CHAPTER

3

Contents

The IEEE 802.11 Standard

R. Pérez-Jiménez, J. M. Riera, and F. J. López-Hernández

3.1 Introduction to IEEE 802.11: General description

If there were a queen of LANs, it would be the Ethernet. The Ethernet is the common, and commercial, name for the ISO 8802-3, the name of the standard assuming the IEEE 802.3 recommendation. There is an enormous family of LAN standards. Almost all of them deal with just the two lower layers of the OSI architecture, the data link layer (DLL) and the PHY. In fact, some sublayers are defined to simplify the implementation of conformant equipment. In this way, the DLL is divided into the logical link control (LLC) and the medium access control (MAC), while the PHY includes the PHY convergence protocol (PLCP) and the PHY medium dependent (PMD). Other management layers have been included to optimize the layer or sublayer coordination.

The standard describes the functionality and relationships between the layers and sublayers but fails to specify how they are to be made. This is an important point, because the rules set out by the standards only deal with the behavior of the equipment, allowing manufacturers to implement it as they wish. Together with the definition of any standard, the conformance tests are included. This is the touchstone to assure interoperability between devices from different manufacturers.

One of the younger members of the IEEE 802 family is the ISO 802.11 WLAN standard. It only describes the specification of the MAC and PHY layers, so wireless devices can use the same LLC developed for other IEEE 802–compliant systems. Figure 3.1 describes this structure.

The main goal of the 802.11 standard is to achieve full functionality for the upper layers without worrying about the quite significant differences between a network based on a reliable cable and another using the air. This benefit is possible because of the careful design of the MAC and PHY. All the issues on security, link losses, node authentication, fading, and large differences in power have been analyzed and incorporated as MAC duties or PHY services.

An 802.11 network can be as simple as two stations interchanging information, and as complex as several buildings with network services in several rooms and with a connection to a backbone-cabled network. Between these two extremes there are many possibilities.

The weak spot in the 802.11 standard is its complexity. Many features of this standard, needed to establish reliable data communication between several nodes, make the system so complex that the cost of the system is much higher than that of cabled networks. It is important to remember that quality is not inexpensive. Moreover, other standards,

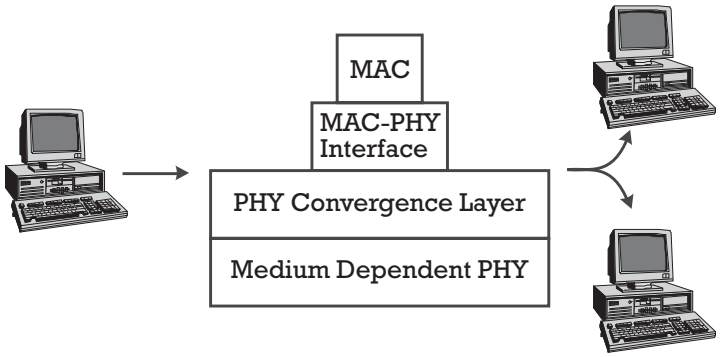

Figure 3.1 Basic layer structure for the IEEE 802.11 standard.

such as IrDA (infrared) or Bluetooth (radio frequency), offer low-cost applications.

The MAC sublayer of a LAN is responsible for correct frame transmission between stations. The IEEE 802.11 standard defines a single MAC protocol for use with all of the PHYs. This supposes that the special characteristics of both RF [1] and IR channels [2–5] are kept in mind. The use of a single MAC protocol better enables chip vendors to achieve high-volume production that will help keep the costs for these systems low. There was considerable debate and compromise before the adoption of the current 802.11 MAC protocol. The MAC protocol defined in the 802.11 is quite complex. The protocol has a few options, as well as several features that can be turned on and off, and combines most of the functionality that was contained in the dozen or so MAC proposals considered by the committee.

The important characteristic of the 802.11 MAC protocol, which is likely to remain unchanged in the final standard, is its ability to support the following.

- The access point (AP)–oriented and ad hoc networking topologies;
- Both asynchronous and time-critical traffic (called time-bounded services in the 802.11);
- Power management.

The structure of an IEEE 802.11 network is described in the MAC section, because the functionality and duty of every station is controlled by this sublayer.

3.2 Medium access control (MAC) for the IEEE 802.11 wireless LANs (WLANs)

This section briefly describes the MAC sublayer of the IEEE 802.11 WLAN standard, summarizing some of the general considerations of the MAC design and discussing the features usually found in a WLAN MAC protocol. Then, the standardized access methods, the distributed coordination function (DCF), with or without handshaking, and the point coordination function (PCF), are described. Finally, a comparison between IEEE 802.11 MAC and HIPERLAN MAC is presented.

3.2.1 Expected features of a WLAN MAC protocol

The 802.11 study group set out some requirements for an appropriate MAC protocol [6–9]. These requirements, which are summarized as follows, may generally be considered as the features expected in any WLAN, not just an IEEE 802.11 LAN.

1. *Throughput:* Since the spectrum is a scarce resource, throughput is definitely one of the most critical considerations in the design of a MAC protocol. The capacity of WLANs should ideally approach that of their wired counterparts. However, due to physical limitations and limited available bandwidth, WLANs are currently targeted to operate at data rates of 1–20 Mbps. The most extended random access protocols belong to the *ALOHA* family, including carrier sense multiple access (CSMA). The ALOHA family suffers from stability problems. That is, the peak throughput is accompanied by a tremendous delay. With Ethernet, and its 10 Mbps of physical transmission and over 80 percent of throughput for CSMA/CD, it is possible to deliver over 8 Mbps of theoretical performance, but, in practice, measurements show that only 3–3.5 Mbps performance is achieved. We should consider not only theoretical throughput but operating throughput (which, practically, is more important). One way of increasing the throughput is by using spread spectrum techniques, which support multiple transmissions simultaneously [10–13].

   Another important consideration about throughput is the impact of unauthorized network access. Neither the MAC nor network management functions can identify any unauthorized access before receiving its transmission, such access inevitably has an impact on network throughput and delay. A successful MAC and network security scheme should reject such unauthorized access and minimize its impact.

2. *Delay:* Delay characteristics are important in every application, but especially in WLANs, since they should serve not only the mandatory asynchronous data service, but also time-bounded multimedia applications such as voice and video. Delay can also cause problems for all data services where the preservation of the sequence of packets is extremely important.

3. *Transparency to different PHY layers:* One of the special requirements for a WLAN MAC is transparency to different physical transmission layers. For IEEE 802.11 LANs, physical transmission layers include direct sequence spread spectrum (DS-SS), frequency-hopped spread spectrum (FH-SS), and diffuse IR. These physical transmission layers are different not only in system design but also in propagation characteristics. However, one MAC must handle all of them. One way of achieving this goal is to have a physical dependent layer, a physical convergence layer, and an appropriate MAC-PHY interface in each station. This architecture is shown in Figure 3.1. Based on architecture currently being adopted by the IEEE 802.11 committee, a single MAC can exchange data transparently with different PHYs via the MAC-PHY interface. Directly related to this item is the limitation of the complexity of the PHY (medium dependent layer, PHY convergence layer, and MAC-PHY interface) to a minimum. The design of a WLAN is an integrated problem, from the PHY up to the network management layer. A MAC design that creates difficulty in other parts/layers of the system is undesirable.

4. *Fairness of access:* The fading characteristics of indoor channels may cause unequal received power at the base station even when power control is enforced. Such a situation may result in unfair access to the network. That is, one mobile node may receive much less power at the base station than another mobile node. When the MAC protocol is operating in the contention mode (necessary for initial registration and often used for uplink traffic), the disadvantaged mobile node may not have a chance to access the channel for a while. A MAC protocol should be able to resolve this situation since it is possible that capture can take place with as small as a 6–9 dB power difference while the dynamic range of fading can be as large as several dBs.

5. *Battery power consumption:* Typically, the 110V (or 220V) electrical supply provided in a building powers device connected to a wired network. Wireless devices, however, are meant to be portable and/or mobile and are typically battery-powered. Therefore, devices must be designed to be very energy-efficient, resulting in "sleep" modes and low-power displays, enabling users to make

cost versus performance and cost versus capability tradeoffs. Many proposed higher-level protocols require mobile nodes to constantly monitor access points or handshake with base stations for the purpose of synchronization, pointer control, or exchanging state information. Therefore, very limited power should be used for packet transmission. Sleep mode should be possible at the receiver front end. The active receive mode may consume more battery power than transmission mode operation since modern commercial digital communication systems may typically have transmission power of 100 mW but need 100 mA of current to support the digital signal processor operation at the receiver.

6. *The maximum number of nodes and maximum coverage area:* According to market studies, a WLAN may need to support hundreds of nodes. Therefore, a MAC should not limit the maximum number of nodes to maintain a satisfactory performance. This feature does not imply that we have an unlimited coverage area that is limited by the delays. The typical coverage area for WLANs ranges from $10m^2$ to $100m^2$, which introduces less than a 1,000 ns propagation delay. WLANs are likely to operate at more than 10–20M chips per second (c/s) for DS-SS and more than 1M symbols per second for other PHYs. Delay in the 500–1,000 ns range can cause big problems for some MACs: a synchronous CDMA system, for instance. We can summarize this property as the ability to work in a wide range of systems, with a MAC design that can handle the geographical size and number of nodes in the LAN.

7. *Robustness vs. cochannel access and interference:* A big challenge to designing a WLAN MAC is to work successfully in the case of collocated networks, which can cause severe cochannel interference. It is quite likely for two or more WLANs to operate in the same region or in some regions where interference between different LANs may occur. Some protocols cannot function normally in this situation. For instance, consider two WLANs operating in two nearby buildings. For certain parts of these LANs, it may be more difficult to communicate with other parts of their own LAN than to communicate with the other LAN. Serious trouble can result from this situation if the MAC uses token passing. It is possible to mistakenly pass the token to a node in the other network. Generally

speaking, there are two concerns for collocated networks, described as follows:

- Security: Other users may illegally break into the network, causing a security alert. This can be solved by an appropriate authentication procedure for new users.
- Interference from collocated networks: For example, if we apply traditional CSMA protocols in WLANs, interference from another network can cause disastrous hidden terminal problems.

These two concerns should be treated in more depth. Interference in wireless communications can be caused by simultaneous transmissions (i.e., collisions) by two or more sources sharing the same frequency band. Collisions are typically the result of multiple stations waiting for the channel to become idle and then beginning transmission at the same time. Collisions are also caused by the "hidden terminal," problem, where a station, believing the channel to be idle, begins transmission without successfully detecting the presence of a transmission already in progress. Interference is also caused by multipath fading, which is characterized by random amplitude and phase fluctuations at the receiver. The reliability of the communications channel is typically measured by the average BER. For packetized voice, packet loss rates on the order of 10^{-2} are generally acceptable; for uncoded data, a BER of 10^{-5} is regarded as acceptable. Automatic repeat request (ARQ) and forward error correction (FEC) are used to increase reliability.

- In infrastructure LANs, multicell coverage is governed by an AP, which is typically a base station or a repeater. The coverage of each cell should overlap the neighboring cell(s) properly; that is, the overlapping region is intended to be minimized to increase system capacity but also kept to a certain proportion so that seamless service is possible. This joint region between cells introduces extra problems. These problems are summarized as follows (see Figure 3.2):
  - Self-interference: When two APs (such as two repeaters) try to transmit a packet to a node in the joint region simultaneously. This causes interference and the packet is likely to be lost.

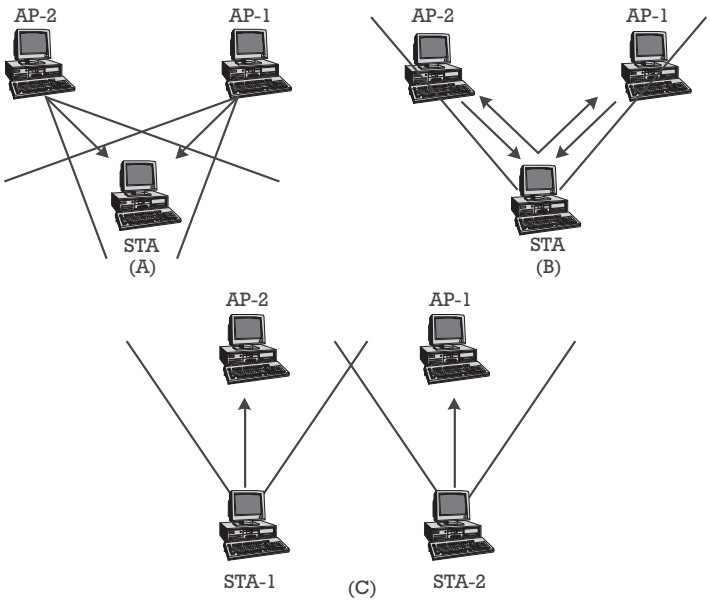

Figure 3.2 Typical cases of interference: (a) self-interference, (b) self-collision, and (c) and up-down collision.

- Self-collision: A node in the joint region transmits a packet that is received by more than one AP. This causes collision or bandwidth waste while routing this packet to its destination(s).
- Up-down collision: A node (A) in one cell is transmitting uplink while another node (B) in another cell is receiving downlink. It is possible that node B may be able to hear (receive) node A's transmission, and this situation results in collision, unless we can perfectly schedule all transmissions. Fortunately, this situation—which is similar to the hidden terminal problem and is a problem as yet unsolved in multicell infrastructure LANs—is very unlikely if the cells are well separated. Since up-down collisions are very destructive, any MAC should take it into account carefully.

The first two problems can be eliminated by careful coordination between APs. However, the up-down collision can only be alleviated; it can never be solved practically. A possible collision window will always exist, although it can be kept to a minimum if

coordination between APs and uplink/downlink takes care of this problem.

The other issue is security. In a wired network, the transmission medium can be physically secured, and access to the network is easily controlled. A wireless network is more difficult to secure, since the transmission medium is open to anyone within the geographical range of the transmitter. Data privacy is usually accomplished in a radio medium using encryption. While encryption of wireless traffic can be achieved, it is usually at the expense of increased cost and decreased performance of the MAC. IEEE 802.11 supports the 802.11 draft standard, which specifies an (optional) data encryption algorithm called the wired equivalency privacy (WEP) algorithm. The WEP algorithm is based on the RC4 PRNG algorithm developed by RSA Data Security, Inc.

8. *Establishing peer-to-peer connectivity without a priori knowledge:* The MAC of a WLAN should support ad hoc networking. Therefore, there should be no requirement for a priori information about network topology (e.g., whether there is communication between all nodes).

9. *The ability to support hand-off/roaming between service areas:* At first, it was thought that a MAC protocol had to support a hand-off function to serve nodes moving from one cell to another. Currently, however, this is not a real limiting consideration because portable computers are not real mobile computers, and users normally work in a fixed place. Nevertheless, a MAC should consider this issue, which is a special feature of WLANs. In indoor environments, due to fast fading, hand-off is not a straightforward problem. For time-bounded services, the ability of a MAC to support hand-off in real time is not an easy task either, especially if we take power consumption into consideration.

10. *The ability to support broadcast (multicast):* Although broadcasting is the natural form of communication for downlink traffic in wireless networks, the MAC should support multicast.

