CHAPTER

UNLICENSED SPECTRUM AND LICENSE-ASSISTED ACCESS

17

Spectrum is fundamental for wireless communication, and there is a never-ending quest for more spectrum to meet the ever-increasing demands of increased capacity and higher data rates. Increasing the amount of spectrum available to LTE is thus highly important. LTE and previous cellular systems were designed for licensed spectrum where the operator has an exclusive license for a certain frequency range. Licensed spectrum offers many benefits since the operator can plan the network and control the interference. It is thus instrumental to providing quality-of-service (QoS) guarantees and wide-area coverage. However, the amount of licensed spectrum an operator has access to may not be sufficient, and there is typically a cost associated with obtaining a spectrum license.

Unlicensed spectrum, on the other hand, is open for anyone to use at no cost, subject to a set of rules, for example, on maximum transmission power. Since anyone can use the spectrum, the interference situation is typically much more unpredictable than for licensed spectrum. Consequently, QoS and availability cannot be guaranteed. Furthermore, the maximum transmission power is modest, making it less suitable for wide-area coverage. Wi-Fi and Bluetooth are two examples of communication systems exploiting unlicensed spectrum.

From the discussion earlier, it is seen that these two spectrum types have different benefits and drawbacks. An attractive option is to combine the two such that licensed spectrum is used to provide wide-area coverage and QoS guarantees with unlicensed spectrum as a local-area complement to increase user data rates and overall capacity without compromising on coverage, availability, and reliability.

Using unlicensed spectrum to complement LTE in licensed spectrum is in itself not new. Many operators are already using Wi-Fi to boost capacity in local areas. In such mixed deployments, the selection of the radio access to use, LTE or Wi-Fi, is currently handled autonomously by the device. This has some drawbacks as the device may select Wi-Fi even if staying on LTE would provide a better user experience. One example of such a situation is when the Wi-Fi network is heavily loaded while the LTE network enjoys a light load. To address such situations, the LTE specifications have been extended in release 12 with means for the network to assist the device in the selection procedure [82]. Basically, the network configures a signal-strength threshold controlling when the device should select LTE or Wi-Fi.

4G, LTE-Advanced Pro and The Road to 5G. http://dx.doi.org/10.1016/B978-0-12-804575-6.00017-0

Furthermore, release 13 also supports LTE—WLAN aggregation where LTE and WLAN are aggregated at the PDCP level using a framework very similar to dual connectivity.

The primary reason for Wi-Fi integration is to cater to operators with existing Wi-Fi deployments. However, a tighter integration between licensed and unlicensed spectrum can provide significant benefits. For example, operating one LTE network covering both spectrum types is simpler than operating two different technologies, one for licensed and the other for unlicensed spectrum. Mobility is another benefit. LTE was designed with mobility in mind from the start while Wi-Fi performs best when the users are more or less stationary. QoS handling and the possibility for increased spectral efficiency in a scheduled system are additional examples of the benefits of using LTE for both licensed and unlicensed spectrum.

With these points in mind, 3GPP has specified *license-assisted access* (LAA) as part of release 13. The basis for LAA is carrier aggregation with some component carriers using licensed spectrum and some unlicensed spectrum, see Figure 17.1. Mobility, critical control signaling and services demanding high QoS rely on carriers in licensed spectrum, while (parts of) less demanding traffic can be handled by the carriers in unlicensed spectrum. This is the reasoning behind the name "license-assisted access" where licensed spectrum is used to assist access to unlicensed spectrum.

LAA targets operator-deployed low-power nodes in the 5 GHz band in, for example, dense urban areas, indoor shopping malls, offices, and similar scenarios. It is not intended as a replacement for user-deployed Wi-Fi nodes at home as it requires access to licensed spectrum, nor is it intended for wide-area coverage as the allowed transmission power in unlicensed bands is fairly low. Initially, in release 13, LAA will support downlink traffic with release 14 extending it to also handle uplink traffic.

One important characteristic of LAA is the fair sharing of unlicensed spectrum with other operators and other systems, in particular Wi-Fi. There are several mechanisms that enable this. First, *dynamic frequency selection* (DFS), where the LAA node searches and finds a part

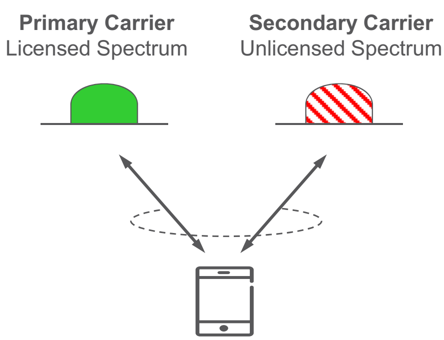

FIGURE 17.1

License-assisted access.

of the unlicensed spectrum with low load, is used, thereby avoiding other systems if possible. Release 13 also supports a *listen-before-talk* (LBT) mechanism,[1] where the transmitter ensures that there are no ongoing transmissions on the carrier frequency prior to transmitting. With these mechanisms, fair coexistence between LAA and Wi-Fi is possible, and LAA can in fact be a better neighbor to Wi-Fi than another Wi-Fi network [56].

17.1 SPECTRUM FOR LAA

Unlicensed spectrum exists in multiple frequency bands. In principle, any unlicensed band can be used for LAA although the 5 GHz band is the main target. One reason is the availability of fairly large amounts of bandwidth in the 5 GHz band and a reasonable load compared to the 2.4 GHz band.

The 5 GHz band is available in most parts of the world, see Figure 17.2, although there are some differences between the different regions. In the following, a brief overview of the regulatory requirements in different parts of the world is given. For a more detailed overview, see [57] and the references therein.

The lower part of the band, 5150−5350 MHz, is typically intended for indoor usage with a maximum transmission power of 23 dBm in most regions. In total 200 MHz is available, divided in two parts of 100 MHz each with regulations calling for DFS and *transmit power control* (TPC) in the 5250−5350 MHz range.

