectrum band they should sense and access if the current one in use is not available. Therefore, in this section, we will review the existing spectrum allocation and sharing approaches that answer these questions.

Before going into details, we would like to first briefly discuss the classification of the current spectrum allocation and sharing schemes. According to different criteria, existing spectrum allocation and sharing schemes can be classified in different types, as summarized in Figure 6.

The first classification is according to the spectrum bands that secondary users are using. Spectrum sharing among the secondary users who access the unlicensed spectrum band is referred to as open spectrum sharing. One example is the open spectrum sharing in the unlicensed industrial, scientific, and medical band. In open spectrum sharing, since no users own spectrum licenses, they all have the same rights in using the unlicensed spectrum. Spectrum sharing among the secondary users and primary users in licensed spectrum bands is referred to as hierarchical access model [85] or licensed spectrum sharing. Primary users, usually not equipped with CR, do not need to perform dynamic/opportunistic spectrum access, since they have priority in using the spectrum band. Whenever they reclaim the spectrum usage, secondary users have to adjust their operating parameters, such as power, frequency, and bandwidth, to avoid interrupting the primary users.

Considering the access technology of the secondary users, licensed spectrum sharing can be further divided in two categories [3] [85].

1) Spectrum underlay: In spectrum underlay secondary users are allowed to transmit their data in the licensed spectrum band when primary users are also transmitting. The interference temperature model is imposed on secondary users’ transmission power so that the interference at a primary user’s receiver is within the interference temperature limit and primary users can deliver their packet to the receiver successfully. Spread spectrum
techniques are usually adopted by secondary users to fully utilize the wide range of spectrum. However, due to the constraints on transmission power, secondary users can only achieve short-range communication. If primary users transmit data all the time in a constant mode, spectrum underlay does not require secondary users to perform spectrum detection to find available spectrum bands.

2) **Spectrum overlay**: Spectrum overlay is also referred to as opportunistic spectrum access. Unlike spectrum underlay, secondary users in spectrum overlay will only use the licensed spectrum when primary users are not transmitting, so there is no interference temperature limit imposed on secondary users’ transmission. Instead, secondary users need to sense the licensed frequency band and detect the spectrum white space, in order to avoid harmful interference to primary users.

The second classification is according to the network architecture [3]. When there exists a central entity that controls and coordinates the spectrum allocation and access of secondary users, then the spectrum allocation is *centralized*. If there is no such a central controller, perhaps because of the high cost of constructing an infrastructure or the ad-hoc nature of the network such as for emergency or military use, that kind of spectrum sharing belongs to *distributed* spectrum sharing. In distributed spectrum sharing, each user makes his/her own decision about his/her spectrum access strategy, mainly based on local observation of the spectrum dynamics.

The third classification is according to the access behavior of secondary users [3]. If all secondary users work towards a common goal, for instance they belong to the same operator or service provider, they will coordinate their allocation and access in order to maximize their social welfare. This is called *cooperative* spectrum sharing. Most centralized spectrum allocation can be considered as cooperative. On the other hand, it is not always the case that all secondary users belong to the same service provider, such as those who access the open spectrum band. Different users have different objectives, and hence they only aim at maximizing their own benefit from using the spectrum resources. Since users are no longer cooperative in achieving the same objective, this kind of spectrum sharing is a *noncooperative* one, and secondary users are selfish in that they pursue their own benefit.

In order to give the readers more insight on how to design efficient spectrum allocation and sharing schemes, we next discuss several important issues in dynamic spectrum allocation and sharing.

### A. Medium Access Control in Cognitive Radio Networks

Medium access control (MAC) refers to the policy that controls how a secondary user should access a licensed spectrum band. Various medium access control protocol have been proposed in wireless networking such as carrier sense multiple access and slotted ALOHA. Due to the new features of CR networks, such as the collision avoidance with a primary user and dynamics in spectrum availability, new medium access protocols need to be designed to address the new challenges in CR networks.

A cognitive medium access protocol with stochastic modelling is proposed in [86], which enhances the coexistence of CR with WLAN systems based on sensing and prediction. A primary-prioritized Markov approach for dynamic spectrum access is proposed in [87], which models the interactions between the primary users and the secondary users as continuous-time Markov chains. By designing appropriate access probabilities for the secondary users, a good tradeoff can be achieved between spectrum efficiency and fairness. A cognitive MAC (C-MAC) protocol for distributed multichannel wireless networks is introduced in [88]. Since the C-MAC operates in multiple channels, it is able to deal with the dynamics of channel availability due to primary users’ activity. A stochastic channel selection algorithm based on learning automata is proposed in [89], which dynamically adapts the probability of access one channel in real time. It is shown that the probability of successful transmissions is maximized using the proposed selection algorithm.