11. *Insensitivity to capture effects:* Although the capture effect can increase throughput, it can also prohibit fair access. One solution is

to enforce insensitivity at the receiver end. A MAC is expected to maintain receiver sensitivity to enhance physical transmission and avoid any potential problems from capture.

12. *Support of priority and non-reciprocal traffic:* In addition to the time-bounded services mentioned earlier, the MAC is expected to support traffic with different priorities. A special feature of WLAN traffic is that the downlink traffic is often much greater than the uplink traffic. A good MAC should definitely support this feature.

Finally, although this is not strictly a MAC concern, another principal consideration should be human safety. Research is currently ongoing to determine whether RF transmissions from radio and cellular phones are linked to human illness. Networks should be designed to minimize the power transmitted by network devices. For IR WLAN systems, optical transmitters must be designed to prevent vision impairment. MAC protocols must be able to work with emission levels low enough to avoid safety complications.

3.2.2 The structure of the IEEE standard MAC protocol

An 802.11 network [14, 15] in general, consists of one or more basic service sets (BSS) that are interconnected with a distribution system (DS). BSS is the fundamental building block of the IEEE 802.11 architecture. A BSS is defined as a group of stations (STAs) that are under the direct control of a single coordination function. This coordination function can be DCF or PCF, both of which are defined below. The geographical area covered by the BSS is known as the basic service area (BSA), which is analogous to a cell in a cellular communications network. Conceptually, all stations in a BSS can communicate directly with all other stations in a BSS. However, transmission medium degradations due to multipath fading, or interference from nearby BSS reusing the same physical-layer characteristics (e.g., frequency and spreading code or hopping pattern) can cause some stations to appear hidden from other stations.

An ad hoc network is the deliberate grouping of stations into a single BSS for the purpose of internetworked communications without the aid of an infrastructure network. Figure 3.3 is an illustration of the components of an IEEE 802.11 and the detail of an independent BSS (IBSS). This is the formal name of an ad hoc network in the IEEE 802.11. Any station can establish a direct communications session with any other

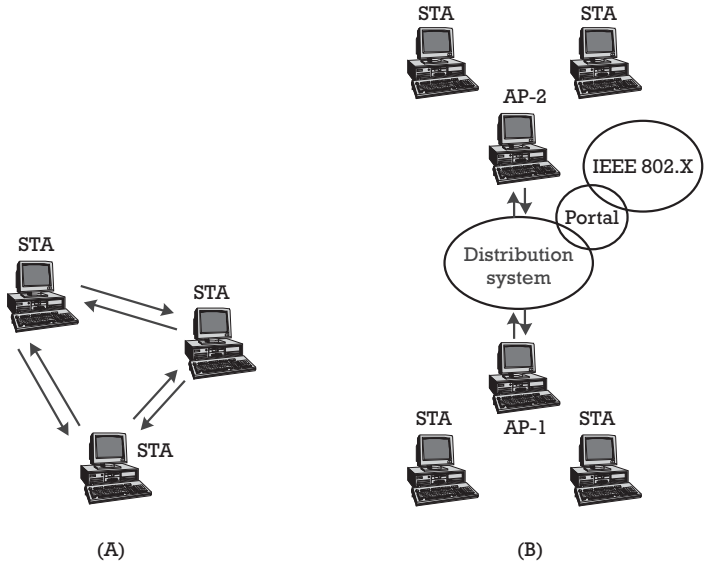

Figure 3.3 Types of connectivity WLAN: (a) Ad hoc, and (b) infrastructured networks.

station in the BSS, without the requirement of channeling all traffic through a centralized AP.

We can summarize the characteristics of an ad hoc architecture as follows:

- Lack of an AP;
- No functionality to support mobility;
- Only support data transfer between stations belonging to the same WLAN.

In contrast to the ad hoc network, infrastructure networks are established to provide wireless users with specific services and range ex- tensions. Infrastructure networks in the context of IEEE 802.11 are established using APs. The AP is analogous to the base station in a cellular communication network. The AP supports range extensions by providing the integration points necessary for network connectivity between multiple BSSs, thus forming an extended service set (ESS). The ESS has the appearance of one large BSS to the LLC sublayer of each STA. The ESS

consists of multiple BSSs that are integrated together using a common DS. The DS can be thought of as a backbone network that is responsible for MAC-level transport of MAC service data units (MSDUs). The DS, as specified by IEEE 802.11, is implementation-independent. Therefore, the DS could be a wired IEEE 802.3 Ethernet LAN, an IEEE 802.3 token bus LAN, an IEEE 802.5 token ring LAN, a FDDI-MAN, or another IEEE 802.11 wireless medium. Note that while the DS could physically be the same transmission medium as the BSS, they are logically different, because the DS is solely used as a transport backbone to transfer packets between different BSSs in the ESS.

An ESS can also provide gateway access for wireless users to a wired network such as the Internet. This is carried out via a device known as a portal. The portal is a logical entity that specifies the integration point in the DS where an IEEE 802.11 network integrates with a non-IEEE 802.11 network. If the network is an IEEE 802.X, the portal incorporates functions that are analogous to a bridge; that is, it provides range extensions and the transfer between different frame formats. Figure 3.3(b) illustrates a simple ESS developed with two BSSs, a DS, and a portal access to a wired LAN.

Coordination functions IEEE 802.11 supports one mandatory and two optional coordination function schemes, which can be summarized as follows:

1. Distributed coordination function (DCF), based on a CSMA with collision avoidance (CSMA/CA) protocol.
2. DCF with handshaking—the request to send (RTS)-clear to send (CTS) procedure—is an optional CF. The use of these two control frames limits the effect of the hidden station problem; some authors [16] call it DFW, as it uses a distribution four-way handshake.
3. Point coordination function (PCF) for distributed time-bounded services (DTBS) in which a point coordinator (or PCF station) has priority control of the medium. That is, when the PCF is active, the PCF station allows only a single station in each cell to have priority access to the medium at any one time.

These CF schemes are described in more detail in the following sections.

Distributed coordination function (DCF) The DCF is the fundamental access method used to support asynchronous data transfer on a best effort basis. As identified in the specification, all stations must support the DCF. The DCF operates solely in the ad hoc network and either operates solely or coexists with the PCF in an infrastructure network. The MAC architecture is depicted in Figure 3.4, where it is shown that the DCF sits directly on top of the PHY and supports contention services. Contention services imply that each station with an MSDU queued for transmission must contend for access to the channel and, once the MSDU is transmitted, must recontend for access to the channel for all subsequent frames. Contention services promote fair access to the channel for all stations.

The DCF is based on CSMA/CA, which is attractive to both vendors and researchers [17–19] due to the popularity of Ethernet. CSMA is a member of the ALOHA family of protocols. ALOHA was the first multiple/random access protocol to be applied to large-scale wireless networks. Pure ALOHA can present unstable behavior when many collisions must be handled. This can result in an unacceptable degradation of the throughput. Improved versions (slotted ALOHA) reduce the possibility of collision duration. CSMA, which senses the status of a channel before transmitting, is the simplest way to improve ALOHA. As shown in many earlier investigations, CSMA demonstrates an increase in throughput in its various versions [20]. Therefore, the IEEE 802.3 committee chose persistent CSMA/CD as the MAC for wired LANs. The success of CSMA/CD in Ethernet relies on the ease of sensing the carrier by measuring the current or voltage in the cable. This is the primary reason why CSMA has been successfully applied in wired networks even though it was originally designed for radio networks. Despite advances in technology, carrier

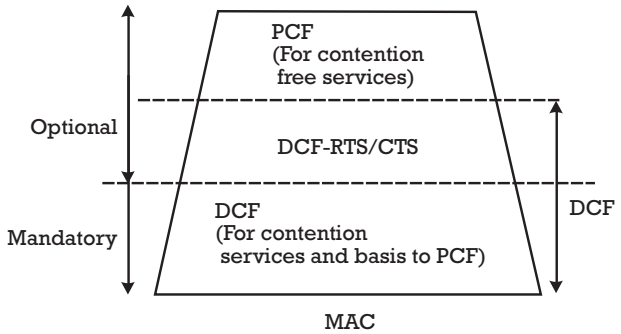

Figure 3.4 MAC architecture for the IEEE 802.11 standard.

sensing is still a major problem for radio networks due to the hidden terminal problem mentioned in Kleinrock and Tohagi's pioneering paper [20]. Reliable carrier (or transmission) sensing is extremely difficult owing to severe channel fading in indoor environments and the use of directional antennas. In IEEE 802.11, carrier sensing is performed at both the air interface, referred to as physical carrier sensing, and at the MAC sublayer, referred to as virtual carrier sensing. Physical carrier sensing detects the presence of other IEEE 802.11 WLAN users by analyzing all detected packets and detects activity in the channel via relative signal strength from other sources. CSMA/CD is not used, because a station is unable to listen to the channel for collisions while transmitting.

In radio systems that depend on the physical sensing of the carrier, a problem called the hidden node problem [8, 21, 22] arises. In this situation, a single receiving station can hear (i.e., is in radio range of) two different transmitters, but the two transmitters cannot hear the carrier signals of one another. In this type of topology, the transmitters send frames without performing a random backoff (because the carrier signal of the other transmitter is never heard). This results in the likelihood of collision. To alleviate the hidden terminal problem and to increase reliability, CSMA/CA is used. In general, it is associated with polling or handshaking, because multipath fading of indoor channels usually lasts for a time equal to a symbol.

CDF with handshaking (RTS-CTS procedure) The 802.11 MAC protocol [14] includes, as an option, a well-known mechanism for solving the hidden node problem. The protocol makes use of two control frames:

- A RTS frame that a potential transmitter sends to a receiver;
- A CTS frame that a receiver sends in response to a transmitter's RTS frame.

The CTS frame gives the requesting station permission to transmit while notifying all stations within radio range not to initiate any transmissions for a given time. This is called the net allocation vector (NAV) in 802.11. NAV indicates the amount of the time that must elapse before the current transmission session is complete and the channel can be sampled again for idle status. The channel is marked busy if either the physical or virtual carrier sensing mechanisms indicates that the channel is busy. Because of the signaling overhead involved, the RTS/CTS feature is not

used for short packets, for which the likelihood of collision and cost (in terms of retransmission time) are both small anyway. If RTS/CTS is not used, the duration field of the data frames actualizes NAV.

K. C. Chen [6] studied the performances of this CSMA-CA with four-way handshake. We assume that the RTS-CTS transmission cycle occupies a short transmission time. The throughput presents a significant increase (10%) when a hidden terminal problem occurs. However, it pays a price in the case where there is no hidden terminal. The throughput is reduced to 63%, a substantial reduction from the original CSMA (about 80% of peak throughput); the cause is, of course, the RTSCTS overhead.

To describe briefly the DCF transmission procedure, it is interesting to see Crow et al. [7] and the standard [14]. A source station performs virtual carrier sensing (NAV actualization) by sending MPDU duration information in the header of RTS, CTS, and/or data frames. An MPDU is a complete data unit that is passed from the MAC sublayer to the physical layer. The MPDU contains information, payload, and a 32-bit CRC. The duration field indicates the amount of time (in microseconds) after the end of the present frame in which the channel will be used to complete successful transmission of the data or management frame.

Priority access to the wireless medium is controlled through the use of the interframe space (IFS), a time interval between the transmission of frames. The IFS intervals are mandatory periods of idleness in the transmission medium. Three IFS intervals are specified in the standard: short IFS (SIFS), point coordination function IFS (PIFS), and DCF IFS (DIFS). The SIFS interval is the smallest IFS, followed by PIFS and DIFS, respectively. Stations only required to wait a SIFS have priority access over those stations required to wait for a PIFS or DIFS before transmitting; therefore, SIFS has the highest-priority access to the communications medium. For the basic access method, when a station senses that the channel is idle, the station transmits an MPDU. The receiving station calculates the checksum and determines whether the packet was received correctly. Upon receipt of a correct packet, the receiving station waits for a SIFS interval and transmits a positive ACK frame to the source station, indicating that the transmission was successful. When the data frame is transmitted, the duration field of the frame is used to let all stations in the BSS know how long the medium will be busy. All stations hearing the data frame adjust their NAV based on the duration field value, which includes the SIFS interval and the ACK following the data frame.

As mentioned previously, since a source station in a BSS cannot hear its own transmissions, when a collision occurs, the source continues transmitting the complete MPDU. If the MPDU is large (e.g., 2,300 octets) a lot of channel bandwidth is wasted due to a corrupt MPDU. To reserve channel bandwidth prior to transmission of an MPDU and to minimize the amount of bandwidth wasted when collisions occur, a station can use RTS and CTS control frames. RTS and CTS control frames are relatively small (RTS is 20 octets, and CTS is 14 octets) when compared to the maximum data frame size (2,346 octets). The source station (after successfully contending for the access to the channel) first transmits the RTS control frames with a data or management frame queued for transmission to a specified destination station. All stations in the BSS hear the RTS packet with a CTS packet after a SIFS idle period has elapsed. Stations hearing the CTS packet look at the duration field and again update their NAV. Upon successful reception of the CTS, the source station is virtually assured that the medium is stable and reserved for successful transmission of the MPDU. Note that stations are capable of updating their NAVs based on the RTS from the source station and CTS from the destination station, which helps to combat the hidden terminal problem. Stations can choose from among the following options:

- Never use RTS/CTS;
- Use RTS/CTS whenever the MSDU exceeds the value of RTS_Threshold (configuration parameter);
- Always use RTS/CTS.

If a collision occurs with an RTS or CTS MPDU, far less bandwidth is wasted when compared to a large data MPDU. However, for a lightly loaded medium, additional delay is imposed by the overhead of the RTS/CTS frames. Figure 3.5 is a timing diagram illustrating the successful transmission of a data frame comparing both cases (using or not the RTS/CTS mechanism).

Large MSDUs handed down from the LLC to the MAC may require fragmentation to increase transmission reliability. To determine whether to perform fragmentation, MPDUs are compared to the manageable parameter Fragmentation_Threshold. If the MPDU size exceeds the value of Fragmentation_Threshold, the MSDU is broken up into multiple fragments. The resulting MPDUs are of size Fragmentation_Threshold, with

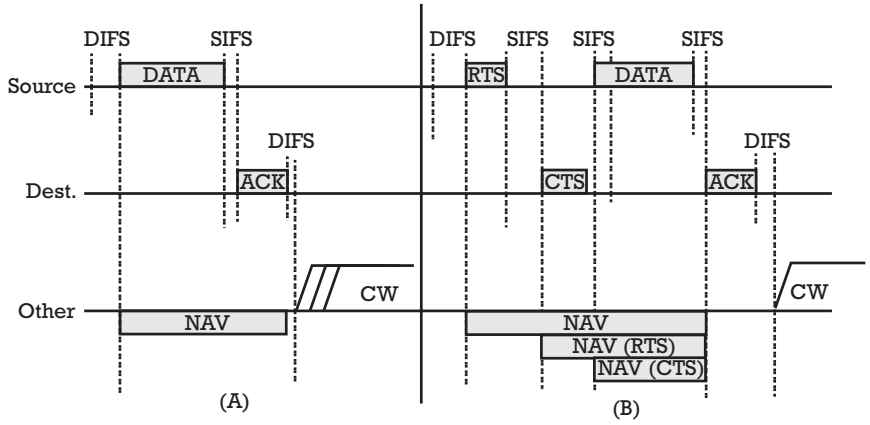

Figure 3.5 The timing diagram of a successful data frame transmission: (a) without handshaking, and (b) using the RTS/CTS mechanism.