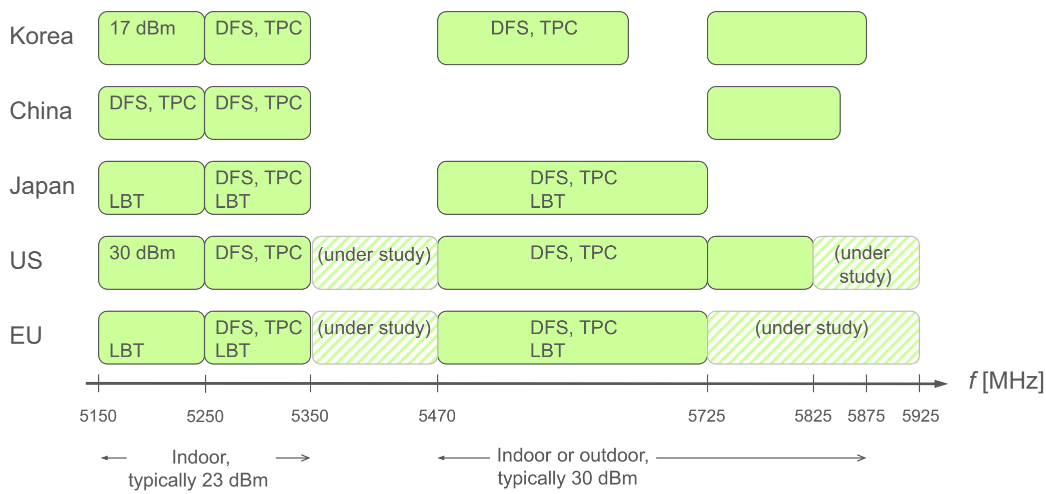

FIGURE 17.2

Overview of unlicensed frequency bands in different regions.

[1]Sometimes the term *clear-channel assessment* (CCA) is used instead of LBT although herein CCA is used for assessing whether the channel is available not including any backoff procedure.

DFS means that the transmitter continuously must assess whether the spectrum is used for other purposes. If such usage is detected, the transmitter must vacate the frequency within a specific time (e.g., 10 s) and not use it again until at least a certain time (e.g., 30 min) has passed. The purpose is to protect other systems, primarily radars, which have higher priority for the usage of unlicensed spectrum. TPC means that a transmitter should be able to reduce its transmission power below the maximum allowed power with the intention to reduce the overall interference level when needed.

In the part of the band above 5470 MHz, transmission powers up to 30 dBm and outdoor usage is allowed in many regions. The amount of spectrum differs across regions but up to 255 MHz can be available. DFS and TPC are mandatory.

In several regions, particularly Europe and Japan, LBT is mandatory. LBT is a mechanism where the transmitter listens to any activity on the channel prior to each transmission in order not to transmit if the channel is occupied. It is thus a much more dynamic coexistence mechanism than DFS. Other regions, for example, the United States, do not have any LBT requirements.

There may also be requirements on the minimum transmission bandwidth, how long a single transmitter may use the channel, and the fraction of time a transmitter must leave the channel idle. For example, a channel occupancy time of at most 6 ms[2] is allowed within Europe while the corresponding number for Japan is 4 ms. The channel occupancy time sets a limit on how long a transmission burst can be. Europe also has two sets of rules for unlicensed spectrum usage described in [58], one for frame-based equipment and one for load-based equipment. The two sets of rules were specified to be applicable to the now-defunct Hiperlan/2 standard and Wi-Fi, respectively. LAA is designed with the load-based rules in mind.

17.2 WI-FI BASICS

Wi-Fi is a well-known system operating in unlicensed spectrum. Although unlicensed spectrum is not allocated to Wi-Fi but free to use by anyone following the regulatory requirements, Wi-Fi is, and is likely to continue to be, a common radio-access technology in unlicensed spectrum. A brief overview of the Wi-Fi behavior and the underlying 802.11 standard with focus on 802.11ac is therefore given in the following in order to better understand some of the LAA design choices.

Wi-Fi divides the available spectrum into several 20 MHz frequency channels. Transmissions use one, or in case of channel bonding, multiple such frequency channels. Coordinating those transmissions across nodes can be done either in a centralized or in a distributed manner. In practice, centralized coordination is seldom used in Wi-Fi and distributed coordination is by far the most common. The following description therefore focuses on *enhanced distributed channel access* (EDCA), an enhancement providing QoS

[2]Under certain conditions, 8 ms or 10 ms is allowed.

enhancements to the *distributed coordination function* (DCF) part of the original 802.11 specifications.

Nodes using EDCA use LBT which includes a backoff procedure prior to transmitting. First, the transmitter listens and waits until the frequency channel is available during a period of time known as *arbitration inter-frame space* (AIFS). A frequency channel is declared available if the power level is lower than −62 dBm and no Wi-Fi preamble with a power level of −82 dBm or higher is detected, otherwise it is unavailable.

Once the frequency channel has been declared available during (at least) AIFS, the transmitter starts the backoff procedure. A backoff timer is initialized with a random number, representing the duration in multiples of the 9 μs slot time the channel must be available before a transmission can take place. The backoff timer is decreased by one for each 9 μs slot the channel is sensed idle, whereas whenever the channel is sensed busy the backoff timer is put on hold until the channel has been idle for a period of AIFS.

Once the backoff timer has expired, the node has acquired a *transmission opportunity* (TXOP). Multiple packets can be transmitted back to back during the TXOP without any need for LBT between them as long as the maximum TXOP duration is not violated. If the TXOP duration is set to zero, only a single packet is allowed, and a new backoff procedure has to be used for each packet.