<table>
<thead>
<tr>
<th>Classification criterion</th>
<th>Type 1</th>
<th>Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectrum bands that secondary users are using</td>
<td>Open spectrum sharing: access unlicensed spectrum band only</td>
<td>Hierarchical access/licensed spectrum sharing: also access licensed spectrum band</td>
</tr>
<tr>
<td>Access technology of licensed spectrum sharing</td>
<td>Spectrum underlay: secondary users transmit concurrently with primary users subject to interference constraints</td>
<td>Spectrum overlay: secondary users only use the licensed spectrum when primary users are not transmitting</td>
</tr>
<tr>
<td>Network architecture</td>
<td>Centralized: a central entity controls and coordinates the spectrum allocation and access</td>
<td>Distributed: each users makes his/her own decision on the spectrum access strategy</td>
</tr>
<tr>
<td>Access behaviors</td>
<td>Cooperative: all secondary users work towards a common goal</td>
<td>Noncooperative: different users have different objectives</td>
</tr>
</tbody>
</table>

Fig. 6: Classification of spectrum allocation and sharing schemes.
A MultiMAC protocol that can dynamically reconfigure MAC and physical layer properties based on per-node and per-flow statistics is proposed in [90]. Considering the limited capability of spectrum sensing and limited bandwidth, a hardware-constrained cognitive MAC is proposed in [91], which optimizes the spectrum sensing decision by formulating sensing as an optimal stopping problem.

Secondary users also need to be aware of their surrounding environment in allocating and accessing the spectrum. Considering that each node’s spectrum usage is unpredictable and unstable, the work in [92] proposes to integrate interference-aware statistical admission control with stability-oriented spectrum allocation. The nodes’ spectrum demand is regulated to allow efficient statistical multiplexing while the outage is minimized. Since secondary users operating in different frequency bands at different locations are constrained by different interference requirements, a distance-dependent MAC protocol is proposed [93] to optimize the CR network throughput subject to a power mask constraint to protect the primary user. The idea of how to utilize location awareness to facilitate spectrum sharing between secondary and primary users is illustrated in [94]. An aggregation-aware spectrum assignment scheme is proposed in [95] to optimize the spectrum assignment when the available spectrum band is not contiguous. Collision probability and overlapping time are introduced in [96] to evaluate the protection of a primary user. Different spectrum access schemes using different sensing, back-off, and transmission mechanism are considered, which reveal the impact of several important design criteria, such as sensing, packet length distribution, back-off time, packet overhead, and grouping.

B. Spectrum Handoff

When the current channel conditions become worse, or the primary user appears and reclaims his assigned channel, secondary users need to stop transmitting data and find other available channels to resume their transmission. This kind of handoff in CR networks is termed as spectrum handoff [3]. Since the transmissions of secondary users are suspended during a spectrum handoff, they will experience longer packet delay. Therefore, a good spectrum handoff mechanism should provide with secondary users with smooth frequency shift with the least latency.

A good way to alleviate the performance degradation due to long delay is to reserve a certain number of channels for potential spectrum handoff [97]. When secondary users need to switch to another frequency, they can immediately pick one channel from the reserved bands. However, if a secondary user reserves too much bandwidth for spectrum handoff, the throughput may be unnecessarily low, because the primary user may not reclaims his licensed band very frequently. Therefore, there is a tradeoff in optimizing the channel reservation. By optimizing the number of channels reserved for spectrum handoff, the blocking probability can be minimized and the secondary users’ throughput is maximized. A location-assisted handover algorithm is proposed in [98], where the secondary users equipped with the location estimation and sensing devices can report their locations back to the secondary base station. Whenever a handoff becomes a must, secondary users can switch their frequency to one of the candidate channels depending on their locations. A joint spectrum handoff scheduling and routing protocol in multi-hop multi-radio CR networks is proposed in [99], which can minimize the total handoff latency under the constraint on network connectivity. The protocol extends the spectrum handoff of a single link to that of multiple links.

In order to achieve a reliable continuous communication among secondary users in the presence of random reclaims from a primary user, secondary users should select their channels from different licensed bands owned by different primary users [100] [101]. The multi-band spectrum diversity helps to reduce the impact of the appearance of a primary user and improve the reliability of secondary spectrum access.