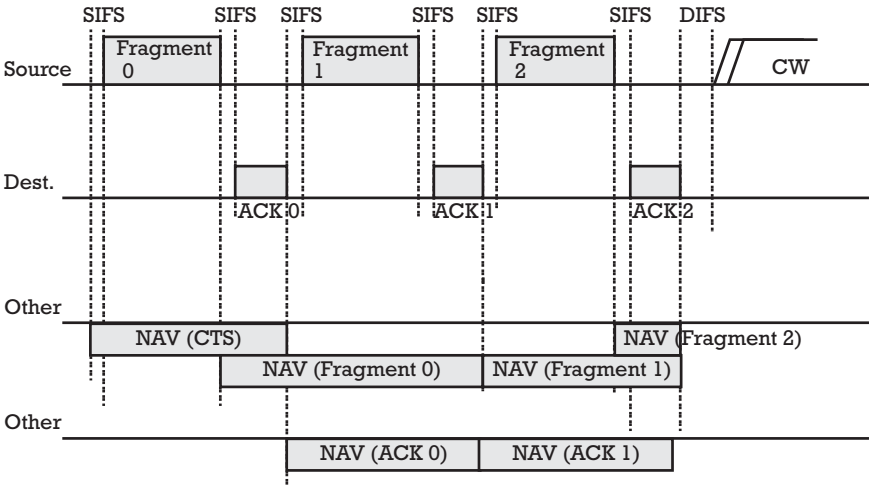

Figure 3.6 The timing diagram of a successful fragmented data frame transmission.

the exception of the last MPDU, which is of variable size not exceeding Fragmentation_Threshold. When an MSDU is fragmented, all fragments are transmitted sequentially (Figure 3.6). The channel is not released until the complete MSDU has been transmitted successfully, or the source station fails to receive an acknowledgment for a transmitted

fragment. The destination station positively acknowledges each successfully received fragment by sending a DCF ACK back to the source station. The source station keeps control over the channel throughout the transmission of the MSDU by waiting only for a SIFS period after receiving an ACK and transmitting the next fragment. When an ACK is not received for a previously transmitted frame, the source station halts transmission and recontends for the channel. Upon gaining access to the channel, the source starts transmitting beginning with the last unacknowledged fragment.

If RTS and CTS are used, only the first fragment is sent using the handshaking mechanism. The duration value of RTS and CTS only accounts for the transmission of the first fragment through the receipt of its ACK. Stations in the BSS thereafter maintain their NAV by extracting the duration information from all subsequent fragments.

Next, the DCF collision avoidance (basic access) procedure is presented. The collision avoidance portion of CSMA/CA is carried out through a random backoff procedure. If the medium is busy, the station defers until after a DIFS is detected and then generates a random backoff period for an additional defer interval before transmitting. Referred as the contention window (CW), this minimizes collisions during contention between multiple stations. The backoff period is the unit of measurement used by the backoff timer. The backoff timer is decreased only when the medium is idle; it is frozen when the medium is busy. After a busy period the decreasing of the backoff timer resumes only after the medium has been free longer than the DIFS. A station initiates a transmission when the backoff timer reaches zero. The backoff interval is chosen following $\text{Int}[2^{2+i}\cdot\text{ranf}()]\cdot\text{Slot\_Time}$, where i is the number of consecutive times a station attempts to send an MPDU, ranf() is a uniform random variable in (0,1), and Int[x] represents the largest integer less than or equal to x. ranf() is a pseudo-random number between 0 and 1, while Slot_Time is a time period. To reduce the probability of collisions, after each unsuccessful transmission the contention window takes the next value in the series until it reaches CW_{max}. The contention window will remain at CW_{max} for the remaining retries.

If a station with a frame to transmit initially senses that the channel is busy, then the station waits until the channel becomes idle for a DIFS period and then computes a random backoff time. For IEEE 802.11, time is slotted in time periods that correspond to a Slot_Time. Unlike slotted ALOHA, where the slot time is equal to the transmission time of one packet, the Slot_Time used in IEEE 802.11 is much smaller than an MPDU and is used to define the IFS intervals and determine the backoff

time for stations in the contention period (CP). The Slot_Time is different for each physical-layer implementation. It corresponds to the sum of clear channel assessment time (time required to determine that the channel is idle), Rx_Tx turnaround time (time required for the modem to change from a receiving to transmitting configuration and vice versa), and air propagation time. The contention window takes an initial value CW_{min} for each frame queued for transmission.

The advantage of this channel access method is that it promotes fairness between stations, but its weakness is that it probably could not support DTBS. Fairness is maintained because each station must recontend for the channel after every transmission of an MSDU. All stations have equal probability of gaining access to the channel after each DIFS interval. Time-bounded services typically support applications that must be maintained with a specified minimum delay, such as packetized voice or video. With DCF, there is no mechanism to guarantee minimum delay to stations supporting time-bounded services [23, 24].

Point coordination function (PCF) The PCF is an optional capability that is connection-oriented and provides contention-free frame transfer. The PCF relies on the point coordinator (PC) to perform polling, enabling polled stations to transmit without contending for the channel. The AP within each BSS performs the function of the point coordinator. Stations within the BSS that are capable of operating in the contention-free period (CFP) are known as CF-aware stations. The method by which the polling tables are maintained and the polling sequence is determined is left to the implementers. The main applications of this capability are time-bounded services (usually packetized voice or video) in which packet order is a main concern.

The PCF is required to coexist with the DCF and logically sits on top of the DCF (see Figure 3.4). The CFP repetition interval (CFP_Rate) is used to determine the frequency with which the PCF occurs. Within a repetition interval, a portion of the time is allotted to contention-free traffic, and the remainder is provided for contention-based traffic. A beacon frame initiates the CFP repetition interval, where the AP transmits the beacon frame. One of its primary functions is synchronization and timing. The duration of the CFP repetition interval is a manageable parameter that is always an integer number of beacon frames. Once the CFP_Rate is established, the duration of the CFP is determined. The maximum size of the CFP is determined by the manageable parameter CFP_Max_Duration. It varies between a minimum (time required to

transmit two maximum-sized MPDUs, including overhead, the initial beacon frame, and a CF-End frame) and a maximum (CFP repetition interval minus the time required to successfully transmit a maximum-sized MPDU during the CP, including time for RTS/CTS handshaking and ACK). Therefore, time must be allotted for at least one MPDU to be transmitted during the CP. It is up to the AP to determine how long to operate the CFP during any given repetition interval. If traffic is very light, the AP may shorten the CFP and provide the remainder of the repetition interval for the DCF. The CFP may also be shortened if DCF traffic from the previous repetition interval carries over into the current interval.

The maximum amount of delay for a frame or fragment on the CFP interval is the time needed to transmit an RTS/CTS handshake, a maximum length MPDU, and ACK. Figure 3.7 shows the CFP repetition interval, illustrating the coexistence of the PCF and DCF. At the nominal beginning of each CFP repetition interval, all stations in the BSS update their NAV to the maximum length of the CFP (i.e., CFP_Max_Duration). During the CFP, the only time stations are permitted to transmit is in response to a poll from the point coordinator or for transmission of an ACK (a SIFS interval after receipt of an MPDU).

Now, the point coordination funciton (PCF) transmission procedure is presented. At the nominal start of the CFP, the point coordinator senses the medium. If the medium remains idle for a PIFS interval, the point coordinator transmits a beacon frame to initiate the CFP. The point

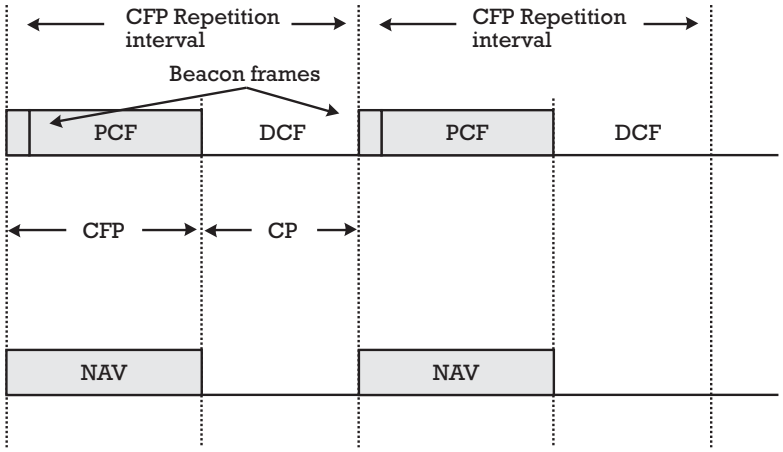

Figure. 3.7 CFP repetition interval.

coordinator starts contention-free (CF) transmission a SIFS interval after the beacon frame is transmitted by sending a CF-Poll (no data), Data, or Data+CF-Poll frame. The point coordinator can immediately terminate the CFP by transmitting a CF-End frame, which is common if the network is lightly loaded and the point coordinator has no traffic buffered. If a CF-aware station receives a CF-Poll (no data) frame from the point coordinator, the STA can respond to the point coordinator after a SIFS idle period, with a CF-ACK (no data) or a Data+CF-ACK frame. If the point coordinator receives a Data+CF- ACK frame from a station, the point coordinator can send a Data+ CF-ACK+CF-Poll frame to a different station, where the CF-ACK portion of the frame is used to acknowledge receipt of the previous data frame. The ability to combine polling and acknowledgment frames with data frames, transmitted between stations and the point coordinator, was designed to improve efficiency. If the point coordinator transmits a CF-Poll (no data) frame and the destination station does not have a data frame to transmit, the station sends a null-function (no data) frame back to the point coordinator. Figure 3.8 illustrates both frame transmission cases, that between the point coordinator and a station and station-to-station. If the point coordinator fails to receive an ACK for a transmitted data frame, the point coordinator waits a PIFS interval and continues transmitting to the next station in the polling list.

After receiving the poll from the point coordinator, the station may choose to transmit a frame to another station in the BSS. When the destination station receives the frame, a DCF ACK is returned to the source station, and the point coordinator waits a PIFS interval following the ACK frame before transmitting any additional frames. The point coordinator may also choose to transmit a frame to a non-CF-aware STA. Upon successful receipt of the frame, the STA would wait a SIFS interval and reply to the point coordinator with a standard contention-period ACK frame.

Fragmentation and reassembly are also accommodated within the Fragmentation_Threshold value used to determine whether MSDUs are fragmented prior to transmission. It is the responsibility of the destination station to reassemble the fragments to form the original MSDU.

3.2.3 Comparison with the MAC protocol of other WLANs: HIPERLAN

This section compares the performances of the MAC protocol with the protocol of its main potential commercial competitor: HIPERLAN. HIPERLAN is extensively described in Chapter 4; here we summarize the most important characteristics of this standard.

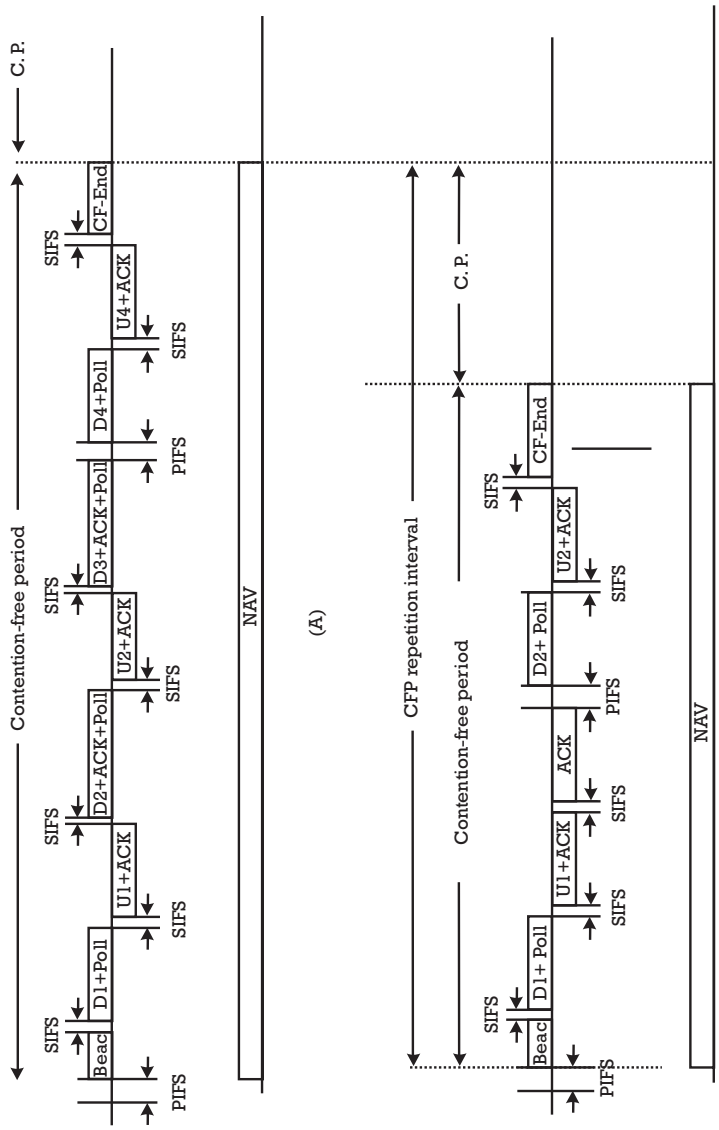

Figure 3.8 The timing diagram of the frame transmission (a) between the point coordinator and a station, and (b) station-to-station in PCF.

The HIPERLAN MAC protocol [21, 25–28] is based on a carrier-sensing mechanism but is quite different in its details from that used in the IEEE 802.3 standard (Ethernet) or the IEEE 802.11 standard. If the medium has been sensed as free for a sufficient length of time, 1,700 bit times in this case, immediate transmission is allowed. If not, the channel access, in the terminology used in the HIPERLAN standard, consists of three phases: prioritization, elimination, and yield. The actions of each node in these three phases are described below and in Figure 3.9.

The prioritization phase aims at allowing only those nodes having packets of the highest available priority to contend further for channel access. This phase consists of a number of slots, with a node having a packet with priority p transmitting a burst[1] in slot $p + 1$ if it has heard no higher-priority burst. At the end of the first burst on the channel, the prioritization phase ends and the elimination phase begins. During the elimination phase, nodes that transmitted a burst during the prioritization phase now contend for the channel. This is achieved by each node transmitting a burst for a geometrically distributed number of slots and then listening to the channel for one time slot. If another burst is heard while listening to the channel, the node stops contending for the channel. Thus, only the node(s) with the longest burst will, in the absence of the hidden node problem, be allowed to further contend for the channel. Immediately after the longest burst and listening period of the elimination phase, the yield phase is started. In this phase, each of the surviving

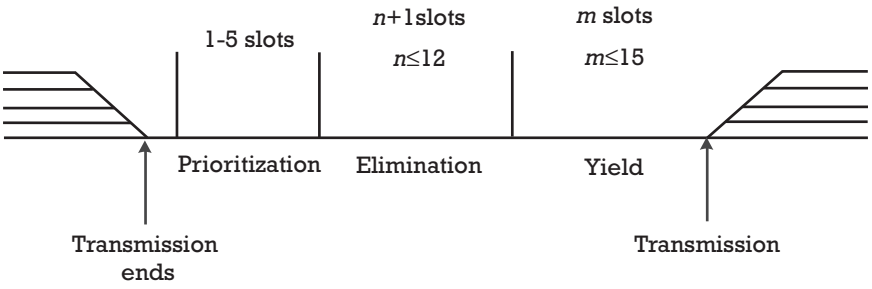

Figure 3.9 Three phases of HIPERLAN channel access.

