Part I

Cognitive radio communications and cooperation

1 Introduction to cognitive radios

With the rapid deployment of new wireless devices and applications, the last decade has witnessed a growing demand for wireless radio spectrum. However, the policy of fixed spectrum assignment produces a bottleneck for more efficient spectrum utilization, such that a great portion of the licensed spectrum is severely under-utilized. The inefficient usage of the limited spectrum resources has motivated the regulatory bodies to review their policy and start to seek innovative communication technology that can exploit the wireless spectrum in a more intelligent and flexible way. The concept of cognitive radio was proposed to address the issue of spectrum efficiency and has been receiving increasing attention in recent years, since it equips wireless users with 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 concerning cognitive radios. In this chapter, the fundamentals of cognitive radio technology, including the architecture of a cognitive radio network and its applications, are introduced. The existing works on spectrum sensing are reviewed, and important issues in dynamic spectrum allocation and sharing are discussed in detail. Finally, an overview on implementation of cognitive radio platforms and standards for cognitive radio technology is provided.

1.1 Introduction

The usage of radio spectrum resources and the regulation of radio emissions are coordinated by national regulatory bodies such as 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 1.1 [114]. 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 [116], through the use of *cognitive radio* technology [284].

Cognitive radio is the key enabling technology that enables next-generation (xG) communication networks, also known as dynamic spectrum access (DSA) networks, to utilize the spectrum more efficiently in an opportunistic fashion without interfering

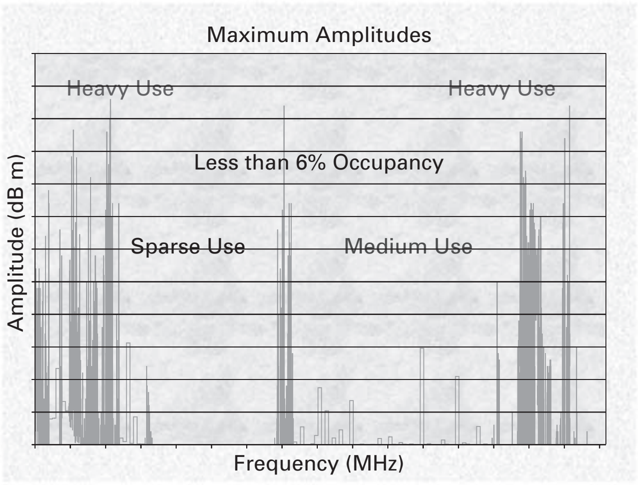

Figure 1.1 Spectrum usage, from FCC Report [7].

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 [114]. It differs from conventional radio devices in that a cognitive radio can equip users with *cognitive capability* and *reconfigurability* [160] [7]. Cognitive capability refers to the ability to sense and gather information from the surrounding environment, such as information about the 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 portions 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 and spectrum-access protocols are no longer applicable. New approaches to spectrum management need to be developed to solve new challenges in research related to cognitive radio, specifically in spectrum sensing and dynamic spectrum sharing.

Since 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, the interference temperature limit should not be violated [65]. If secondary users are allowed to transmit only 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. Owing to noise uncertainty, shadowing, and multipath effects, the detection performance of single-user sensing is pretty limited. Cooperative sensing has been considered effective in improving detection accuracy by taking advantage of the spatial and multiuser diversity. In cooperative spectrum sensing, how to select proper users for sensing, how to fuse an 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 avoiding collision with a primary user. When a primary user reappears 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 relating to cognitive radios in the past few years. In Section 1.2, we overview the fundamentals of cognitive radio technology, including the architecture of a cognitive radio network and its applications. In Section 1.3, we review existing works on spectrum sensing, including interference temperature, different types of detection techniques, and cooperative spectrum sensing. In Section 1.4 we discuss several important issues in dynamic spectrum allocation and sharing. Finally, we present in Section 1.5 several cognitive radio platforms that have been developed in research institutes and industry, and standards on cognitive radio technology.

1.2 Fundamentals

1.2.1 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 [114]. Before CRs adjust their operating mode to environment variations, they must first gain necessary information from the radio environment. This kind of characteristic is referred to as *cognitive capability* [160], 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 have gathered the 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* [160]. For instance, the frequencies of available spectrum bands may keep changing, due to primary users' transmission. Secondary users equipped with CR will know which portion of the spectrum is not occupied by sensing the spectrum, and tune their transmitting frequencies to the spectrum white space.

1.2.2 Cognitive radio functions

A typical duty cycle of CR, as illustrated in Figure 1.2, 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 1.3), i.e., a portion of the 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.