C. Cognitive Relaying

Utilizing the broadcasting nature of wireless networks, cooperative relaying is proposed in recent years [102] to improve the network performance through spatial and multi-user diversity. Combined with CR technology, cooperative relaying can offer more significant performance gain, because cognitive relay nodes can forward a source node’s data by using the spectrum white space they have detected.

A cognitive multiple access strategy in the presence of a cooperating relay is proposed in [103]. Since the cognitive relay only forwards data when the source is not transmitting, no extra channel resources are allocated for cooperation at the relay, and hence the proposed protocols provide significant performance gains over conventional relaying strategies. A frequency sharing multi-hop CR network is studied in [104]. By recognizing the radio environment in each relay node, the system can autonomously avoid the transmission in an interference area. In [105], an infrastructure-based secondary network architecture is proposed to leverage relay-assisted contiguous OFDM data transmission. A relay nodes which can bridge the source and the destination using its common channels between the two nodes will be selected, and relay selection and spectrum allocation is jointly optimized.

D. Spectrum Sensing and Access

Due to energy and hardware constraints, a secondary user may not be able to sense the entire spectrum space and can only access a limited number of channels from those it has sensed. To optimize spectrum access while considering physical layer spectrum sensing and primary user’s traffic statistics, a decision-theoretic approach based on partially observable Markov decision process (POMDP) is proposed in [106], which can optimize secondary users’ performance, accommodate spectrum sensing error, and protect primary users from harmful interference. A separation principle [107] reveals the optimality of myopic policies for the spectrum sensor design and access strategy, and reduces the complexity of the POMDP formulation by decoupling the design of sensing strategy from the design of the access strategy. An extension of [106] is presented in [108] which incorporates
the secondary user’s residual energy and buffer state in the POMDP formulation for spectrum sensing and access.

Cognitive medium access has also been modeled as a multi-armed bandit problem in [109], and an efficient access strategy is developed that achieves a good balance between exploring the availability of other free bands and exploiting the opportunities that have been identified. Multi-cognitive user scenario is also considered, which is modeled as a game.

E. Power Control in a CR Network

In order to manage the interference among secondary users, or avoid harmful interference to primary users due to secondary spectrum usage, various power control schemes are considered in CR networks to coordinate spectrum sharing.

Power control in opportunistic spectrum access (OSA) is studied in [110], which models the packet transmission from source to destination in OSA as crossing a multi-lane highway. If a secondary user tries to use high transmission power to reach the destination in one hop, it has to wait until the primary user is inactive; on the other hand, it can take more advantage of the spectrum opportunities with lower transmission while relying on the intermediate users on the path to destination. The impact of transmission power on the occurrence of spectrum opportunities is investigated in [110], and it is shown that the optimal transmission power of secondary users decreases monotonically with the traffic load of the primary network. Dynamic programming has been used in designing optimal power and rate control strategy, in order to maximize the long-term average rate for a secondary user [111]. An opportunistic power control strategy is proposed in [112], which enables the cognitive user to maximize its transmission rate while guaranteeing that the outage probability of the primary user is not degraded. A collaborative spectrum sensing scheme that considers signal strength, localization and collaboration in the presence of multiple co-channel primary and secondary transmitters is proposed in [113]. The allowed maximum transmitter power of a secondary user in a given channel is determined, using a distributed database containing co-channel transmitter information including location, error estimates, power, and etc.

Conflicts graph is commonly adopted to describe the interference constraints among users, where the node in a graph represents a user, and an edge between a pair of nodes represents the existence of interference. A systematic framework to produce conflict graphs based on physical interference model is presented in [114], which characterizes the cumulative effect of interference while making it possible to use graph theory to solve spectrum allocation problems under physical interference constraints.

F. Control Channel Management

The majority of the DSA systems use a dedicated global control channel to coordinate the spectrum allocation. However, this assumption is not realistic in opportunistic spectrum access since there may be no permanent channel available for secondary users. A distributed group coordination solution is proposed in [115], where a common control channel is only required locally by the neighbor nodes sharing common channels. A cluster based approach is presented in [116], where a dynamic one-hop cluster is formed by users sharing common channels and the spectrum is managed by cluster heads. A distributed swarm intelligence based control channel assignment scheme is proposed in [117], which selects local common control channels among a local group of secondary users according to the quality of the detected spectrum holes and the choice of the neighboring users.