1. Roughly speaking, a burst entails transmitting the carrier frequency. More precisely, there is a particular bit sequence that is repeated for the duration of a burst, but all receivers respond only to the received signal strength and not to the particular bit sequence.

nodes defers transmission for a geometrically distributed number of slots, while listening to the channel. However, if they hear any transmission, they defer transmission altogether. The purpose of the elimination phase is to bring the number of contenders down to small number, and then the yield phase tries to ensure that only one node eventually transmits. As a result, the chances of actual collisions for data are negligible (less than 3%).

The HIPERLAN technical committee explicitly wanted to support a QoS for packet delivery. QoS support is provided via two mechanisms: the initial value in both cases being assigned by the application using the HIPERLAN services and the priority of a packet (high or normal) and the packet lifetime measured in integral milliseconds with a range of 0-32,767 ms (default value: 500 ms). The residual lifetime of a packet together with its priority are used to determine its channel access priority. Channel access priority can fall into one of five categories and is used for the prioritization phase described above. No other explicit mechanism is used to support the desired QoS, unlike the time-bounded services of the IEEE 802.11 standard.

The committee envisioned that a pure cellular architecture would not be sufficient for the system, hence allowing HIPERLAN nodes to forward packets destined for other nodes. This, of course, requires the maintenance of routing databases at nodes and the dynamic updating of these databases. Methods for this topology maintenance have also been addressed in the standard, for both the databases at each node and broadcasting the information to other nodes. However, it is optional for a node to forward packets; hence, a node can also choose to forego this function, becoming a nonforwarder in the terminology. An interesting discussion of some of the issues involved in this process can be found in [25].

There is support for packet encryption in the HIPERLAN packet transmission mechanism. The standard stays away from defining the particular encryption method used but defines methods to inform the receiver which of a particular set of encryption keys has been used to encrypt the packet. The standard defines a small set of such keys and how they are kept at nodes. It does not, however, define any key distribution strategy, which would be a management function on top of the basic services. Another ETSI committee is working on a security standard for HIPERLAN that will be required for conformance. Table 3.1 summarizes and compares the main characteristics of the HIPERLAN 1 and basic IEEE 802.11 protocols.

Table 3.1
A Comparison of the Main Characteristics of the HIPERLAN 1 and IEEE 802.11 Protocols

	HIPERLAN 1 MAC	IEEE 802.11 MAC
Allows ad hoc networks	Yes	Yes
Supports AP (bridge to wired networks)	Yes	Yes
Power-saving mode	Yes	Yes
Encryption	Yes	Yes
Authentication	No	Yes
Association	No	Yes
QOS	Yes	Nonstandardized

3.2.4 Conclusions

Some conclusions can be obtained on the IEEE 802.11 MAC structure and performance. First, studies on efficiency [7, 23, 24] show that DCF is superior to PCF for short MSDU and low values of BER. On the other hand, if MSDU length increases, PCF performances are superior, as in packetized services (such as voice or video). The 802.11 standard MAC structure is a compromise between several proposals. It means that MAC has to cooperate with different PHY layers and leave decision-making capabilities to implementers. We believe that some optional capabilities (especially handshaking in DCF) will become mandatory in standards.

3.3 Physical layer for IEEE 802.11 wireless LANs: Radio systems

3.3.1 Introduction

There are two specifications for radio systems within the IEEE 802.11 physical-layer definition: the FH-SS physical layer and the DS-SS physical layer. Both use spread spectrum techniques and employ radio transmission in unlicensed spectrum bands at around 2.4 GHz.

Unlicensed frequency bands have increased in interest and importance over the last decade because of the technological advances that have allowed the development of compact and cheap radio transceivers. Many applications (preexistent or new) can be implemented using these transceivers to communicate data of a different kind.

Traditionally, radio spectrum has been considered a scarce natural resource whose use has to be regulated by national administrations. These administrations decide on the users that are granted a license for spectrum use and impose some kind of payment for this use. This scheme works well for traditional users (public operators, broadcasters, and government agencies) but cannot be efficiently applied when the number of potential users of a particular application could number millions. This is not only the case for WLANs but also for other applications such as the ubiquitous garage door opener.

Unlicensed bands, such as the ones known as ISM (industrial, scientific, and medical) bands, can be employed by anyone using equipment that complies with some technical specifications. The regulation applies to the transceivers that have to be designed to meet some characteristics usually to limit the amount of interference with other users. As there is no frequency planning, some degree of interference cannot be avoided. The user can freely employ the equipment. There is no possibility of claiming protection from interference caused by other users. Potential problems must be solved by private agreements.

To reduce the probability of interference, transmission power is strictly limited within these bands, in many cases at levels as low as 10 mW or 100 mW. This limitation guarantees that the potential interference sources are physically in the vicinity of the receiver. This is sufficient for many applications. For example, within industrial buildings, it is very unlikely that external transmitters can interfere with the factory communications systems. The same can be said of systems inside hospitals as well as for most indoor systems. Outdoor systems may suffer from interference from transmitters situated in the vicinity, but many of them, like the aforementioned garage opener, are also protected by the discontinuous nature of the transmission.

In 1985, the FCC decided to augment the transmission level in some unlicensed bands to 1W to boost the development of new applications requiring a greater range. To limit the amount of potential interference, the agency established tight limits in the transmission power spectral density, so that the use of spread spectrum systems was promoted: The greater the band, the greater the transmission power. The new rules, known as Part 15, included regulations for direct sequence (DS) and frequency hopping (FH) systems.

The 802.11 PHY radio systems can be seen, in a sense, as a consequence of this regulation. It is adapted to the FCC specifications for the 2.4-GHz ISM band. As many countries have adopted similar regulations,

the standard can be used worldwide. Some limitations are included in the standard to take into account differences in the regulations of particular countries.

Thus, there are technical and regulatory issues behind the specifications of physical layer spread spectrum systems. Regulatory issues have also limited the available band and bit rate. It is foreseen that new specifications for higher bit rates will be developed within the standard, according to regulations for other unlicensed bands.

Section 3.3.2 discusses the technical characteristics of spread spectrum systems and their potential advantages. In addition, Section 3.3.2 reviews the fundamentals of FH and DS transmission systems and describes the main specifications of the 802.11 PHY, briefly noting the advantages and problems of each of the systems. Finally, a comparison of the two systems is made.

3.3.2 Spread spectrum techniques

Spread spectrum systems use more bandwidth than that needed for transmission. To determine how much bandwidth is needed for a transmission, it is necessary to consider the modulation format. Many people would agree that a system could be considered spread spectrum if the occupied bandwidth is intentionally made greater than that needed, given the bit rate and the modulation format. If the system uses roughly the bandwidth needed then it is not a spread spectrum system and should be considered to be narrowband. (Narrowband will be used in this context as the opposite of spread spectrum.)

To achieve a larger bandwidth, spread spectrum systems use a code in the transmitter, independent of the data, prior to modulation, that must be known by the receiver. A receiver unaware of the code would be unable to decode the transmitted data.

When compared to narrowband transmissions, spread spectrum transmissions are more difficult to detect, intercept, or decode. Thus the main applications of these systems were initially military. However, the same properties have advantages in commercial systems as they are less sensitive to interference from other users and less likely to interfere with others. This is particularly true when the interfering/interfered user uses narrowband transmission. Both kinds of systems can coexist in the same frequency band with little disturbance to each other.

There are no differences between narrowband and spread spectrum systems in noisy environments. Their performance in additive white

Gaussian noise (AWGN) channels is exactly the same; thus, their range is equal.

Spread spectrum systems are in general more complex and thus have been adopted to commercial systems, only when technological advances have allowed the integration of powerful digital signal processors that can be manufactured in large quantities at very low costs.

The most frequently used spread spectrum techniques are FH and DS. As both have been included in the IEEE 802.11 standard, Sections 3.3.3 and 3.3.4 present their main features. More detailed descriptions of spread spectrum systems and their applications can be found in [29–32].

3.3.3 Frequency hopping techniques

FH systems use conventional modulation techniques, but the carrier frequency is changed at a given rate, following a given sequence. This sequence is the code of these systems. A receiver that does not know the code cannot follow the frequency hops and can only occasionally detect some data. If the hopping rate is faster than the bit rate, the system is fast FH (FFH). If the hopping rate is slower than the bit rate, the system is slow FH (SFH). Commercial systems are always SFH because of the complexity of FFH systems.

In an SFH system the bit stream is split into packets, which are each transmitted in a burst with a different carrier frequency. Within a given burst, the transmission is narrowband, using just the bandwidth needed according to the bit rate and the modulation format. Figure 3.10 shows an example of this transmission.

Disregarding for the moment the problem of generating the frequency hops in transmitter and receiver, and synchronizing them, it is obvious that there are no basic differences in performance in noisy environments compared to narrowband systems. In fact, the transmission is narrowband itself. Where are the advantages in FH systems to be found? In an interference environment, there are two, described as follows:

- If there is a narrowband interfering transmitter, the interference affects only those bursts whose carrier coincides with the carrier of the other transmitter. On the other hand, the interference with other narrowband systems is mitigated by the fact that the transmission does not always occupy the same bandwidth. The interference only occurs at times.

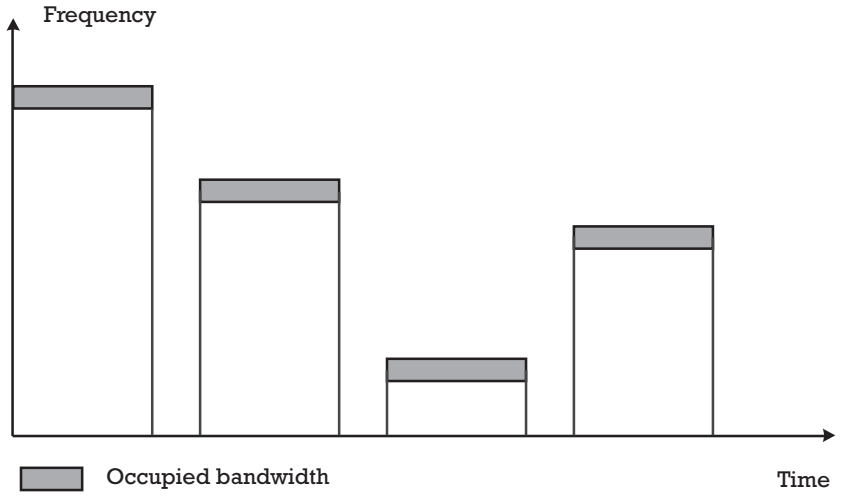

Figure 3.10 FH transmission.

- If two FH systems working on the same band have overlapping coverage areas, interference will occur every time the bursts of both systems coincide in the same carrier.

In addition, there is a third advantage that is very important in multipath environments. When the signal arrives at the receiver through different paths, with different delays, frequency-selective fading occurs: Some frequencies become heavily attenuated because of the destructive combination of the signals coming from different paths. As the resulting phase differences depend on the carrier frequency, for other frequencies the combination is not destructive and can even be constructive, reinforcing the received power. In an FH system some bursts would suffer from fading, but others would be received perfectly. In a multipath propagation environment, only some bursts are affected by frequency-selective fading.

As a consequence, these systems usually assume that some of the bursts are received with very low quality (high BER) or are completely lost because of propagation impairments or because of a powerful interfering signal. However, this affects only a small percentage of the transmission and can be recovered, either by forward error coding or by retransmissions. FH systems are usually considered to work on a pass-fail basis. Some bursts are received perfectly, and others are lost. The system

works as long as the rate of lost bursts is small. This can be achieved in two ways:

- Fight against interference. Usually the higher the number of available frequencies, the smaller the number of lost bursts. This is true if the spectrum is not greatly used.
- However, to limit the effects of selective fading, it is not enough just to use the highest possible number of frequencies. It is also necessary to limit the probability that many of them are simultaneously affected by fades. This leads to the concept of the channel correlation bandwidth, or the amount of frequency displacement for which there is a high probability that fades occur at the same time (in some sense, it is the bandwidth of fades). The FH system will tolerate selective fading with few problems if the system bandwidth is much larger than the channel correlation bandwidth.

The coherence correlation bandwidth is roughly inversely proportional to the rms delay spread, which is a measure of the difference between the time of arrival of the first and the last significant components of the signal. An estimation of the coherence bandwidth B_c makes it equal to $1/(2\pi D)$, where D is the root mean squared (rms) delay spread. Thus, it can be very small in outdoor mountainous areas, with different rays traveling very different distances. However, it is very large in indoor environments, where the distance differences are measured in meters and delay spreads in nanoseconds. In median rooms rms delay spread has been measured to be in the range of 25–50 ns [33]. The coherence bandwidth would be between 3 and 6 MHz. Larger rooms would show larger delay spreads and smaller coherence bandwidths.

It is clear that FH systems are not appropriate when the spectrum is heavily used. It requires that most of the bursts be properly received, free of interference, meaning that most of the channels must be free. However, this is not a problem in unlicensed bands. It is important to keep in mind that as there is neither frequency planning nor user control, unlicensed frequency bands are never used as much as licensed ones, no matter which transmission system is selected.

When frequency use is more important, as in licensed bands, orthogonal FH can be used. This is a technique that allows a group of transmitters to use the same frequency band, all of them with FH, but with their hop sequences tightly synchronized, so that it is impossible for

two transmitters to use the same frequency at the same time. This is used, for example, with the same cell transmitters in the cellular system GSM.

In a scenario of two independent FH transmitters that are working in unlicensed bands, the interference between them will occur when their frequencies coincide. On average $1/N$ of the bursts would be lost, N being the number of available frequencies. Actually, as the hop times would not be synchronized, the number of lost bursts would be between $1/N$ and $2/N$ on average. In any case, it is clear that the greater the number of frequencies, the higher the level of protection against interference. As N increases, the bandwidth also increases. In fact, N can be regarded as the ratio between the bandwidths for the spread spectrum system B_{ss} and the narrowband one B_{nb}.

$$N = \frac{B_{ss}}{B_{nb}} \tag{3.1}$$

Because it is possible to combine the modulation with the hopping sequence, FH systems usually (but not necessarily) employ frequency modulation, normally some variant of FSK with two or four levels. BPSK and QPSK have also been used.

FH transmitters and receivers are technologically complex as they have to rapidly change the carrier frequency. The switching time is lost for transmission and thus has to be made as small as possible. Also, it has to be very small compared to the length of the burst. Currently, commercial PLL frequency synthesizers can change frequency on the order of 100 μs, and thus the length of the bursts can be of only a few milliseconds. Hop rates in commercial systems usually are between 1 and 300 hops per second.