Upon reception of a packet (or a set of contiguous packets[3]) the receiver responds with an acknowledgment message. The acknowledgment is transmitted a *short inter-frame space* (SIFS) duration of 16 μs after the reception of the packet. Since the SIFS is shorter than the AIFS, no other Wi-Fi user can grab the channel during this period. If no acknowledgment is received either the data or the acknowledgment itself was lost and a retransmission is performed. After completing the TXOP and prior to transmitting the next packet(s) from the transmission buffer, regardless of whether it is a retransmission or a new packet, a random backoff is performed by using the same procedure as described earlier. The reason for the backoff procedure is to avoid collisions between multiple transmitters. Without the random backoff, two nodes waiting for the channel to become available would start transmitting at the same time, resulting in a collision and most likely both transmissions being corrupted. With the random backoff, the likelihood of multiple transmitters simultaneously trying to access the channel is greatly reduced.

The random value used to initialize the backoff timer must be within the *contention window* and is drawn from a uniform distribution with exponentially increasing size for each retransmission attempt. For the nth retransmission attempt, the backoff time is drawn from the uniform distribution $[0, \min(2^{n-1}CW_{\min}, CW_{\max})]$. The larger the contention window, the larger the average backoff value, and the lower the likelihood of collisions.

[3]In the original 802.11 standard the receiver responds with an acknowledgment after each packet, while the 802.11e amendment, used in 802.11ac, introduced block acknowledgments enabling a single acknowledgment message to cover multiple packets. This is often used in combination with TXOP > 0.

Table 17.1 Default Parameters for the Different Access Classes (for an access point)

Priority Class	CW_{min}	CW_{max}	AIFS	TXOP
Voice	3	7	25 μs	3.008 ms
Video	7	15	25 μs	1.504 ms
Best effort	15	63	43 μs	0
Background	15	1023	79 μs	0
Legacy DCF	15	1023	34 μs	0

The original 802.11 standard relied on the distributed coordination function and did not support handling of different traffic priorities; all traffic was treated with the same priority. This was addressed by the introduction of EDCA where one of the major enhancements was handling of different traffic priorities. This is done using priority-class-dependent values for CW_{min} and CW_{max} as shown in Table 17.1 for the four priority classes.[4] High-priority traffic uses a smaller contention window to get faster access to the channel while low-priority traffic uses a larger contention window, increasing the likelihood of high-priority data being transmitted before low-priority data. Likewise, different duration of AIFS are used for the different priority classes, resulting in high-priority traffic sensing the channel for a shorter period of time and grabbing the channel quicker than low-priority traffic. For comparison, the last row in Table 17.1 shows the corresponding values when using the legacy DCF functionality instead of EDCA.

LBT and the associated backoff procedures are illustrated in Figure 17.3 for three different users. The first user gets access to the channel relatively quick as there are no other users actively transmitting. Once the backoff timer has expired, the packet is transmitted. The third user, for which a packet arrived while the first user is transmitting, found the channel to be occupied. The backoff timer is held and not decremented until the channel is available again. However, in the meantime the second user grabbed the channel and again the third user's backoff timer is put on hold, deferring the transmission further. Once the backoff timer for the third user expires the data are transmitted.

One benefit of EDCA lies in its distributed nature—any device can communicate with any other device without the need for a centralized coordination node. However, the use of LBT with backoff timers implies a certain overhead. At higher loads such a distributed protocol is less efficient and higher efficiency would be possible with a centralized scheduling function. This would also be well in line with the common scenario of multiple devices communicating with a central access point.

[4]The numbers given are for an access point, clients use slightly different values (larger AIFS for voice and video and larger CW_{max} for best effort). See 802.11e for a description on how the parameters relate. For 802.11ac, the parameters are given as AIFS = SIFS + $n\sigma$ where n is an access-class dependent parameter and $\sigma = 9$ μs is the slot time.

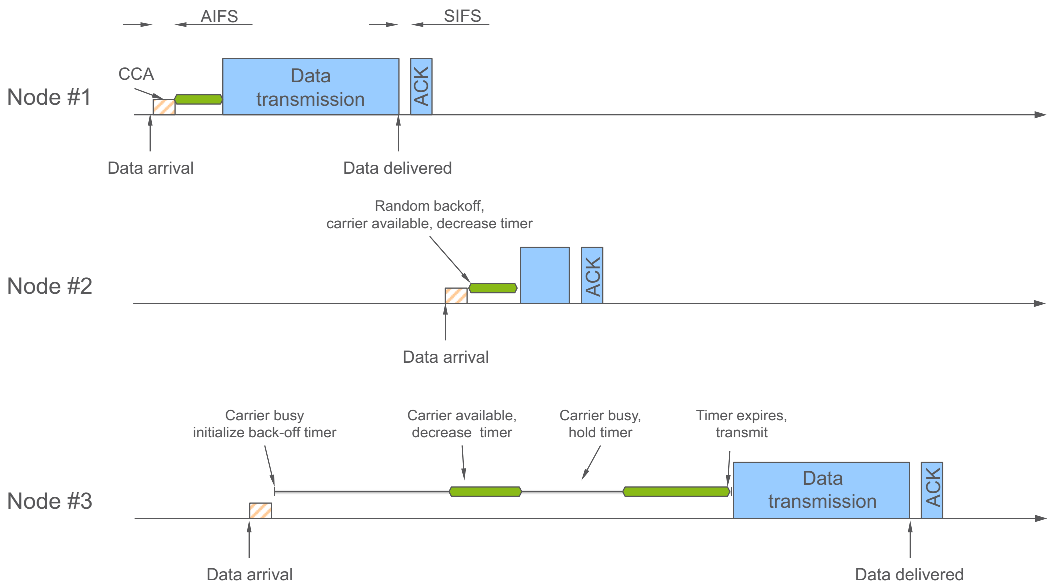

FIGURE 17.3

Example of LBT in Wi-Fi.