Potential control channel saturation will degrade the network performance severely. An alternative MAC protocol without requiring a common control channel for multi-hop CR networks is proposed in [118]. By dividing the time into fixed-time intervals and having all users listen to a channel at the beginning of each slot, the proposed protocol ensures that control signals can be exchanged among users.

G. Distributed Spectrum Sharing

In centralized spectrum allocation, a lot of information needs to be exchanged among the central controller and network users to coordinate their spectrum usage, and this results in a large amount of signaling overhead. Therefore, distributed spectrum sharing is preferred where users can make their decision on how to use the spectrum solely based on local information.

A distributed spectrum management scheme is proposed in [119], where nodes take independent actions and share spectrum resource fairly. An adaptive approach to manage spectrum usage in dynamic spectrum access networks is investigated in [120]. Considering the frequency agility and adaptive bandwidth, the concept of time-spectrum block is introduced in [121], and a distributed protocol is developed to solve the spectrum allocation problem which enables each node to dynamically choose the best time-spectrum block based on local information. Based on the adaptive task allocation model in insect colonies, a biologically-inspired spectrum sharing algorithm is introduced in [122]. The proposed algorithm enables secondary users to distributively determine the appropriate channels to use with no spectrum handoff latency due to coordination, and achieves efficient spectrum sharing. A distributed resource-management algorithm that allows network nodes to exchange information and learn the actions of interfering nodes using multi-agent learning approach is proposed in [123].

H. Spectrum Sharing Game

Game theory is a well-developed mathematical tool that studies the intelligent behaviors of rational decision makers in strategic interactions, such as cooperation and competition. In dynamic spectrum sharing, secondary users compete for the limited spectrum resources. If they do not belong to the same network entity, secondary users only aim at maximizing their own benefit from utilizing the spectrum resources. Therefore, their strategies in dynamic spectrum sharing can be well analyzed via game theoretical approaches [124].

A game theoretic modeling is presented in [125] that analyzes the behavior of cognitive users in distributed adaptive channel allocation. Both cooperative and non-cooperative
scenarios are considered, and a no-regret learning approach is proposed. In [126], a repeated game approach for spectrum allocations is proposed, in which the spectrum sharing strategy could be enforced using the Nash Equilibrium of dynamic games. Mechanism design is proposed in [127] [128] to suppress the cheating behavior of secondary users in open spectrum sharing by introducing a transfer function to user’s utility. Spectrum pricing problem with analysis of the market equilibrium is studied in [129] and [130]. Correlated equilibrium concept is used in [131] that can achieve better spectrum sharing performance than non-cooperative Nash equilibrium in terms of spectrum utilization efficiency and fairness. A game-theoretic overview for dynamic spectrum sharing is provided in [132].

Auction mechanisms for spectrum sharing have also been proposed in [133], where the utility of each user is defined as a function of the received signal-to-noise-and-interference ratio (SINR). Considering the potential price of anarchy due to the non-cooperative nature of selfish users, the spectrum manager charges each user a unit price for their received SINR or power, so that the auction mechanism achieves the maximum social utility as well as maximal individual utility. A real-time spectrum auction framework is proposed in [134] to assign spectrum packages to proper wireless users under interference constraints. In [135] [136], a belief-assisted double auction mechanism is proposed to achieve efficient dynamic spectrum allocation, with collision-resistant strategies that combat possible user collusive behavior using optimal reserve prices. A scalable multi-winner spectrum auction scheme is proposed in [137] that awards one spectrum band to multiple secondary users with negligible mutual interference. Effective mechanisms to suppress dishonest/collusive behaviors are also considered, in case secondary users distort their valuations about spectrum resources and interference relationships. Other truthful and efficient spectrum auction mechanisms have been studied in [138] and [139].

I. Routing in a CR Network

In traditional wireless networks, all network nodes will be provided with a certain fixed spectrum band for use. For instance, WLAN uses 2.4 and 5 GHz bands, and GSM uses 900 and 1800 MHz bands. In DSA networks, however, there may be no such pre-allocated spectrum that can be used by every node at any time, and the frequency spectrum that can be used for communication may vary from node to node. This new feature of DSA network imposes even greater challenges on wireless networking, especially on routing. If two neighboring nodes do not have a common channel, or they have common channels but do not tune to the same frequency, then multi-hop communication will be infeasible. Thus, new routing algorithms are needed to accommodate the spectrum dynamics and ensure satisfying network performance such as high network capacity and throughput, short latency and low packet loss.