The receivers have to synchronize their frequency hop sequence with the one embedded in the received signal. Unless some kind of centralized control informs all the nodes of the times to hop, the receiver must usually spend a considerable amount of time (one or several cycles of the hopping sequence) to acquire synchronism. There are several techniques for acquiring and keeping the hop synchronism, and its applicability depends on the hop rate and the modulation format. A detailed description of some of these techniques can be found in [29, 31].

3.3.4 Direct sequence systems

In DS systems the modulation rate is intentionally increased to spread the spectrum. This is achieved by combining the bit sequence with a higher-rate binary sequence (called the chip sequence) to obtain a new sequence

with the chip rate. This one is then used to modulate the carrier. Figure 3.11 shows a basic scheme for a DS transmitter.

The inverse operation is performed at the receiver side. The signal is demodulated, and then it is recombined with the same chip sequence to restore the original data. Figure 3.12 shows a much simplified diagram of a DS receiver.

Although these are the basic ideas behind the concept of DS systems, practical implementations require further considerations on the kind of chip sequences used, ways to combine them with the data sequence,

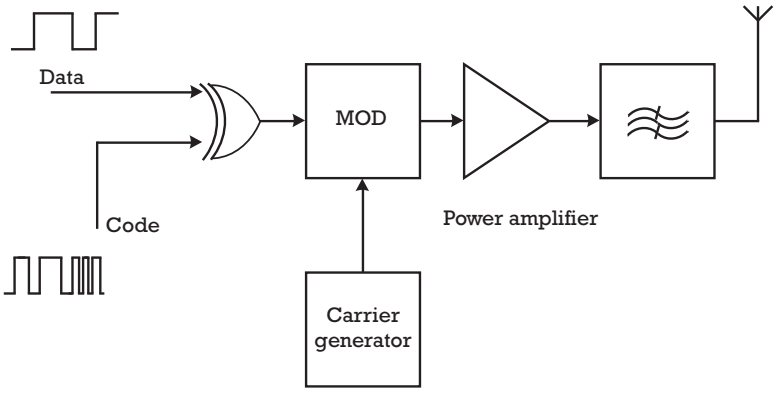

Figure 3.11 Basic direct sequence transmitter.

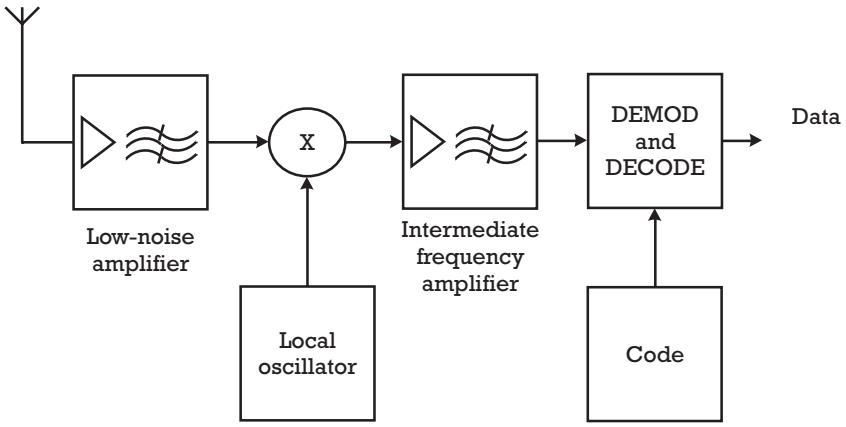

Figure 3.12 Simplified diagram of a direct sequence receiver.

modulation formats employed, techniques to demodulate the signal, synchronization issues, and data recoverers. This section addresses all of these aspects. The advantages of these systems compared to narrowband systems are described as follows:

- The power spectral density, measured in power per unit of bandwidth, is much lower because of the larger bandwidth over which the power is spread. This influences the low probability of interception and the lower capacity of interfering with other systems.
- To recover the data, the receiver must know the chip sequence and carry out the operation of combining the received sequence with the chip sequence. This adds privacy to the communication, because any unwelcome listener would be unable to recover the bit sequence if the chip sequence is kept private.
- In the receiver, the operation to combine the signal with the chip sequence restores the data to their original bit rate, which is much smaller than the chip rate. As a consequence, the signal bandwidth is reduced, and the components lying outside this smaller bandwidth can be filtered out. As any other signals, including noise and interference of any kind, will have a wide bandwidth after their combination with the chip sequence, most of their power will be filtered. Note that narrowband interference would be spread throughout the band after combination with the chip sequence. Thus, DS receivers present a certain degree of noise rejection and interference.
- Several communications can share the same spectrum with direct sequence communications, as long as they use an uncorrelated chip sequence. This is the base of CDMA, a multiple access system in which each user is assigned a spreading code, so he/she is unable to interfere with one another, or mutual interference can be kept at controlled low levels.

Not all of these advantages are used in the 802.11 standard. In particular, it is not a CDMA system. The DS method does not play a role in the privacy features of this standard, as the chip sequence is public and equal for all users. The main reason for selecting this method of transmission is its ability to share the spectrum with other systems at low levels of mutual interference.

Codes in direct sequence systems The selection of codes is one of the most interesting aspects of designing a DS system. Codes are always periodic, but their period can be as short as a symbol period, or as long as several thousands of symbols or even more. Some codes used in the cellular system IS-95 have a period of almost one century. Short codes facilitate the acquisition of synchronism in the receiver, while long codes are needed to guarantee the privacy of the communications. There are two characteristics that are met by all the codes used in DS systems:

- Balanced codes, with roughly the same number of 0s and 1s. These are needed to avoid the presence of a DC component. Pseudorandom sequences with roughly equal probabilities of appearance of both symbols ('0's and '1's) are very often used, particularly in long codes.
- Because the chip rate is always a multiple of the bit rate, in a bit period there is an integer number of chips. Both sequences are synchronized. This reduces the amount of high-frequency energy that would appear if a transition in the bit sequence were followed shortly afterward by a transition in the chip sequence. The ratio of the chip rate (V_c) over the bit rate (V_b) is called the processing gain (G_p); it influences all the system's performance. It is also equal to the ratio of the spread spectrum system bandwidth (B_{ss}) to the bandwidth of a narrowband system (B_{nb}) with the same bit rate.

$$G_p = \frac{V_c}{V_b} = \frac{B_{ss}}{B_{nb}} \tag{3.2}$$

Codes can be classified as orthogonal, quasi-orthogonal, and uncorrelated depending on their behavior in relation to other codes of the same kind. This classification can also be applied to systems, according to the type of codes used.

Orthogonal codes are those whose correlation with codes of the same family is exactly zero. This is a very good property, as it means that different users using these codes can share the same frequency band with zero interference among them. Unfortunately, this property is only met when the codes are combined with the same time reference. Even small displacements between the time origin of codes usually lead to an uncontrolled amount of interference. This means that the system must maintain perfect synchronism at the chip level among all users, which is not easy to achieve. Orthogonal codes are usually employed for the

transmission from a central node to different remote receivers. As the signals are generated in the same transmitter, there is no difficulty in synchronizing them at chip level. One example of orthogonal codes is the Walsh sequences used in the IS-95 cellular system [32].

Quasi-orthogonal codes are those that, while keeping a very small level of correlation when combined in phase, can tolerate displacements to some degree with no significant degradation in their mutual interference. They are the preferred choice when there is a synchronization system that can maintain alignment of the received signals during a few chip intervals. Some subsets of Gold codes have this property and are used in CDMA satellite communications [34].

Uncorrelated codes present low levels of correlation with one another, for any time displacement. They are used when no time synchronization between different users is performed by the system. Different users employ different codes, and the interference with each other is kept small on average, with only occasional peaks that would lead to bit errors. Examples of uncorrelated codes are maximal length codes and Gold codes [31].

Orthogonal and quasi-orthogonal codes are usually short codes, as this facilitates the synchronization of all the signals. They are obtained from tables in the transmitter and receiver. On the contrary, uncorrelated codes are usually long codes. This helps to limit the level of correlation between different codes and to enlarge the time between interference peaks. They are usually obtained from pseudorandom generators, implemented with displacement registers with some linear feedback. This technique is the same used in data scramblers, or in cryptography systems. Different versions of the same code, obtained from a given structure but with different initialization words, are also employed for different users, as they feature low correlation. Gold codes use two pseudorandom sequence generators, whose outputs are combined. The selection of two words for initialization allows for the generation of completely different codes, with low correlation properties but with the same generation structure.

Transmission and reception in direct sequence systems The combination of the chip sequence with the bit sequence is usually made by addition (mod. 2) with an XOR gate. Several modulation formats can be used. One possibility is an I-Q modulation, such as BPSK or QPSK. QPSK has the advantage that it spreads the power further between the two orthogonal carriers with the same frequency (sine and cosine components). This

improves the performance of the system in the presence of narrowband interference signals. BPSK is easier to implement and has a higher immunity to noise. Some systems use BPSK for the data and QPSK for the spreading sequences to combine the advantages of both modulation formats.

DS transmitters are not too different from the basic one shown in Figure 3.12. Practical implementation can vary depending on how much processing is made in digital form and how much is analog processing and whether the modulation is made on the RF frequency or in an intermediate frequency. In this case a frequency converter that transforms the signal to RF and delivers it to the power amplifier follows the modulator. The transmitter structure is quite similar to a digital transmitter working at the chip rate, whose data input signal is the combination of data and chip sequences.

However, DS receivers are not at all digital receivers working at the chip rate, followed by a despread module. In fact, individual chips need not be detected in a DS receiver, and a receiver designed to detect them would present a degraded threshold, because it would have to cope with all the noise and perturbations present in the spread spectrum bandwidth.

A more realistic structure of a DS receiver is shown in Figure 3.13. The signal is first down-converted and pass-band filtered within its RF bandwidth. This is the spread spectrum bandwidth B_{ss}. The filtered signal is then XOR-combined with the code. Subsequently, it is filtered with the narrowband bandwidth B_{nw}. As this is smaller than B_{ss} in the processing gain G_p, noise and perturbations are rejected in this proportion. After this filter, the signal is conceptually identical to a signal received in a narrowband system; it undergoes the usual operations of carrier and clock synchronization and detection.

What is specific to a direct sequence receiver is the need for synchronization of the code with the received signal. This is performed with a local code generator controlled by a local clock. This is in practice a voltage controlled oscillator (VCO). When the code is in phase with the one used in the transmitted signals, higher levels are detected in the correlator output. The power detected at this point is filtered and fed back to control the VCO frequency. The structure is very much like a phase lock loop, but as the controlled variable is the delay between the transmission and reception codes, it is called a delay lock loop (DLL). There are several variants of this scheme. The basic delay lock loop is shown in Figure 3.14. Further reading on delay lock loops can be found in [29, 31]. Strategies for combining the operations of code, carrier, and clock synchronization

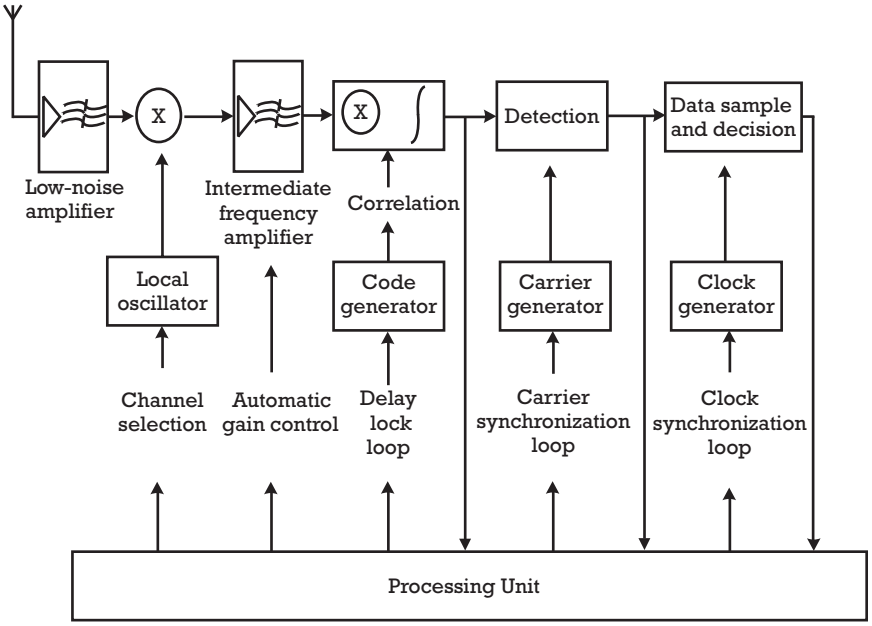

Figure 3.13 Typical structure of a direct sequence receiver.

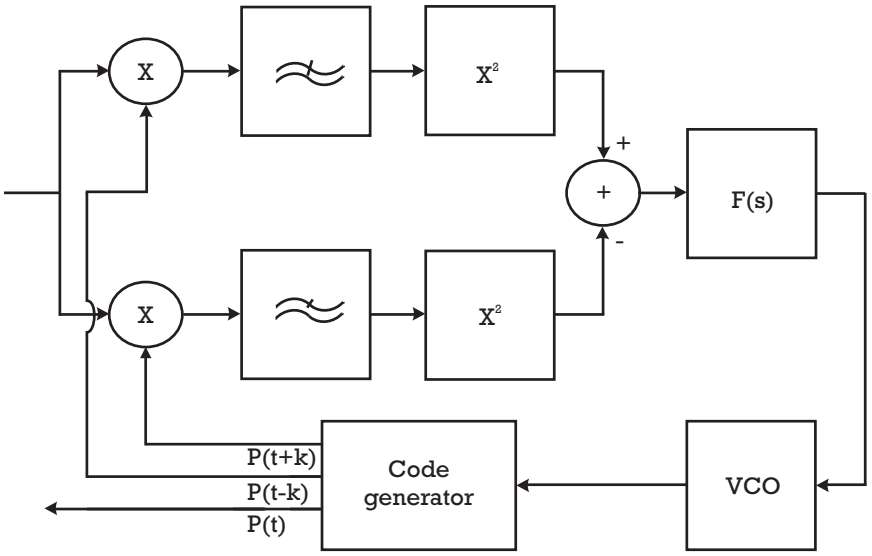

Figure 3.14 Delay lock loop.

as well as the automatic gain control are the keys in designing DS receivers.

Performance in direct sequence systems The process of filtering the signal within the spread spectrum bandwidth, combining it with the code, and further filtering within the narrowband bandwidth affects all the signals present in the system, including the noise, by reducing its power by roughly the process gain. It can be shown that:

$$\left(\frac{S}{N}\right)_{nb} = G_p \times \left(\frac{S}{N}\right)_{ss} \tag{3.3}$$

However, for most of the interferences, I, the $(S/I)_{nb}$ is:

$$\left(\frac{S}{I}\right)_{nb} \approx G_p \times \left(\frac{S}{I}\right)_{ss} \tag{3.4}$$

The approximation becomes equality for interferences having a low degree of correlation with the signal. This includes other users' interference if they use low correlation codes but can be applied to almost any other signal that can be present in the signal bandwidth.

Another interesting property is that by applying the central limit theorem, it can be shown that any interfering signal, and the sum of all of them, has an amplitude distribution that is approximately Gaussian after the correlation process. The combination of all the interfering signals and noise after the correlation is noise-like. This is the optimum perturbation from the information theory [35] and, as a result, the protection against interference is maximum.