Another aspect of the LBT mechanism in Wi-Fi is that it, in essence, leads to a time reuse larger than one across multiple access points, each providing coverage in a certain area. A transmission from one access point is typically detected at the neighboring access points, all of which will find the frequency channel occupied and therefore defer their transmissions. This is less efficient from a capacity perspective as it is known that reuse one results in higher capacity.

The overview in the preceding paragraphs is brief as the intention is to provide some background to some of the LAA design choices. For a more detailed description of Wi-Fi and aspects such as the hidden node problem occurring in distributed schemes, the reader is referred to [59].

17.3 TECHNOLOGY COMPONENTS FOR LAA

LAA is, as already mentioned, based on the carrier aggregation framework in LTE. The primary component carrier, and optionally, one or more secondary component carriers operate in licensed spectrum, aggregated with one or more secondary component carriers operating in unlicensed spectrum. With the carrier aggregation enhancements in release 13 enabling aggregation of up to 32 carriers, up to 620 MHz of unlicensed spectrum can be exploited by a single device in combination with 20 MHz of licensed spectrum. In practice, the use of carrier aggregation implies that licensed and unlicensed spectrum is handled by the same node, possibly with remote radio heads connected over low-latency backhaul. To handle separate nodes for licensed and unlicensed spectrum, interconnected with nonideal backhaul, LAA could be based on the dual-connectivity framework, but this is not part of release 13 and is left for potential introduction in future releases.

No changes are required for the component carriers in licensed spectrum as this is the type of spectrum for which LTE was designed. Many of the LTE design choices hold for the component carriers in unlicensed spectrum as well, although some aspects are different compared to licensed spectrum, primarily due to multiple operators and/or systems using the same spectrum. DFS, TPC, LBT, discontinuous transmission, and radio-resource management are some of the larger areas affected by the introduction of LAA, although some smaller enhancements were made in other areas as well. A quick overview of these components is given with a more detailed description in the following sections.

DFS is used to vacate the frequency channel upon detecting interference from radar systems. This is a requirement in some frequency bands. DFS is also used when activating the node, for example, at power up, in order to find an unused or lightly used portion of the spectrum for future transmissions. No specification enhancements are needed to support DFS; implementation-specific algorithms in the eNodeB are sufficient.

TPC is required in some bands and regions, requiring the transmitter to be able to lower the power by 3 or 6 dB relative to the maximum output power. This is purely an implementation aspect and is not visible in the specifications.

LBT ensures that the carrier is free to use prior to transmission. It is a vital feature that allows fair sharing of the spectrum between LAA and other technologies such as Wi-Fi. In some regions, in particular Europe and Japan, it is a mandatory feature. The introduction of LBT to LTE impacts the specifications and is a completely new feature in LTE which has been extensively discussed in 3GPP.

Downlink discontinuous transmission is required not only to comply with regulations, since some regions limit the maximum transmission duration, but also to nicely coexist with other users of the unlicensed spectrum. Only when the channel is declared available should transmissions take place. In particular, continuous transmission of cell-specific reference signals, as originally required by LTE, is not possible, impacting not only data demodulation but also RRM functionality, and requiring additions to the specifications. To some extent, discontinuous transmission can be seen as small-cell on/off operating on a subframe basis. The introduction of discontinuous transmission will impact time and frequency synchronization, *automatic gain control* (AGC) setting, and CSI measurements as these functionalities typically rely on certain reference signals always being present.

The fact that multiple operators in the same area may use the same carrier frequency, unlike the licensed case where the carrier frequency is unique to an operator, also needs to be considered. In addition, multiple operators may also end up using the same physical-layer cell identity. The structure of many signals and channels are linked to the physical-layer cell identity (e.g., CSI-RS, downlink scrambling, and CRS sequences). This would lead to the situation that a device connected to operator A may successfully receive signals and channels originating from operator B. However, the likelihood of this is very small [57] and the eNodeB implementation can avoid this by trying to detect the other physical-layer cell identities being used on the targeted carrier and by selecting an unused cell identity.

17.3.1 DYNAMIC FREQUENCY SELECTION

The purpose of DFS is to determine the carrier frequencies for the secondary carriers in order to find an available or at least lightly loaded carrier frequency. If this is successful, there are no coexistence problems with other systems using the unlicensed spectrum. Since around 25 frequency channels, each 20 MHz wide, are part of the 5 GHz band, and the output power is fairly low, there is a reasonably high likelihood to find unused or lightly loaded frequency channels.

DFS is performed at power-up of an LAA cell. In addition to power-up, DFS can also be performed on an event-triggered basis. For example, the base station can periodically measure the interference or power level when not transmitting in order to detect whether the carrier frequency is used for other purposes and if a more suitable carrier frequency is available. If this is the case, the base station can reconfigure the secondary component carriers to a different frequency range (essentially an inter-frequency handover).

DFS is, as already mentioned, a regulatory requirement for some frequency bands in many regions. One example motivating DFS being mandated is radar systems, which often have priority over other usage of the spectrum. If the LAA base station detects radar usage, it must stop using this carrier frequency within a certain time (typically 10 s). The carrier frequency is not to be used again until at least 30 min has passed.

The details of DFS are up to the implementation of the base station, and there is no need to mandate any particular solution in the specifications. RSSI measurement from the device can be used by the implementation-specific DFS algorithm.

17.3.2 LISTEN BEFORE TALK

In LTE, all transmissions are scheduled and the scheduler is in complete control of when transmissions occur on a carrier in licensed spectrum. Scheduling is also used for LTE transmissions in unlicensed bands, but an inherent consequence of unlicensed spectrum is that there can be multiple transmitters (potentially belonging to different operators) using the same spectrum without any coordination between them. Although channel selection aims to find one or more frequency channels with no or very light usage, and in many cases succeeds in this, simultaneous usage of the same frequency channel by multiple systems cannot be precluded. LBT refers to the mechanism used by LAA to check the availability of a channel prior to using it. In LAA, the transmitter listens to potential transmission activity on the channel prior to transmitting.