Due to the heterogeneity of spectrum availability among nodes, routing problem can not be well solved without considering the spectrum allocation. In [140], the inter-dependence between route selection and spectrum management is studied, where the network layer selects the packet route as well as decides a time schedule of a conflict-free channel usage. In [141], the topology formation and routing in DSA networks is studied. DSA network nodes first identify spectrum opportunities by detection, and then the detected spectrum opportunities are associated to the radio interfaces of each node. A layered graph model is proposed to help assign the spectrum opportunities to the radio interfaces.

A MAC-layer configuration algorithm is proposed [142], which enables nodes to dynamically discover the global network topology and node location, and identify common channels for communication. New routing metrics are introduced [142] [143], such as the number of channel switches along a path, frequency of channel switches on a link, and switching delay [142].

Spectrum-aware on-demand routing protocol is proposed in [144], which select routes according to the switching delay between channels and the backoff delay within a channel. A probabilistic path selection approach is proposed for a multi-channel CR network [145]. The source node first computes the route that has the highest probability to satisfying a required demand, and then verifies whether the capacity of the potential path indeed meets the demand. If not, extra channels are judiciously added to the links of the route until the augmented route satisfies the demand at a confidence level.

Considering the channel switching constraints due to primary users’ activity, [146] proposes analytical models for channel assignment in a general multi-hop CR network, studies the impact of the constraints on network performance, and investigates the connectivity and transport capacity of the network.

J. Cooperation Stimulation and Enforcement

In the last decade there has been tremendous developments in cooperative communications. A thorough and detailed discussion about various issues on cooperative communications and networking can be found in [147], in which the underlying assumption is that all nodes are unconditionally cooperative. However, this assumption is often not valid because a rational node is considered selfish with a tendency to maximize its own utility. Cooperation is very important when a cognitive user cannot complete his/her task by him/herself alone. Often a cognitive radio network is not centralized but autonomous, and users are either individuals or belong to different authorities and pursue different goals, so cooperation among these selfish users cannot be taken for granted. In [148], cooperation stimulation and enforcement among cognitive nodes on routing a packet in autonomous networks is investigated. Each node keeps a record on the cooperation degree between any other nodes in the network, which represents the difference between the contributions of two nodes to each other. When the cooperation degree is very low, no cooperation could happen among the nodes. By maintaining a proper positive cooperation degree, cooperation among selfish nodes can be enforced. A credit mechanism is further introduced in [149], where the number of packets forwarded by one node for another node
is bounded by a threshold (credit line), and the threshold is adapted to the number of the forwarding requests between the pair of nodes in real time. Cooperation enforcement under noise and imperfect observation is further studied in [150]. Since the nodes need to infer the future actions of others based on their own imperfect observation in a noisy environment, a belief evaluation framework is considered, where Bayes’ rule is used to assign and update each node’s belief values.

K. Security in CR Networks

Due to their new characteristics, such as the requirement on the awareness of the surrounding environment and internal state, reasoning and learning from observations and previous experience to recognize environment variations, adaptation to the environment, and coordination with other users/devices for better operation, CR networks face their unique security challenges. In [151], awareness spoofing and its impact on different phases of a cognitive cycle has been studied. Through spoofing, the malicious attackers can cause an erroneously perceived environment, introduce biases to CR decision-making process, and manipulate secondary users’ adaptation. In [46], the authors have investigated the primary user emulation attack, where the cognitive attackers mimic the primary signal to prevent secondary users from accessing the licensed spectrum. Localization-based defense mechanism is proposed, which verifies the source of the detected signals by observing the signal characteristics and estimating its location from the received signal energy. The work in [79] has investigated the spectrum sensing data falsification attack, and proposed a weighted sequential probability ratio test to alleviate the performance degradation due to sensing error. In the proposed approach, individual sensing reports are compared with the final decision. Users whose reports are identical to the final decision will have high reputation values, and their reports will then carry more weight in future decision fusion. Several types of denial of service attacks in CR networks have been discussed in [152], such as spectrum occupancy failures where secondary users are induced to interfere with primary users, policy failures that affect spectrum coordination, location failures, sensor failures, transmitter/receiver failures, compromised cooperative CR, and common control channel attacks. Simple countermeasures are also discussed. How to secure a CR network by understanding identity, earning and using trust for individual devices, and extending the usage of trust to networking has been discussed in [153].

V. COGNITIVE RADIO PLATFORMS AND STANDARDS

Although many works have been proposed to improve the performance of spectrum sensing and dynamic spectrum access and sharing, most of them only focus on the theoretical modeling and analysis and few of them have been verified in a practical system. Therefore, CR platforms need to be developed as a real-world testbed that can verify the theoretical analysis. In this section, we will first review some existing testbed/platforms and their features, followed by a brief discussion about standardization of CR techniques.