The performance of a DS receiver in a particular environment can be easily analyzed, in a first approximation, by obtaining the total perturbation power within the spread spectrum bandwidth as:

$$P_{ss} = \sum I_i + N_{ss} \tag{3.5}$$

and then finding the signal to perturbation ratio in the narrowband bandwidth as:

$$\left(\frac{S}{P}\right)_{nb} \approx G_p \times \left(\frac{S}{P}\right)_{ss} \tag{3.6}$$

Then, because the sum of perturbations after the correlation process is noise-like, the relationship that, for the modulation format, relates the BER to the signal-to-noise ratio in the detector can be applied. Although

this method is only approximate and should be applied carefully, it allows us to remark on the main advantage of DS transmission: It allows extending the threshold as much as the processing gain. For example, a system whose threshold is 6 dB, with a processing gain of 10 dB can work with perturbations 4 dB more powerful than the received signal.

Thus, signal-to-noise ratio can be –4 dB at the receiver input. However, this is not an advantage, as the broader bandwidth implicitly has a higher noise power in the receiver input. For this reason, this DS is transparent with regard to the system range. On the other hand, this is an advantage with regard to interference, as the system can tolerate 10 dB (in this example) more interference power than a narrowband system. Next, the physical layer for PH is presented.

3.3.5 IEEE 802.11 frequency hopping physical layer

Radio transmission Transmission in the FHSS physical layer is defined in two modes: with bit rates of 1 Mbps and 2 Mbps. In both cases, the band between 2.4 and 2.5 GHz is structured in channels. The carrier frequency for channel n can be calculated as:

$$f_n = 2400 + n \quad \text{MHz} \tag{3.7}$$

Because of different regulations in different geographical areas, North America and most of Europe allows the use of 79 channels, but Japan, Spain, and France can use only about 30 channels. The range of frequencies is presented in Table 3.2.

Channels 2 to 80 are thus available in North America and most of Europe. In Japan, channels are from 73 to 95. In France, channels range from 47 to 73, and in Spain they range from 48 to 82.

Table 3.2
Frequency Bands Available in Different Countries

Geographic Region	Band Limits	Channels
North America and Europe	2.402–2.480 GHz	2–80
France	2.448–2.482 GHz	48–82
Spain	2.447–2.473 GHz	47–73
Japan	2.473–2.495 GHz	73–95

FH sequences are defined in the standard [14] in the form of patterns that are permutations of all the frequencies available in the particular geographical area. The number of different patterns is equal to the number of channels, shown in Table 3.1. Every pattern uses all the available channels in a period.

Patterns are described with a parameter x, according to the following relationships:

$$f_x(i) = [b(i) + x] \bmod(79) + 2$$

in North America and Europe,

$$f_x(i) = [(i - 1)^* x] \bmod(23) + 73$$

in Japan,

$$f_x(i) = [b(i) + x] \bmod(27) + 47$$

in Spain, and

$$f_x(i) = [b(i) + x] \bmod(35) + 48$$

in France, where $b(i)$ are the base-hopping sequences, shown in Tables 3.3–3.5.

In North America and Europe, the minimum hop size is 6 MHz; it is 5 MHz in Japan. This is to minimize the correlation between channels used in subsequent bursts.

Three sets are defined for each geographical region. Each set is made up of 26 patterns in North America, 4 patterns in Japan, 9 in Spain, and 11 in France. Within each set, long periods of collisions are avoided.

Table 3.3
Base-Hopping Sequence for North America and Europe

0	23	62	8	43	16	71	47	19	61	76	29	59	22	52	63
26	77	31	2	18	11	36	72	54	69	21	3	37	10	34	66
7	68	75	4	60	27	12	25	14	57	41	74	32	70	9	58
78	45	20	73	64	39	13	33	65	50	56	42	48	15	5	17
6	67	49	40	1	28	55	35	53	24	44	51	38	30	46	

Table 3.4
Base-Hopping Sequence for Spain

13	4	24	18	5	12	3	10	25	19
8	23	15	22	9	21	0	6	14	1
20	7	16	2	11	17	26			

Table 3.5
Base-Hopping Sequence for France

17	5	18	32	23	7	16	4	13
33	26	10	31	20	29	22	12	6
28	14	25	0	8	1	15	3	11
30	24	9	27	19	2	21	34	

Thus, network overlapping on the same area can use patterns of the same set to minimize the interaction.

The modulation employed is Gaussian frequency shift keying (GFSK) for the two data rates. In both cases the symbol rate is 1 Msymbol/s. Two levels of GFSK are employed for 1 Mbps, and four levels of GFSK are used to transmit 2 Mbps. The Gaussian filter bandwidth multiplied by the symbol period gives BT = 0.5 for both speeds. The nominal frequency deviations are shown in Table 3.6.

For 2-GFSK, h2 = 0.32. The nominal frequency deviation from the carrier is 160 KHz. This is the maximum deviation measured after a

Table 3.6
Frequency Deviation for 1 Mbit/s and 2 Mbit/s.

Modulation	**Symbol**	**Carrier Deviation (MHz)**
2-GFSK (1 Mbps)	1	1/2*h2
	0	–1/2*h2
4-GFSK (2 Mbps)	10	3/2*h4
	11	1/2*h4
	01	–1/2*h4
	00	–3/2*h4

certain number of consecutive ones or zeroes to eliminate the effect of the Gaussian filter. For 4-GFSK, h4 = 0.144. Nominal frequency deviation for the external symbols is 216 KHz. These are nominal values. Maximums are defined by national regulations.

Channel switching, defined as the time to settle to within 60 KHz of a new channel nominal frequency, is established at 224 s. For transmission efficiency, a minimum dwell time of 2 ms (maximum hop rate of 500 hop/second) would probably be used in practice. Maximum dwell time is defined by regional regulations. In the FCC it is defined as 400 ms, which gives a minimum hop rate of 1.25 hops/second.

Transmit center frequency accuracy is established at ±60 KHz. All receivers must be able to work with an input signal whose center frequency is in this range from the nominal channel frequency.

Transmitter power must be stabilized within 2 dB of its nominal value for the time of whole frame. The maximum time to switch from on to off and from off to on is 8 μs. The time to switch from reception to transmission is 19 μs.

Reference receiver sensitivity is set at −80 dBm for the 1 Mbps and −75 dBm for the 2 Mbps receiver. The threshold is established at a frame error ratio (FER) of 3% for the MPDUs of 400 octets. The maximum received signal level is −20 dBm for the correct recovery of the data.

The receiver threshold for the clear channel assessment (CCA) is fixed at −85 dBm for the preamble and −65 dBm for the data for an 802.11-compliant signal. This is for transmission power lower than or equal to 100 mW. A device with transmission power greater than 100 mW should have these thresholds decreased in 5*log (P_t/100 mW).

Transmitter power (P_t) depends on national regulations and implementations. The maximum transmitter equivalent radiated isotropic power (EIRP) should exceed 10 mW, except when forbidden by national regulations. Power control is mandatory if the maximum EIRP exceeds 100 mW, with at least two levels: the maximum level and another lower level or a level equal to 100 mW. This is the maximum power allowed in many countries outside North America, where power control is not mandatory. In North America the maximum power is 1W, and if utilized, power control would be mandatory.

FHSS physical layer protocols and functions The FHSS physical layer consists of two protocol functions: One of them, the physical layer convergence function, provides the interface with the MAC and adapts the frame format of the MPDUs to the suitable format for FH transmission. The

second function, a physical medium–dependent function, defines the characteristics of transmission and reception of the data via radio.

This separation is intended to simplify the interface between the MAC and the PHY, and to allow the MAC to operate with minimum dependencies on the transmission medium.

Three functional entities are defined within the FHSS PHY. They are described as follows.

- PMD sublayer: Provides the transmission interface between the transmitting and the receiving stations;
- PHY layer management entity (LME): Provides the management of the local PHY functions in conjunction with the MAC management entity;
- PLCP sublayer: Simplifies the interface of the MAC services with the PHY services.

The reference model for the physical layer is shown in Figure 3.15. Detailed descriptions of the PHY entities, protocols, and exchanged primitives are outside the scope of this text. The interested reader should refer to the standard [14].

The PHY functions can be grouped into two categories, described as follows.

- Those that provide a frame structure for the transmission and reception of the MPDU and associated information. This is done by the PLCP.
- Those that transmit and receive the frames with the physical transmission characteristics defined in the previous section. The PMD

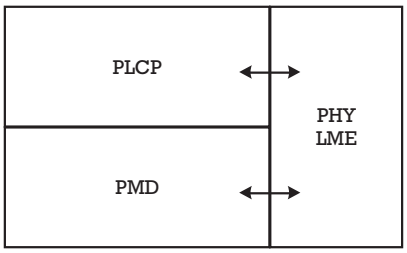

Figure 3.15 Reference model for the PHY.

sublayer does this. The LME helps to synchronize the hopping sequences in all the stations within a given network.

The MPDU length can be between 0 and 4,095 octets. This is a parameter needed both in transmission and reception. The required bit rate is a parameter needed in the transmission side, while the *RSSI* is needed in the reception side.

The PLCP frame format is shown in Figure 3.16. The PLCP preamble and header are transmitted always at the basic speed of 1 Mbps. The PLCP_PDU can be transmitted at 1 Mbps or 2 Mbps.

The fields are described as follows.

- SYNC: This is an alternating zero-one sequence starting with zero and ending with one. It contains 80 bits and allows the receiver to detect a received signal, to synchronize the carrier and the clock, and to select an antenna if frequency diversity is employed.
- Start frame delimiter (SFD): This is a unique word that allows for the frame synchronism. The 16-bit pattern is: 0000 1100 1011 1101.
- PLCP-PDU length word (PLW): This indicates the number of octets contained in the MPDU packet. Valid states are from 0 to 4,095, with 12 bits.
- PLCP-PDU signaling field (PSF): Composed of four bits. The first is reserved for future applications. The three others indicate the bit rate: 000 is for 1 Mbps, and 010 is for 2 Mbps. The other six combinations are for the bit rates of 1.5, 2.5, 3, 3.5, 4, and 4.5 Mbps, but, at this moment, the physical layer specifications do not allow these speeds.
- Header error check (HEC): 16-bit field, obtained with a UITT CRC-16 generator polynomial.

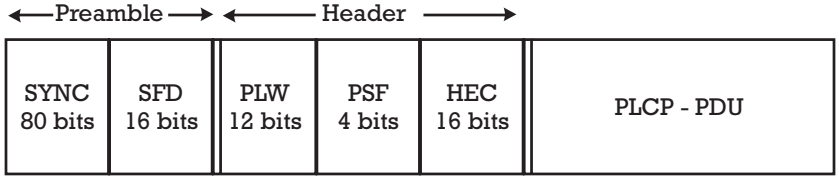

Figure 3.16 PLCP frame format.

- PLCP-PDU: Contains the data in the MPDU, scrambled and unbiased.

Data are scrambled using a synchronous scrambler of 127 bits in length. Its structure is shown in Figure 3.17. The same structure is used for descrambling in the receiver. Both scramblers are initialized at the beginning of the PLCP-PDU with an all-ones word.

Scrambler data are grouped into 32-symbol packets. Stuff symbols (0 for 1 Mbps and 00 for 2 Mbps) are added at the beginning of each packet. To reduce the amount of bias (residual DC component due to imbalance between the number of zeroes and ones) a procedure is established to invert all the bits within one packet when bias is accumulating. Next, the physical layer for DS is presented.

3.3.6 IEEE 802.11 direct sequence physical layer

Radio transmission Transmission in the DSSS physical layer is defined in two modes: 1 Mbps and 2 Mbps. In both cases, the band between 2.4 and 2.5 GHz is structured in channels whose center frequency is shown in Table 3.7 for North America, Europe, and Japan.

The symbol rate is always 1 Ms/s. The bit rate can be 1 Mbps, using BPSK modulation, or 2 Mbps with QPSK modulation. The chip rate is 11 Mchip/s. The spreading sequence is unique for all implementations, the Barker sequence being:

$$+1, -1, +1, +1, -1, +1, +1, +1, -1, -1, -1$$

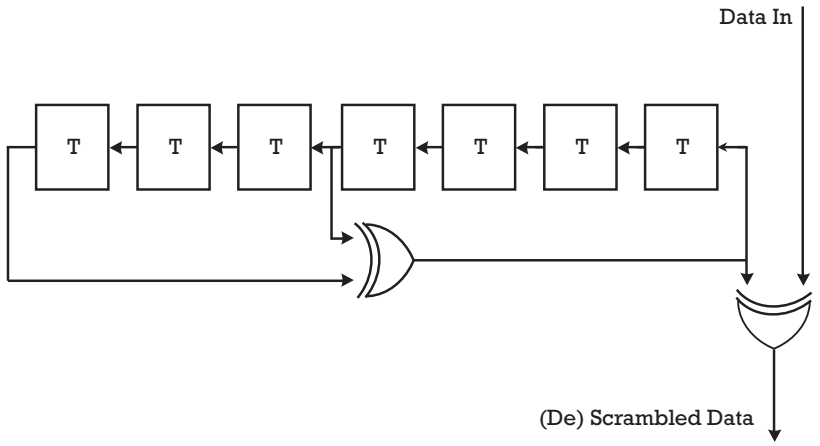

Figure 3.17 Structure of the scrambler/descrambler.

Table 3.7
DSSS PHY Channels

Channel	North America	Europe	Japan
1	2.412 GHz	—	—
2	2.417 GHz	—	—
3	2.422 GHz	2.422 GHz	—
4	2.427 GHz	2.427 GHz	—
5	2.432 GHz	2.432 GHz	—
6	2.437 GHz	2.437 GHz	—
7	2.442 GHz	2.442 GHz	—
8	2.447 GHz	2.447 GHz	—
9	2.452 GHz	2.452 GHz	—
10	2.457 GHz	2.457 GHz	—
11	2.462 GHz	2.462 GHz	—
12	—	—	2.484 GHz

The left-most chip shall be sent first in time and shall be aligned at the start of a transmitted symbol. As the spreading sequence period is equal to the symbol period, implementation is quite easy: For BPSK, the sequence is transmitted without changes when a '1' is sent, and the sequence is inverted to send a '0.' The same can be said for QPSK, since each bit must modulate one of the two quadrature carriers. Synchronization is facilitated by the use of a very short code.

Processing gain is equal to 11 or 10.4 dB. When several independent networks must operate in the same scenario, the processing gain states that the receiver can tolerate 10.4 dB more interference power than the threshold given by the modulation. Also, networks can share a scenario by using different channels separated by at least 30 MHz.

More importantly, the wider bandwidth on the order of 22 MHz helps to improve the reception in the presence of multipath selective fading. A RAKE receiver would be able to distinguish signals arriving with time differences greater than 90 ns, to separate them, and to optimally combine them. No equalizers are needed. The modulation encoder for BPSK or QPSK is shown in Table 3.8.