Regulatory requirements vary across regions with some regions, for example, Japan and Europe, mandating the use of LBT in unlicensed bands while some other regions being more relaxed. In regions where there are no regulatory requirements on LBT, one could in principle deploy LTE release 12 and use small-cell on/off of the secondary carriers in unlicensed spectrum to realize a fractional loading mechanism for Wi-Fi coexistence. When the secondary carriers are active, Wi-Fi would detect the channel as busy and not transmit. Similarly, when the secondary carriers are inactive Wi-Fi can use the spectrum. The length of the on and off periods could vary according to the load but may be in the order of 100 ms. Although such a mechanism works and can be tuned for fair spectrum sharing with Wi-Fi, it would not be suitable for global deployment as some regions require LBT. This is the reason why LBT is a vital part of LAA since LAA is targeted to provide a single global solution framework. Note that the intention with LBT is to ensure fair sharing and coexistence with *other* networks, for example, Wi-Fi, and not to coordinate transmissions *within* an LAA cell. Coordinating transmissions between multiple devices in the same LAA cell is done through scheduling in the same way as in licensed spectrum.

Downlink LBT in LAA is based on the same underlying principles as Wi-Fi. Note that LBT is a much more dynamic operation than channel selection because it is performed prior to each transmission burst. It can therefore follow variations in the channel usage on a very fast time scale, basically in milliseconds.

A transmission burst is a contiguous transmission spanning one or more subframes on a given component carrier from a given eNodeB with no transmission immediately before or after. Before transmitting a burst, the eNodeB assesses whether the frequency channel is available or not by performing the LBT procedure with a random backoff as illustrated in Figure 17.4. Whenever the eNodeB has found the channel to be idle after previously being busy, it executes a defer period. The defer period starts with a delay of 16 μs, after which the eNodeB measures the energy for a certain number of 9 μs slots which depends on the priority class as shown in Table 17.2. The total length of the defer period, whose purpose is to avoid collisions with potential Wi-Fi acknowledgments transmitted 16 μs after data reception, is at least equal to 25 μs.[5] For each observation of the channel in a slot during the defer period, the frequency channel is declared available if the received energy is less than a threshold, which among other things depends on the regulatory requirements, the maximum transmission power, and whether the LAA is the only technology using the frequency channel.

After executing the defer period, the eNodeB performs a full random backoff procedure similarly to what was discussed in Section 17.2. A backoff timer is initialized with a random number, drawn from a uniform distribution [0, *CW*], representing the duration in multiples of 9 μs the channel must be available for before transmission can take place. The availability of the channel within a 9 μs time slot is subject to the same rules as in the defer period described

[5]The time 25 μs equals the sum of the 16 μs SIFS and the 9 μs slot duration.

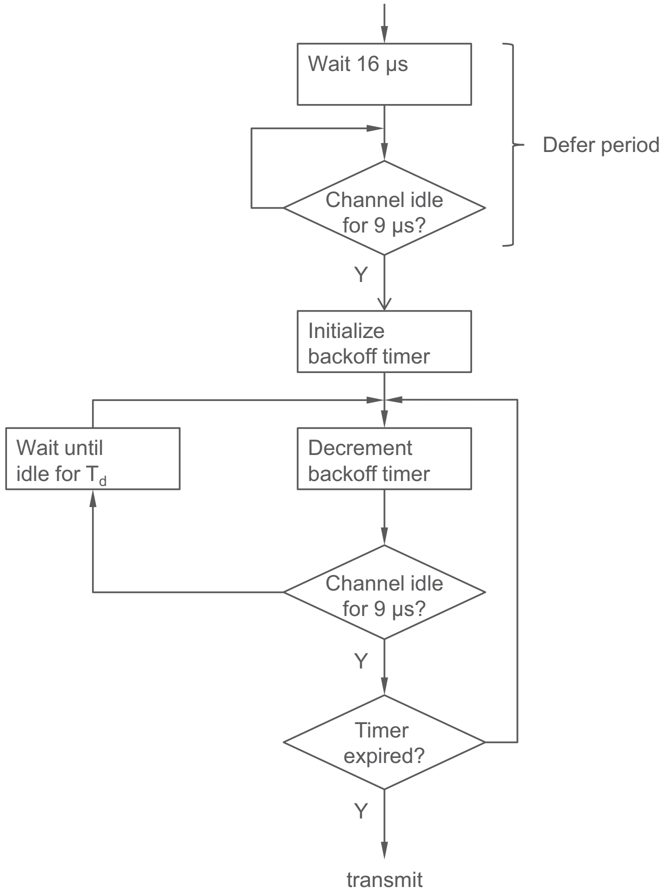

FIGURE 17.4

The LBT procedure for LAA.

Table 17.2 Contention-Windows Sizes for Different Priority Classes

Priority Class		Defer Period	Possible *CW* Values $\{CW_{min}, ..., CW_{max}\}$	Max Burst Length[a]
1	Signaling, voice, real-time gaming	25 µs	{3,7}	2 ms
2	Streaming, interactive gaming	25 µs	{7,15}	3 ms
3	Best-effort data	43 µs	{15,31,63}	10 ms or 8 ms
4	Background traffic	79 µs	{15,31,63,127,255,511,1023}	10 ms or 8 ms

[a] Regulatory requirements may limit the burst length to smaller values than in the table. If no other technology is sharing the frequency channel, 10 ms is used, otherwise 8 ms.