Researchers at the University of California, Berkeley have proposed an experimental setup based on the Berkeley Emulation Engine 2 (BEE2) platform [154] to compare different sensing techniques and develop metrics and test cases so as to measure the sensing performance. In specific, a good CR system should provide sufficient protection to the PU, which casts certain requirements on a CR testbed, including the capability to support multiple radios, connect various different front-ends to support different frequency ranges, the capability for physical/link layer adaptation and fast information exchange for sensing and cooperation, and the capability to perform rapid prototyping. The BEE2 can meet these requirements, and support the features for a CR testbed. Using the BEE2 platform, research on spectrum sensing using energy detection and sensing with cooperation is tested by experiments [155], which shows the feasibility and practical performance limits of energy detection under real noise and interference in wireless environments. In [156], the feasibility of cyclostationary feature detection is further investigated.

A distributed genetic algorithm based CR engine is proposed in the center for wireless telecommunications at Virginia Tech [157] [158]. The cognitive engine focuses on how to provide CR capability to the physical and MAC data link layers. The cognitive system monitor enables cross layer cognition and adaptation by classifying the observed channel, matching channel behavior with operational goals, and passing the goals to a wireless system genetic algorithm adaptive controller module to gradually optimize radio operation. Researchers at Rutgers University have constructed an Open Access Research Testbed for Next-Generation Wireless Networks (ORBIT) [159] to perform experimentation on CR research. Based on the architectural foundation [160], high performance CR platform with integrated physical and network layer capabilities [161] is under development using the ORBIT testbed.

IEEE 802.22 [162] is proposed to reuse the fallow TV spectrum without harmful interference to TV incumbents. A CR based PHY and MAC for dynamic spectrum sharing of vacant TV channels is evaluated in [163], which studies spectrum sensing, coexistence of primary and secondary users, spectrum management, reliability and QoS, and their impact on the overall network performance. Dynamic frequency hopping (DFH) is recently proposed in IEEE 802.22 [162], where sensing is performed on the intended next working channels in parallel to data transmission in current working channel and no interruption is required for sensing. Efficient and mutual interference-free spectrum usage can only be achieved if multiple users operating in DFH can coordinate their hopping behavior. In [164], DFH communities are proposed so that neighboring secondary users form cooperating communities and coordinate their hopping patterns in DFH. The analysis in [165] quantifies the idle bandwidth in the current TV band assignments, and the statistical analysis shows that secondary users can operate on the contiguous idle spectrum using OFDM. A feature detector design for TV bands is studied in [166].

The IEEE P1900 [167] is a new standard series focusing on next generation radio and spectrum management. One important focus of the standard is to provide reconfigurable
networks and terminals in a heterogeneous wireless environment, where the multi-homing capable terminals enable users to operate multiple links simultaneously. The architectural building blocks include a network reconfiguration management (NRM) module that provides information about the environment, a terminal reconfiguration management (TRM) module that takes information from the NRM and determines the optimal radio resource usage strategies, a radio enabler of reconfiguration management that acts as a link between the NRM and the TRM.

VI. CONCLUSIONS

Cognitive radio technology has been proposed in recent years as a revolutionary solution towards more efficient utilization of the scarce spectrum resources in an adaptive and intelligent way. By tuning the frequency to the temporarily unused licensed band and adapting operating parameters to environment variations, cognitive radio technology provides future wireless devices with additional bandwidth, reliable broadband communications, and versatility for rapidly growing data applications. In this survey, the fundamental concept about cognitive radio characteristics, functions, network architecture and applications are presented, and then various research topics on cognitive radio networks are discussed. We start with the prerequisite requirement on deploying cognitive radio, i.e., spectrum sensing, and review different types of detection techniques and cooperative spectrum sensing protocols. In addition, recently proposed dynamic spectrum management and sharing schemes are reviewed, such as medium access control, spectrum handoff, power control, routing, and cooperation enforcement.

The reviews provided in this survey article demonstrate the promising future of cognitive radio technology in terms of dynamic spectrum selectivity, high-speed seamless communications, and low deployment cost. Meanwhile, the intrinsic features of the new communication technology impose new challenges in the design of efficient spectrum management and sharing schemes [168]. Researchers are expected to come up with novel solutions to higher spectrum efficiency enlightened by this survey.

REFERENCES