Transmit-to-receive switching time is fixed at 10 μs. Receive-to-transmit switching time is 5 μs. Transmitter switching from on to off and

Table 3.8
Modulation Encoding

Modulation	Symbol	Phase Change
DBPSK (1 Mbps)	1	0
	0	π
DQPSK (2 Mbps)	00	0
	01	π/2
	11	π
	10	–π/2

from off to on must be performed in less than 2 μs. Frequency tolerance is ±25 ppm for the channel frequency and for the chip rate.

Transmitter power shall be at least 1 mW. Depending on regional regulations, the maximum transmitter power must be lower than 1W in the United States, lower than 100 mW (EIRP) in Europe, and lower than 10 mW/MHz in Japan. Power control is mandatory if the maximum power is higher than 100 mW. At least one of the power levels must be lower than or equal to 100 mW.

The specified receiver sensitivity is –80 dBm for a BER of 8×10^{-2} with an MPDU length of 1,024 bytes. This is for 2-Mbps modulation. The maximum signal is –4 dBm.

There are three modes of operation for the clear channel assessment (CCA). Receivers must be able to work with at least one of them:

- Mode 1: Energy above threshold. The channel will be considered busy if energy above a given threshold is detected. The threshold is –80 dBm for transmission power greater than 100 mW, –76 dBm for transmission power between 50 and 100 mW, and –70 dBm for transmission power lower than 50 mW.
- Mode 2: Carrier sense only. The channel will be reported as busy if a valid DSSS signal is detected, with independence of its energy. If a PLCP header is detected, the channel will be considered busy until the end of the frame as indicated in the PLCP LENGTH field, even if reception is lost.
- Mode 3: Carrier sense with energy above threshold. This is a combination of both. The channel is considered busy if a valid DSSS signal is detected with energy greater than the threshold, calculated as in mode 1.

DSSS physical layer protocols and functions As in the FHSS PHY, the DSSS PHY consists of two protocol functions: the PLCP function and the PMD function. The first adapts the frame format of the MAC to a format suitable for transmission in this medium. The second performs the modulation and demodulation and associated functions.

Three functional entities are defined within the DSSS PHY:

1. PMD sublayer: Provides the transmission interface between the transmitting and the receiving stations;
2. Physical LME: Provides the management of the local PHY functions in conjunction with the MAC management entity;
3. PLCP sublayer: Simplifies the interface of the MAC services with the PHY services.

The PHY functions can be grouped into two categories:

1. Functions that provide a frame structure for the transmission and reception of the MPDUs and the associated information. This is done by the PLCP.
2. Transmission and reception of the frames with the physical transmission characteristics defined in the last paragraphs. This is done by the PMD. The LME helps the MAC in obtaining information on the state of the channel or the synchronization of the signals.

The MPDU length can be between 0 and 4,095 octets. This parameter is exchanged with the MAC in transmission and in reception. The required bit rate is a parameter needed in the transmission side, while the RSSI is needed in the reception side. The PLCP frame format is shown in Figure 3.18.

The PLCP preamble and header are always transmitted at the basic speed of 1 Mbps. The PLCP_PDU can be transmitted at 1 Mbps or 2 Mbps.

The fields are described as follows.

- SYNC: This is composed of a sequence of 128 bits equal to '1's. It is intended to provide all the synchronization needed.
- SFD: This is the 16-bit word F3A0H.

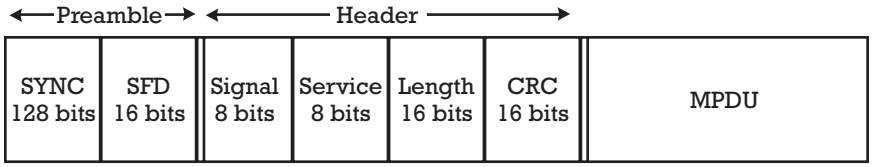

Figure 3.18 PLCP frame format.

- SIGNAL: This 8-bit field provides the receiver with the information on the modulation that will be used. The data rate is equal to the SIGNAL multiplied by 100 Kbps. Currently, only two possible values of data rate are specified. Thus, there are only two possible values of this field: 0AH for 1 Mbps and 14H for the 2 Mbps.
- SERVICE: This field is reserved for future use. The value 00H signifies an 802.11-device compliance.
- LENGTH: This field represents a 16-bit integer number that indicates the number of microseconds needed to transmit the MPDU.
- CRC: A UIT-R CRC 16 polynomial generator is used to provide error protection for the PLCP header.
- MPDU: This field contains all the data in the MPDU, in the same order.

Prior to transmission, all the bits in the PLCP frame are scrambled with a self-synchronized sequence with a length equal to 127 bits. The scrambler structure is shown in Figure 3.19, and the descrambler is shown in Figure 3.20. The scrambler is initialized to any word (except all zeroes) at the beginning of the frame.

3.3.7 Comparison of the FHSS and DSSS physical layers

FH and DS are two methods of spreading the spectrum in transmission systems. Both share some of the advantages of spread spectrum transmission, such as the lower probability of detection and interception, the higher resistance to interference, and the lower levels of interference over other systems. However, they are quite different in their modes of operation.

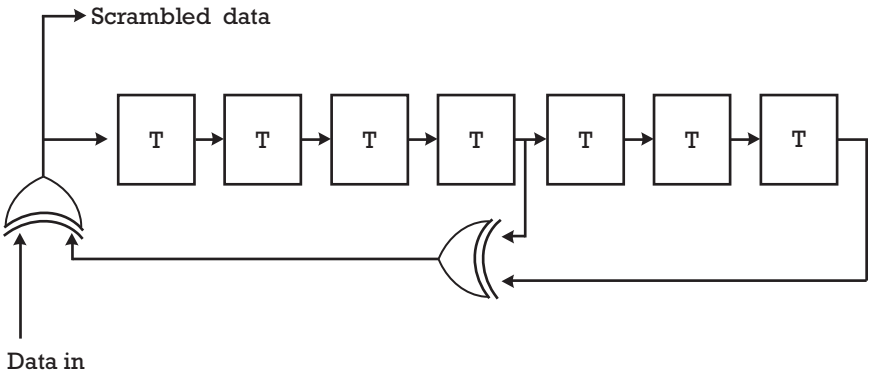

Figure 3.19 DSSS data scrambler.

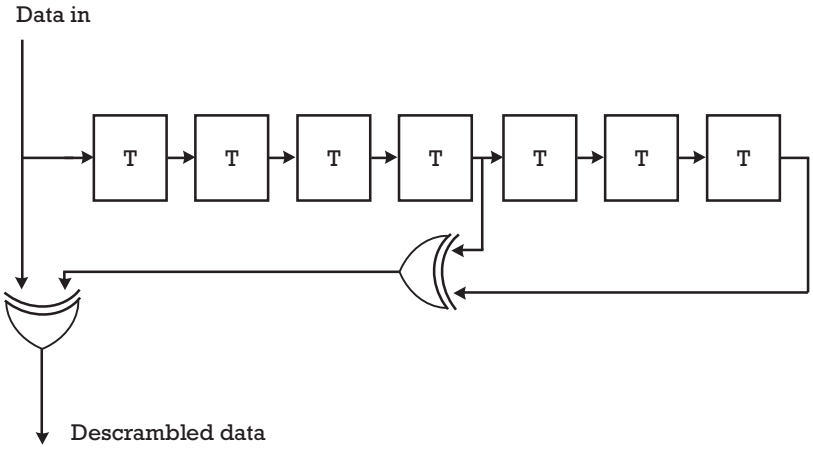

Figure 3.20 DSSS data descrambler.

Many different arguments can be raised in favor of one or the other system. This section compares their applications in different scenarios to highlight their different behavior.

First, let's review the different ways in which they combat channel perturbations arising from interference or from propagation impairments. FH relies on the assumption that perturbations will only affect some of the channels. When transmission is made in the perturbed channels, data will be received with poor or unacceptable quality. However, data transmitted in the unperturbed channels will be received without any difficulty. If the number of perturbed channels is small, by changing

the transmission frequency from time to time most of the data will be delivered without errors. In some systems the lost data can be recovered with the aid of interleaving and forward error coding. As interleaving is not considered in IEEE 802.11, the lost data must be retransmitted. Within a specified channel, the transmission is equivalent to narrowband and has no special protection against perturbations.

DS, on the contrary, provides higher levels of protection within its transmission channels. First, it tolerates interference levels higher than the equivalent narrowband transmission in an amount equal to the processing gain. This is 10.4 dB in this standard. Second, it provides protection against multipath through the wider bandwidth and the possibility of the separation of signals arriving with delays higher than the chip period.

Isolated networks can operate in DSSS or FHSS depending only on the propagation scenario. Small rooms can have multipath with a very small delay spread, and thus the channel coherence bandwidth can be larger than the DSSS bandwidth. In this case, FHSS might be preferable. Larger rooms, such as offices or industrial installations, are good scenarios for DSSS. Delayed paths can be separated in the receiver, and the extra protection against interference can help to provide constantly a very low BER even in the case of occasional interference from other systems. FHSS would suffer from the possibility of fading in some channels and the permanent need for retransmission. In addition—but not as important—transmission time in DSSS is shorter, because there is no need to spend time changing the frequency of the channel.

Networks deployed in scenarios where other equipment (802.11 or other types) operate in the same frequency band present the problem of controlling the amount of interference. If it can be guaranteed that in any receiver the interference level is below its threshold (with the processing gain taken into account), then DSSS is still a good alternative. However, interference levels higher than this threshold would make the receiver deaf to the desired signals, with no possibility of recovering.

Thus, in complex scenarios without the possibility of controlling or even modeling the received interference power from other systems, FHSS must be the choice. Some bursts will be completely lost, but some of the data will be correctly received.

To summarize, when DSSS works, it works well in a wide range of conditions. However, when it fails, as a result of propagation impairments affecting the full channel bandwidth, or because of high levels of interference, it fails dramatically, losing all the ability to receive signals. FHSS behavior is smoother, and more robust in some scenarios.

However, even in the best of conditions it presents the probability of losing some bursts, requiring retransmission.

3.4 Physical layer for IEEE 802.11 wireless LANS: Infrared systems

IR radiation is a good alternative for sending information in closed areas. The IEEE-802.11 standard defines a PHY layer using this technology. Based on the common IEEE-802.11 MAC, the communication is based on a diffuse channel, although a LOS link can be used, with one or more receivers. This is the main difference with IrDA, which is only intended for point-to-point links.

There are three main transmission modes for the emitter-to-receiver wireless optical link: LOS, one or several reflections on the walls and furniture, or by being received and resent by an active repeater. As the power reaching the receiver has a wide dynamic range, no value is set within the standard for the communication range, but in usual rooms or offices, about 10m is expected. Several people with their portable computers or PDS sitting around a common table, or in a not-too-large classroom, are the scenarios taken into account by the standard developers. Other scenarios are well-suited, too—for example, an airport hall or museum, airplane or train cabins, and bank offices.

The network protocol is able to manage several nodes without the necessity of physical connections. If an external connection is needed, an AP should be implemented. Furthermore, a meeting room may have one or more active repeaters installed to improve the signal level anywhere inside it. This is the alternative to offering several RJ or coaxial sockets for the participants.

The reduced range can become advantageous if the data should be confined to the room or if confidentiality is a goal. This is the case in bank offices or meeting rooms.

Due to its intrinsic short range, no regulation has been made on IR frequency allocation for transmission. On the other hand, several eye-safety regulations limit the power emission of IR devices.

3.4.1 Description

The PHY is divided into two sublayers: PLCP and PMD. The PLCP interfaces with MAC to convert MAC packets into the format used by the circuitry of the PMD. The PMD sublayer performs no data processing, so if

new hardware techniques or devices are developed, they can be easily implemented. Both sublayers can be implemented together without a physical frontier between them. For example, one chip can set up the synchronization frame, CRC checksum, PPM encoding, and energy detection. Sections 3.4.2 and 3.4.3 discuss the main characteristics of both sublayers.

3.4.2 The physical layer convergence procedure (IR-PLCP)

This sublayer is where MPDUs are converted to electrical signals to be applied to PMD devices. It has two purposes: to assure an error-free transmission, and to simplify the reception procedure. Inserting, before the MPDU, a preamble and a header makes this. The preamble consists of a long series of pulses to synchronize the receiver clock, to achieve good data extraction. On the other hand, the header includes all the information needed on the receiver PHY.

The basic optical signal is a 250-ns pulse. This period is known as *slot* time. Although two data rates are defined, the pulse length is the same so the optical receiver can be optimized at this pulse duration. Most of the information is coded in PPM format, but other signals, mainly in the PLCP preamble and header cannot be modulated. In any case, the pulse duration is always the same. For example, synchronism pulses are a clock signal of 2 MHz (i.e., a train of many 250-ns pulses) used to synchronize the receiver clock; if the signal were PPM-codified, it could not be decoded because the receiver is out of synchronism at this time.

The frame sent to PMD, or received from it, has three parts: the PLCP preamble, the PLCP header, and the PSDU. This frame is named the PLCPDU.

PLCP preamble This includes some pulses to synchronize the receiver (SYNC) and the SFD. The SYNC part is a clock square signal of 2 MHz that lasts between 57 and 73 slots. The last slot has to be empty (i.e., have no pulse). Its main goal is to allow the receiver clock to synchronize, although it can also be used to estimate the signal-to-noise ratio and automatic gain control, or to choose the receiver if diversity is implemented. It is followed by the SFD, which is always the nibble 1001. This signal marks the beginning of the frame and allows the symbol synchronization. These signals are made of single pulses; they are not modulated in PPM format. In fact, there is no PPM symbol with these sequences of pulses.

PLCP header The header gives the information needed to translate the PPM symbols into bytes. It includes the data rate (DR), the DC level adjustment (DCLA), the length of the data in PSDU (LENGTH), and a checksum of the LENGTH parameter (CRC).

The DR block is three slots long and can take the values listed in Table 3.9. The DR block values are not modulated in PPM.

After DR there is a sequence of pulses to allow the receiver to adjust the DC level of the remaining signals. This is called DC level adjustment (DCLA), and it is needed because the mean value of 4PPM and 16PPM is not the same. The DC level of these signals would be the same as an average PPM signal at the data rate used. Their values are listed in Table 3.10, and only these two values are permitted.

The third block is LENGTH, that is the number of bytes of the MPDU. It is a 16-bit unsigned integer and is the first field to be PPM-modulated. The LSB is transmitted first.

To assure that the previous value is correctly received a checksum of 16 bits is sent. This is calculated using the polynomial $x^{16} + x^{12} + x^5 + 1$. This field is also PPM-modulated.

PSDU This is the data coming from the MAC. Its length is defined by the LENGTH field, can vary between 0 and 2,500 bytes, and is PPM-modulated. The LSB is sent first. The process of transforming bits to PPM symbols is described in Section 3.4.3.

Table 3.9
DR Block Values

Data Rate	Value
1 Mbps	000
2 Mbps	001

Table 3.10
DCLA Values for Both 1 and 2 Mbps Data Rates

Data Rate	Value
1 Mbps	0000 0000 1000 0000 0000 0000 1000 0000
2 Mbps	0010 0010 0010 0010 0010 0010 0010 0010

3.4.3 The IR physical medium sublayer (IR-PMD)

This layer only describes the signals format and the minimum specifications needed for communication between two IEEE 802.11-IR–conformant devices.