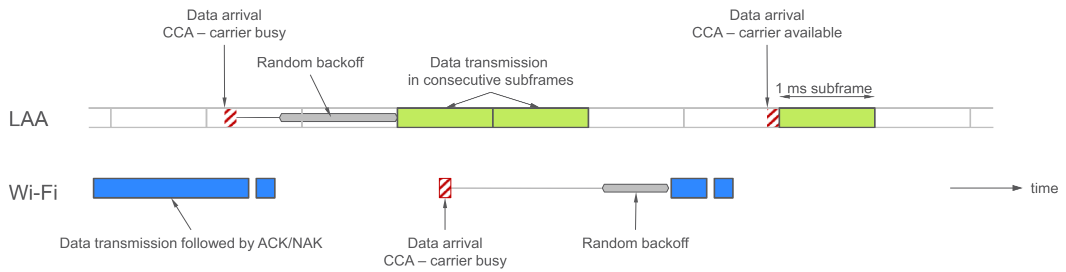

FIGURE 17.5

Example of coexistence between Wi-Fi and LAA.

earlier. Once the timer has expired, the random backoff is completed and the burst is transmitted. When the timer expires, the node may postpone transmission and transmit at a later time if the channel is found to be free for a period of 25 μs. If the channel is found to be busy during that period of 25 μs, the node executes another defer period and then performs a full random backoff as described earlier.

The size of the contention window is adjusted based on hybrid-ARQ acknowledgments received from the device with the contention window CW (approximately) doubling up to a limit $CW_{\max}$ if a negative hybrid-ARQ is received.[6] If the hybrid-ARQ acknowledgment is positive, the contention window is reset to its minimum value, $CW = CW_{\min}$.

There are four different priority classes defined, each with individual contention windows and with different maximum and minimum values of the contention window as depicted in Table 17.2. The intention with different priority classes is to use a shorter backoff for high-priority data, thereby taking priority over low-priority transmissions. The high-priority classes also have a shorter maximum burst length to not block the channel for too long.

The purpose of LBT is, as already discussed, to avoid transmitting when the channel is already in use. One reason is to handle coexistence with other radio-access technologies using the unlicensed spectrum, for example, Wi-Fi as illustrated in Figure 17.5. However, as already discussed in conjunction with the Wi-Fi overview in Section 17.2, LBT may detect transmissions also in neighboring cells, in essence resulting in a time-reuse scheme as can be observed in dense Wi-Fi networks. To maintain the benefits of reuse-one operation, for which LTE was originally designed and which in many scenarios leads to higher capacity, LBT should preferably be blind to LAA transmissions in neighboring cells of the same network but still monitor activity from other radio-access technologies and other operators. One possibility to achieve this is to time synchronize neighboring cells and select a common start time for the LBT procedure such that all cells that are part of the same network perform LBT at the

[6]The description is slightly simplified; the specifications describe in detail how to handle different cases of bundled acknowledgments, see [26] for details.

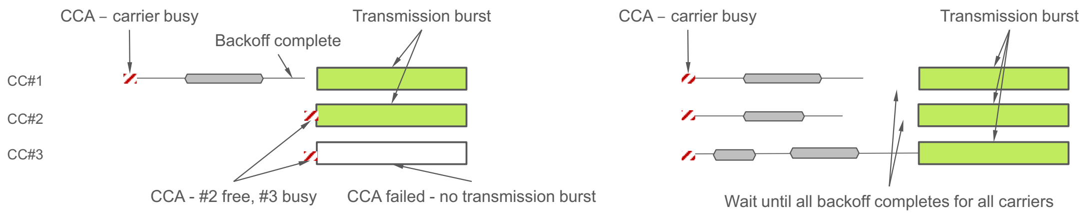

FIGURE 17.6

LBT for multiple carriers, single backoff counter (left), individual backoff counters (right).

same time. The pseudo-random generators should also be aligned across nodes to ensure the same backoff duration in neighboring cells.

The LBT procedure above has been described for the case of one carrier using unlicensed spectrum. If there are multiple carriers using unlicensed spectrum, backoff in conjunction with LBT can be handled in two ways illustrated in Figure 17.6:

- A single backoff value, valid for one of the component carriers in unlicensed spectrum. When the counter has expired, transmissions may take place on all component carriers subject to passing a clear-channel assessment (CCA) of duration 25 μs on the other carriers prior to transmission.
- Multiple backoff values, one per carrier. Transmissions take place once all backoff counters have reached zero. Note that the backoff completes at different time instants for the different carriers. However, "early" carriers cannot start to transmit as this would make listening on the other carriers impossible and the eNodeB therefore has to wait for the longest backoff counter.

Uplink transmissions in unlicensed spectrum is not part of release 13 but is planned for release 14. In principle, if the uplink transmission follows immediately after a downlink transmission, there is no need for LBT in the uplink as long as the maximum channel occupancy is not exceeded. A 25 μs CCA check prior to transmission would be sufficient, assuming the maximum channel occupancy is not exceeded. When needed, LBT in the uplink can be handled in a similar way as for the downlink. This would imply that a device sometimes may need to ignore an uplink grant received if the device finds the frequency channel to be busy when uplink transmission is to take place.

17.3.3 FRAME STRUCTURE AND BURST TRANSMISSION

When the channel is found to be available, the eNodeB can initiate a transmission burst the format of which is described later. However, before going into the details, a description of the LAA frame structure as such is beneficial.