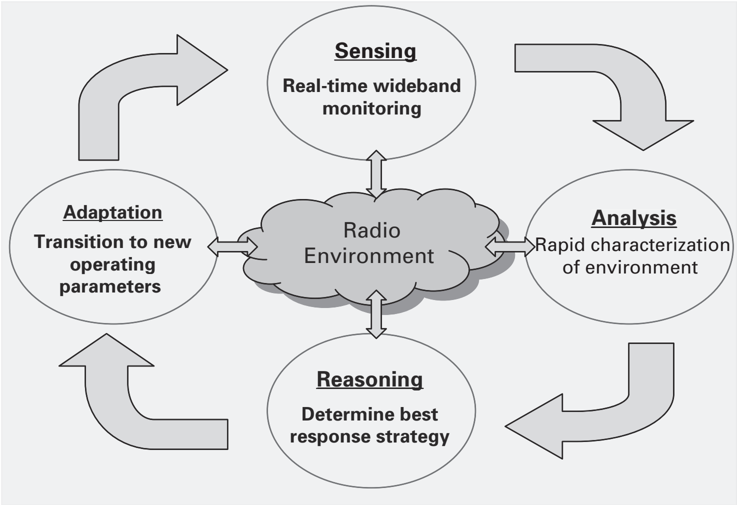

Figure 1.2 The cognitive cycle.

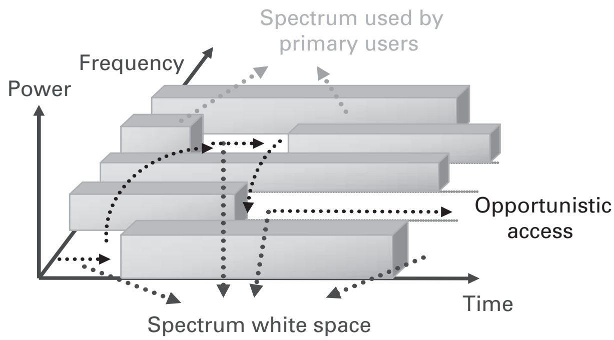

Figure 1.3 Spectrum white space.

After recognizing the spectrum white space by sensing, the 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 [7]. For instance, when a primary user reclaims his/her frequency band, the secondary user 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, etc.

In dynamic spectrum access, a secondary user may share the spectrum resources with primary users, other secondary users, or both. Hence, a good mechanism for spectrum allocation and sharing is critical in order to achieve high spectrum efficiency. Since primary users own the spectrum rights, when secondary users coexist 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 share a frequency band, their access should be coordinated in order to alleviate collisions and interference.

1.2.3 Network architecture and applications

With the development of CR technologies, secondary users who have not been allocated spectrum-usage rights can utilize the temporally unused licensed bands owned by the primary users. Therefore, in a CR network architecture, the components include both a secondary network and a primary network, as shown in Figure 1.4.

A secondary network is a network composed of a set of secondary users and one or more secondary base stations. Secondary users can access the licensed spectrum only 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 a *spectrum broker* [372]. The spectrum broker collects operation information from each secondary network, and allocates the network resources in such a way as to achieve efficient and fair spectrum sharing.

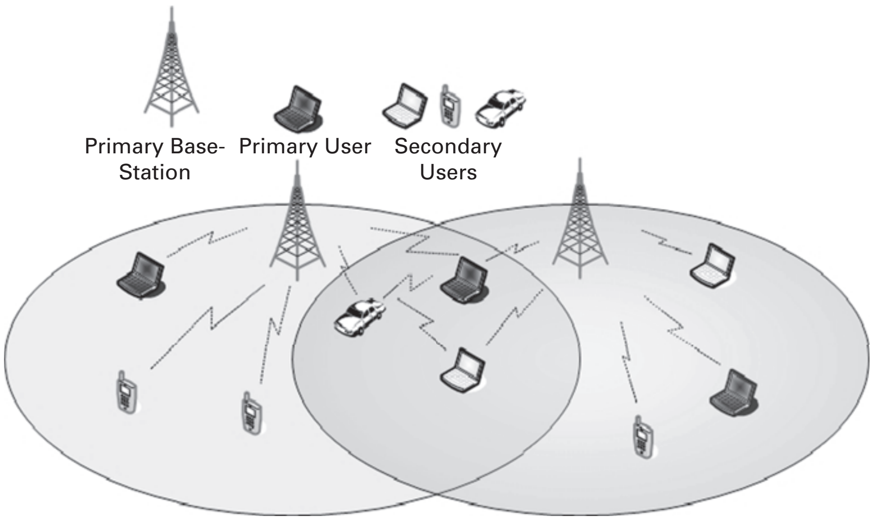

Figure 1.4 Dynamic spectrum sharing.

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 with by secondary networks. Primary users and primary base stations are in general not equipped with CR functions. Therefore, if a secondary network shares 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 [284]. Therefore, CR can provide military users 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 to aid the search and rescue effort becomes indispensable. Since 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, 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, CR constantly scans the entire band to look for and avoid interference from other users; whenever a channel in use by CR is reclaimed by a primary user or interfered with by another secondary user, the CR will instantly select a free channel from its constantly updated free-channel list. The adaptive and dynamic channel switching can help avoid spectrum conflict and expensive redeployment. In addition, since CR can utilize a wide range of frequencies, some of which have excellent propagation characteristics, CR devices are less susceptible to fading related to growing foliage, buildings, terrain, and weather. A CR configuration can also support mobile applications with low cost. 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.