Two data rates are defined: 1 and 2 Mbps. Both data rates are implemented for the detectors, but 2 Mbps is optional for emitters. As the emitter power consumption depends heavily on the data rate, the choice is based on battery life for portable devices. The most common implementation includes a speed and rate that are selected depending on the battery charge. Also, at the lower speed the receiver sensitivity is larger so that one conversation can be started at one speed, and if the channel characteristics change, the other one can be used.

Characteristics of the signal used The electrical and optical signals are based on a 250-ns pulse. This defines the slot time both for modulated and unmodulated signals. By using just one fixed-length pulse data, the receiver and emitter can be optimized for this signal, independent of the data rate. Two PPM schemes (see the next section) are used for sending the data, one for each rate. Nevertheless, other signals have to be sent unmodulated, at the most obvious data rate. (The data cannot be extracted until the PPM format is established.) A good synchronization is required for PPM data recovering, so a long clock signal is sent at the beginning of each frame.

Pulse position modulation (PPM) This is the best modulation method for low- and medium-speed optical signals. In this method, a large pulse is sent. Its amplitude and shape are not important. The important factor is its delay relative to a symbol clock.

The maximum current that can be applied to an LED has two values: maximum DC current and maximum pulsed current. The first value is due to the thermal dissipation capacity of the device: the larger the current, the larger the heat generated and the device temperature. On the other hand, short and long current pulses can damage the device because of current channeling, and other processes. In any case, the maximum pulsed current is usually 5 to 10 times larger than DC maximum current. Of course, the LED should be switched off for a period long enough to lose the heat generated while in the "on" state. PPM offers a duty cycle low enough to surpass the DC maximum current. Using, for example, a 16 PPM, the mean current will be about one-sixteenth of the peak current. So a 100 mA LED can be driven with 1A pulses, without degrading the device. On the receiver's photodiode, the signal will be almost 10

times larger. Unfortunately, the bandwidth needed will be larger too, so the benefits are not so important on the signal-to-noise ratio.

There is another factor to be taken into account: the electrical power needed to turn on the LED. Usually wireless communications are intended to be used on battery-powered equipment. If large currents are switched on and off in short pulses, the battery life and the charge duration will be reduced, and the electromagnetic interference (EMI) propagated through power supply connections could be a problem. To maintain an almost constant discharge regime, the LED driver charges a large capacitor while in the off state; this capacitor supplies the large current pulses while in the "on" state. With this technique, the pulse shape cannot be controlled too tightly, but this is not important in a PPM scheme. Figure 3.21 presents the electrical spectrum of a PPM signal.

In Figure 3.21, the electrical bandwidth is much larger than the data rate, and this modulation cannot be used if the bandwidth is a scarce resource, as in RF. Fortunately, the optical spectrum is unregulated because the optical signals are blocked by walls, so a large amount of interference is impossible. As more and more IR equipment is used, some interference will be unavoidable, so three options are feasible for the forthcoming developments. The first one, now used by IrDA, is to establish the communication based on LOS links; the second one is to use another wavelength range; and the third is to use a carrier modulation over 20 MHz. The second option is handicapped by the high cost of optoelectronic devices for longer wavelengths; as the short history of electronic technology teaches us, the prices will fall as the number of devices increases. Accordingly, this problem will diminish if the new applications are successful. Going to a larger wavelength has a second advantage: The optical power level can be increased because the sensitivity and transparency of the human eye is lower. As a result, they are safer than shorter wavelengths. The carrier-modulated option will take advantage of all the circuitry developed for RF, but it will need faster, and again more expensive, emitters and detectors. In conclusion, IEEE 802.11-IR has taken the most convenient position. The problems will exist for future IR standards. With this in mind, PPM is the best option.

PPM modulation is described as follows. Let T_0 be the bit period (i.e., the inverse of the data rate). The time needed to send k bits will be $k \cdot T_0$ seconds. If we divide this time in 2^k slices (slots), every one will represent the value of every k-bit combination. The duration of one slot is $T_s = k \cdot T_0 / 2^k$ seconds, so no more than a few bits can be grouped before the slot time becomes too short.

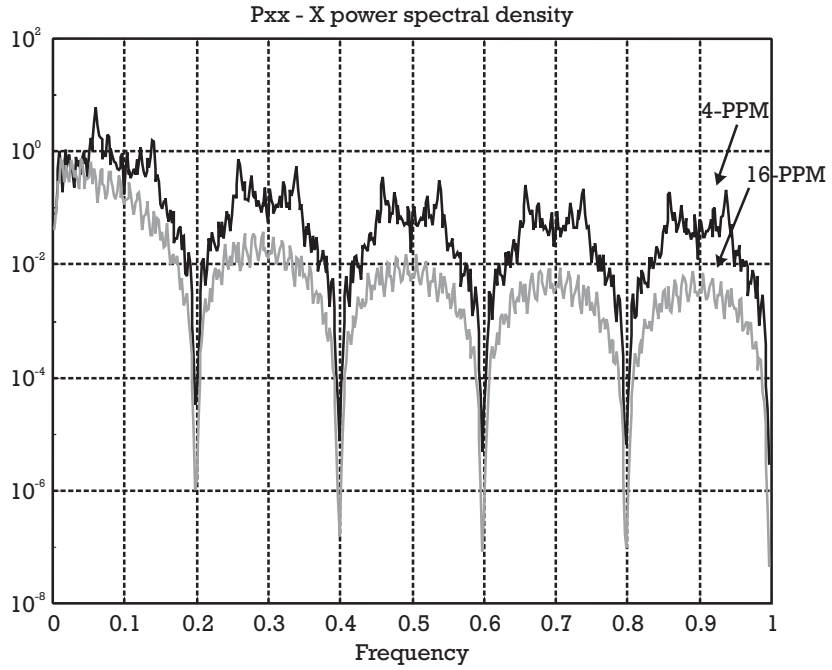

Figure 3.21 Electrical spectrum of a PPM signal.

A single pulse placed into a time slot represents the value of the k bits. Because of this, synchronization is very important: An error in the delay measurement will produce a different bit combination.

IEEE 802.11-IR uses 4-PPM and 16-PPM modulation for 2- and 1-Mbps transmission. The number indicates the number of slots. So one symbol (pulse) defines two bits at 2-Mbps, and four bits at 1-Mbps.

With the synchronization being so critical, the pulse position is not just the plain bit group value. Instead, a Gray coding scheme is used so that one slot timing error forces one bit error.

With these figures, it is easily seen that T_s is 250 ns in both cases. What is different is the duty cycle: one-sixteenth at 1-Mbps, and a quarter at 2-Mbps. If a given amount of data is to be sent, the energy requirement and the time needed for transmission are different. Table 3.11 presents these requirements, relative to 1 Mbps.

The basic data unit is the byte (octet), but the symbols are composed of two or four bits, so bytes have to be broken up before their transmission. The rule defined is: "Send blocks of bits starting with the less significant ones, but keeping the internal order inside the block."

Table 3.11
Energy and Time Requirements

Data Rate	Energy	Time
1 Mbps	1	1
2 Mbps	4	0.5

Figure 3.22 shows the order for 2-Mbps data rate. Notice that the pair order is kept, so the second symbol transmitted will be the one corresponding to the values of the bit pair 32. The same rule is applied to 1 Mbps, but two 4-bit blocks are set up for each byte.

Optical transmitters Regarding the optical parameters, two kinds of devices are described: The first is mainly oriented toward active repeaters on the ceiling, while the second is to be implemented in the computers, perhaps as a PCMCIA card. They differ only in the emission properties, having common receiver characteristics.

The first one has a high power level (2W); this value will exhaust a portable computer battery in a few minutes. Nevertheless, it is not intended for mobile equipment but rather for being placed on the ceiling to act as an active repeater (AP), so it can be powered from the mains by a power supply.

The beam pattern has a torus-like shape. A spherical or Lambertian profile would send too much power to the floor lying just under the emitter and much less to the walls. Because the most probable placement of receivers is on a table at a fixed height and several meters apart from the

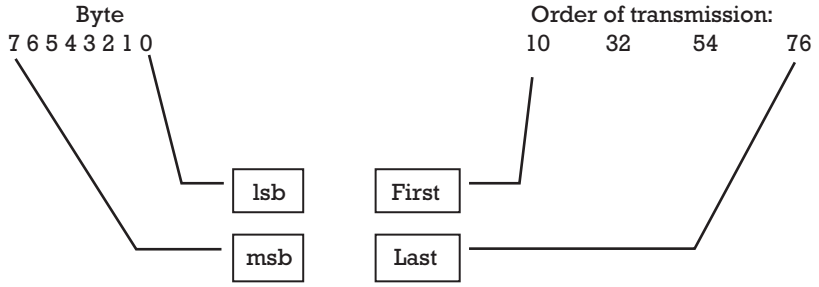

Figure 3.22 Bit ordering in a 2-Mbps transmission.

vertical of the emitter (see Figure 3.23), the beam has been designed to cover these positions without wasting power in other directions. Nevertheless, as the emitted power is high, reflections on walls and furniture can reach the receivers with enough power to be detected.

The second profile (or mask as it is defined in the standard) emits less peak power (500 mW), and the beam is closer to a cone of 60 degrees. This beam can be aimed at the repeater or at other devices; in fact, it can be oriented to any white surface that serves as a passive reflector or diffuser.

This one is intended to be used in a portable computer, PDS, or any handheld device. Although the power is not so low, it should be kept in mind that only when the device is emitting is high power drawn. In the common use of this equipment, traffic is heavily asymmetric (i.e., portable devices are to receive most of the information), and only several ACKs are returned. The option to reduce the data rate to 1 Mbps (with a lower duty cycle) while sending could be a good opportunity for battery-powered devices, while receiving at 2 Mbps. The wavelength used is in the 850–950-nm range, the one used by IrDA and TV remote controllers, so LED devices are very cheap. To avoid interference with other (future) devices operating at higher speed, the electrical power spectrum of the pulses should fall by 20 dB at 15 MHz.

Optical receivers Two power-related parameters are defined for the receiver: sensitivity and dynamical range. The sensitivity is defined as the minimum power density on the receiver surface to achieve a FER of $4 \cdot 10^{-5}$

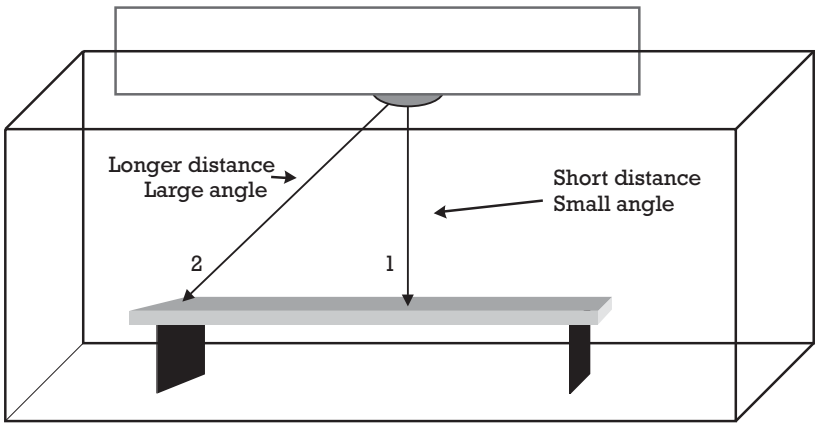

Figure 3.23 The optical intensity for direction 2 has to be larger than for 1.

in a 512-byte frame. This optical power is disturbed by an unmodulated background IR source, so the receiver should be able to recover the signal while other sources, mainly the IR spectrum of illumination lamps, are present.

The sensitivity levels are $2 \cdot 10^{-5}$ mW/cm^2 at 1 Mbps and $8 \cdot 10^{-5}$ mW/cm^2 at 2 Mbps. These values are minima and can be improved to achieve longer distances.

The optical power on the receiver can take a wide range of values and depends on the relative orientation between emitter and receiver, furniture, walls, and the people using the system. Based on this, a large (30-dB) dynamical range is mandatory for the receiver circuit. That fact is also responsible for no range distance having been defined in the standard.

The PHY uses three signals to describe the average availability. These signals are energy detect (ED), carrier sense, and CCA.

ED is triggered when the optical energy changes more than 1 W/cm^2 in the 1–10-MHz band. Carrier sense is asserted when a preamble signal has locked the receiver clock. The receiver clock circuitry may be able to lock it, although the level of the signal is not enough to trigger ED; in this case carrier sense will be asserted, and ED will not.

The CCA signal has two states: IDLE if the channel is free and BUSY if the opposite is true. When both ED and carrier sense are not asserted, CCA will assume the IDLE state. Nevertheless, other systems (remote controls, electrically switched fluorescent lamps, etc.) can maintain ED asserted although the channel is free. In this case, after a period of time with ED asserted but without carry detection, the CCA will change to IDLE. The state of CCA is sent to MAC, so it can manage all the possible events.

3.5 Conclusions and applications

The standard IEEE 802.11 is the first serious and universally accepted standard for WLAN. It covers the high-quality area of mobile data communications. As a living standard, several improvements have been developed. The 802.11a and 802.11b versions take advantage of new RF channels to increase the data rate to 20 or 25 Mbps in the 5.2- and 17.1-GHz bands.

Because the main objectives of WLAN are PDAs and portable computers, a great number of these products are implemented in PCMCIA

cards. As the features of these operating systems promote networking, many services such as file transfer, e-mail, and, of course, Internet connection are available to users. Other equipment can also take advantage of mobile networking. For example, moving robots can be controlled from a central computer in a more flexible way. Other scenarios include meeting rooms, quality control in industries, academic rooms, hospitals, and libraries.

The home scenario is not well suited for 802.11 networks because of the price. Simpler and cheaper solutions in which all the appliances are controlled by a central unit, are under development for home use.

The possibility of e-mailing and Web browsing offered by new mobile telephones is not, in any sense, an alternative to wireless networking. IEEE 802.11 networks are intended to be implemented as a facility by companies for their workers. It should never be forgotten that they are based on a LAN standard. On the other hand, smart mobile phones have improved information processing capabilities. They do not allow users to modify or resend documents of several pages, including figures, graphs, and tables.

The main makers of WLAN products have set up the WLAN Association (WLANA) to extend and publicize the capabilities of these networks.

References

[1] Pahlavan, K., and A. Levesque, *Wireless Information Networks*, New York: J. Wiley & Sons, 1995.

[2] Barry, J., *Wireless Infrared Communications,* Dordrecht, Netherlands: Kluwer Academic Publishing, 1994.

[3] Kahn, J., "Wireless Infrared Communications," *Tutorial at PIMRC '96*, Taipei, Taiwan, ROC.

[4] Kahn, J. M., and J. R. Barry, "Wireless Infrared Communications," *Proceedings of the IEEE*, Vol. 85, No. 2, Feb. 1997, pp. 265–298.

[5] Pérez-Jiménez, R., V. M. Melián, and M. J. Betancor, "Analysis of Multipath Impulse Response of Diffuse and Quasi-Diffuse Optical Links for IR-WLAN," *Proceedings IEEE INFOCOM'95*, April 4–6, 1995, Boston, MA, pp. 7d.3.1–7d.3.7.

[6] Chen, K.C., "Medium Access Control of Wireless LANs for Mobile Computing," *IEEE Network*, Sept.–Oct. 1994, pp. 50–63.