The unlicensed 5 GHz band is an unpaired band and hence TDD is the relevant duplex scheme, making frame structure type 1 unsuitable for LAA. However, since LBT is used and

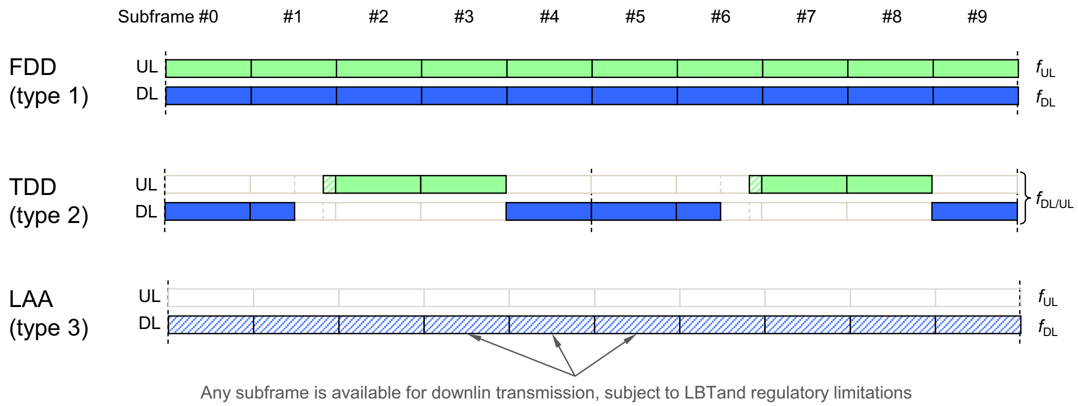

FIGURE 17.7

Frame structure type 3.

transmission burst can start in any subframe, frame structure type 2 with its fixed[7] split into uplink and downlink is not suitable either, especially since release 13 is supporting downlink only. Hence, a frame structure allowing downlink transmissions to start in any subframe, subject to LBT, and not enforcing a fixed allocation of uplink and downlink subframes is needed. For this reason, frame structure type 3, illustrated in Figure 17.7 was introduced in release 13. From most perspectives, frame structure type 3 has the same mapping of signals and channels as frame structure type 1 and is treated as such in various procedures: for example, carrier aggregation procedures for transmitting hybrid-ARQ acknowledgments in the uplink.

As described in the previous section, transmissions may start immediately after the LBT procedure is finished, alternatively at a later point in time as long as the channel is found available immediately before the transmission. Being able to start data transmissions only at one instant every 1 ms clearly has limitations in terms of LAA operation in highly loaded channels as another transmitter may grab the channel in the meantime. One possibility to avoid this would be for an eNodeB implementation to start transmission of an arbitrary "reservation signal" until the subframe starts to ensure the channel is available once the data transmission starts. To reduce the impact from the limitation of data transmissions starting at subframe boundaries only, the possibility to start data transmission also at the slot boundary inside a subframe, a partially filled subframe, is supported for LAA, see Figure 17.8. If the possibility to start downlink transmission at the slot boundary is enabled by RRC signaling, any transmissions starting in the second slot uses the same mapping for the second slot as otherwise used in the first slot of a normal subframe—that is, there is a control region at the

[7]With the introduction of eIMTA some flexibility in the uplink–downlink allocation is possible although only on a frame basis.

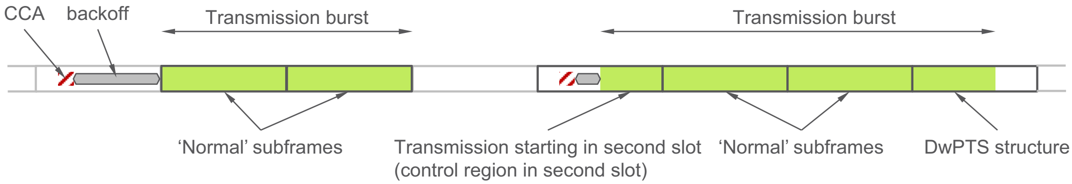

FIGURE 17.8

Illustration of transmission bursts partially filled subframes.

beginning of the second slot in this case. Neither synchronization signals, nor CSI-IM resources are assumed to be part of such a partially filled subframe. A device configured to support PDSCH transmissions starting in the second slot obviously need to monitor PDCCH/EPDCCH in the second slot in addition to the first slot.

The length of a transmission burst depends on the amount of data to transmit and regulatory requirements. As a consequence, PDSCH transmission may need to end prior to the end of a subframe. The last subframe in a burst can therefore use one of the DwPTS structures, in addition to occupying a full subframe. Unlike TDD operation in licensed spectrum, where the device knows in advance where the DwPTS is located from the semi-static uplink–downlink configuration, this is not the case in LAA as the burst length is a dynamic quantity. Instead, the eNodeB signals to the LAA terminal the occurrence of an ending partial subframe in that subframe and in the prior subframe. The signaling uses DCI format 1C and a reserved RNTI value, the CC-RNTI.

Data demodulation can be based on DM-RS or CRS, although the device can only assume these signals are present when a burst is transmitted. Transmission modes 1, 2, 3, 4, 8, 9, and 10 are supported. Only normal cyclic prefix can be used for LAA which is reasonable as the transmit power limitations for unlicensed spectrum implies relatively small cells with modest time dispersion.

17.3.4 REFERENCE SIGNALS AND DISCONTINUOUS TRANSMISSION

Discontinuous transmission—that is, the possibility for completely blank subframes—is a fundamental requirement and a consequence of LBT. Clearly, if the LBT mechanism declares the channel to be occupied, the base station (or device) should not transmit. Consequently, periodically transmitted signals need special attention. In the downlink this impacts the CRS, CSI-RS, and discovery reference signal (DRS) design.

Cell-specific reference signals cannot occur in every subframe as is the baseline for LTE. Hence, functionality such as AGC setting, time–frequency synchronization, RRM measurements, channel-state feedback, and demodulation of data and control channels need to rely on signals that are part of a transmission burst only. AGC setting and time–frequency synchronization can be handled by the CRS both in subframes carrying PDSCH as well as the discovery signal. The DM-RS may also be additionally used. For data demodulation both

DM-RS and CRS can be used, although the device can only assume these signals are present when a burst is transmitted.