1.3 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 the appearance of spectrum holes. Only with this information can CR adapt its transmitting and receiving parameters, such as transmission power, frequency, modulation schemes, etc., in order to achieve efficient spectrum utilization. Therefore, spectrum sensing and analysis is the first critical step toward 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 spectrum hole detection to determine additional available spectrum resources and compare several detection techniques. Finally, we will discuss cooperative sensing with multiple users or relays' help.

1.3.1 Interference temperature

Secondary users do not have a license for using the spectrum, and can use the licensed spectrum only when they cause no harmful interference to primary users. This requires secondary users to be equipped with CRs, which can detect primary users' appearance and decide which portion of the spectrum is available. Such a decision can be made according to various metrics. The 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 RF emitters, constraining the transmitter power becomes problematic, since unpredictable new sources of interference may appear. To address this issue, the FCC Spectrum Policy Task Force [115] 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 [236]. More specifically, it is defined as the temperature equivalent to the RF power available at a receiving antenna per unit bandwidth [66], i.e.,

$$T_{\mathrm{I}}(f_{\mathrm{c}}, B) = \frac{P_{\mathrm{I}}(f_{\mathrm{c}}, B)}{kB}, \tag{1.1}$$

where $P_{\mathrm{I}}(f_{\mathrm{c}}, B)$ is the average interference power in watts centered at f_{c}, covering bandwidth B measured in hertz, and Boltzmann's constant k is 1.38×10^{-23} J K^{-1}.

With the concept of interference temperature, the 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 will 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 [160]. 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. A subspace-based method to gain knowledge of the quality and usage of a spectrum band has also been proposed [454], in which information about the interference temperature is obtained by eigenvalue decomposition.

Given a particular frequency band in which the interference-temperature limit is not exceeded, that band could be made available for secondary usage. If a regulatory body sets an interference-temperature limit T_{L} for a particular frequency band with bandwidth B, then the secondary transmitter has to keep the average interference below kBT_{L}. Therefore, the interference temperature serves as a cap placed on the potential RF energy that could appear on that band, and there have been some previous studies on how to implement efficient spectrum allocation with the interference-temperature limit.

In [66], two interpretations of the interference-temperature models were analyzed, since there is ambiguity over which signals are considered interference, and which frequency f_c and bandwidth B to use. The first is the *ideal interference-temperature model*, in which interference is limited specifically to primary signals. Assume a secondary transmitter is operating with average power P in a band $[f_c - B/2, f_c + B/2]$, which overlaps n primary signals with frequency f_i and bandwidth B_i. Then, the interference-temperature limit will ensure that

$$T_{\mathrm{I}}(f_i, B_i) + \frac{MP}{kB_i} \le T_{\mathrm{L}}(f_i), \qquad \forall 1 \le i \le n, \tag{1.2}$$

where M represents attenuation due to fading and path loss between the secondary transmitter and the primary receiver. However, it is generally very difficult to distinguish primary signals from secondary signals or measure T_{I} in the presence of a primary signal, unless some a priori information is known about the primary signal. Therefore, a *generalized model* is considered, which requires no a priori knowledge of the RF environment and limits the secondary transmitter's parameters, since the information about the primary receivers is unknown. In the generalized model, the interference-temperature limit is applied to the entire frequency range, i.e.,

$$T_{\mathrm{I}}(f_c, B) + \frac{MP}{kB} \le T_{\mathrm{L}}(f_c). \tag{1.3}$$

With the interference-temperature-limit constraints, secondary users can select the optimal operating frequency, bandwidth, and power to maximize their capacity. Spectrum shaping has been proposed to improve spectrum efficiency [84] in CR networks. More specifically, using interference fitting, a CR senses the shape of the interference power spectrum and creates spectra inversely shaped with respect to the current interference environment in order to take advantage of gaps between the noise floor and the cap of the interference-temperature limit. Another application of spectrum shaping is to create notched power spectra that allow the usage of noncontiguous spectrum segments and avoid primary signals. Dynamic spectrum access with QoS and interference-temperature constraints has been studied in [472]. The objective of the scheme is to maximize the total throughput of all secondary users in a network, constrained by a minimum QoS requirement and a total-received-power requirement at a specified measurement point. Within the framework of the interference-temperature model, the work in [399] presented cooperative algorithms for selecting the most appropriate channel for transmission in a cognitive mesh network. Each mesh node computes a set of channels available for transmission without violating the interference-temperature limit in its interference range, and then uses a per-hop link-cost metric and an end-to-end routing metric to select channels for each hop on the path.

Traditional interference constraints are usually binary and inefficient, since they consider only pair-wise sets of users. Non-binary constraints in line with the interference-temperature model have been studied in [39], which considered the effects of multiple interference sources from across the network. Under these constraints, simultaneous spectrum assignment for a number of secondary transmitters is achievable to improve spectrum utilization, while ensuring that the primary receivers can maintain