Similarly to the CRS, CSI-RS may not be possible to transmit periodically. Furthermore, intermittent transmissions as a result of LBT as well as other non-LAA usage of the unlicensed spectrum typically lead to a rapidly fluctuating interference environment. Interference measurements when the serving cell is not transmitting may not reflect the interference characteristics when the device is receiving data. Hence, interference measurements for CSI purposes should only occur when the serving cell is transmitting. Transmission occasions for CSI-RS are configured in the same way as for a licensed carrier—that is, according to a certain repetitive pattern—but only those CSI-RS overlapping with a transmission burst are transmitted. Since the device knows when CSI-RS are transmitted in a burst to that device, it can use these CSI-RS as a basis for its CSI reports in the same way as for a licensed carrier. The CSI reports are then transmitted on PUSCH on a licensed carrier.

The DRS was introduced in release 12 to support small-cell on/off where secondary component carriers can be turned off except for periodic transmission of DRS as described in Chapter 15. The device is configured with a time window within which it expects the DRS to be transmitted. DRS are used also in LAA and are the basis for radio-resource management including cell identification. However, as a consequence of transmitting on unlicensed spectrum, DRS transmission must be preceded by a CCA.

If the DRS are transmitted together with the PDSCH—that is, inside a transmission burst—it is automatically subject to the LBT mechanism performed prior to the burst and the discovery signal can be multiplexed[8] with ongoing data transmission in the burst.

On the other hand, if the discovery signal is not part of a transmission burst, the DRS must be preceded by a CCA spanning 25 μs. The DRS is transmitted only if the channel is detected as being available. Since the exact time when the DRS is transmitted depends on the CCA, the discovery signal may move around in time and the device needs to detect whether a DRS is transmitted or not before basing measurements on it, something that is not necessary in licensed spectrum.

The DRS structure is the same as in previous releases, but the duration of a discovery signal is limited to 12 OFDM symbols. Since the discovery signal is potentially moving around in time as a result of the CCA, the PSS/SSS being part of a discovery signal may occur outside subframe 0 and 5. Consequently, frame timing cannot be obtained from the PSS/SSS on the secondary carrier—only subframe timing can be obtained—however, this is not a problem as the frame timing can be obtained from the licensed carrier. Finally, to simplify the DRS detection in the device, the CRS/CSI-RS/PSS/SSS sequences do not vary with the subframe number but are kept unchanged across subframes 0–4 and 5–9, respectively.

The release 12 DRS structure is not contiguous in time. A consequence of this is that another node in principle can find the channel being available during the unused OFDM

[8]Simultaneous transmission of a discovery signal and PDSCH is possible in subframes 0 and 5 only.

symbols between two CRS symbols and start transmitting. One possibility for an implementation to avoid this is by transmitting a "dummy signal" when needed to make the DRS transmission contiguous in time.

Discontinuous transmission for a future extension to the uplink is straightforward as all transmissions are scheduled. Periodic transmissions of uplink sounding reference signals can easily be avoided by relying on aperiodic SRS only.

17.3.5 SCHEDULING, HYBRID-ARQ, AND RETRANSMISSIONS

Scheduling requires no major changes but is handled the same way as carrier aggregation in general. Both self-scheduling and cross-carrier scheduling can be used. Self-scheduling can be beneficial for downlink transmissions to better spread the control signaling load across all the carriers; if the channel is available for control signaling it will also be available for downlink data transmission and a single LBT is sufficient. Only cross-carrier scheduling from the primary carrier, which operates on licensed spectrum, is supported; scheduling an unlicensed carrier from another unlicensed carrier is neither supported, nor useful. Furthermore, in case of cross-carrier scheduling, the PDSCH always starts in the first subframe—that is, partially filled subframes are not supported.

Hybrid-ARQ requires no changes for the downlink; the existing asynchronous hybrid-ARQ scheme can be used. Hybrid-ARQ retransmissions are subject to LBT in the same way as an original transmission. Depending on the transmission duration, retransmissions may follow in the same burst as the original transmission. A device transmits hybrid-ARQ acknowledgments on the uplink primary carrier using the same mechanism as in carrier aggregation in general.

For uplink transmission, which is not part of release 13 but to be considered in later releases, cross-carrier scheduling from licensed spectrum can be beneficial as self-scheduling may require two successful LBTs, one for the downlink control signaling and the other for the actual uplink data transmission. Furthermore, asynchronous hybrid-ARQ operation is required in the uplink, calling for enhancements to the uplink hybrid-ARQ protocol. A synchronous protocol implies a fixed timing for the retransmission and the channel may not be available at a fixed retransmission instant. Asynchronous operation requires the hybrid-ARQ process to be signaled. Consequently, retransmissions must be scheduled with PDCCH or EPDCCH and not PHICH in this case.

17.3.6 RADIO BEARER MAPPING AND QOS CONTROL

QoS handling in LTE is, as discussed in Chapter 4, handled by different radio bearers. Multiple radio bearers are multiplexed and, in case of carrier aggregation, the multiplexed data stream is distributed across the component carriers. Carrier aggregation is thus invisible to the RLC and PDCP layers and a certain radio bearer may be transmitted on an arbitrary subset of component carriers. In case all component carriers are transmitted in licensed spectrum, which was the assumption when carrier aggregation was developed, this is not an issue as all component carriers have essentially the same conditions.

For LAA, where some component carriers are transmitted in unlicensed spectrum, the situation is quite different. Depending on the interference situation, other usage of the unlicensed spectrum, and the outcome of LBT, the component carriers in unlicensed spectrum may have substantially different conditions than the ones in licensed spectrum. At some points, the unlicensed component carriers may have very low availability or be subject to a long latency due to LBT. Hence, for LAA there is a need to control on which component carriers a certain radio bearer is mapped.

In the downlink this is an implementation issue. The scheduler in the eNodeB can control which data from the different radio bearers is mapped to the different component carriers, thus controlling which data is transmitted on licensed spectrum and which is transmitted on unlicensed spectrum.

In the uplink, which is not part of release 13, the situation is more complex. The scheduler cannot control which component carrier a certain piece of data is transmitted upon. Critical data may therefore end up on less reliable unlicensed spectrum, possibly resulting in guaranteed bitrate requirements not being fulfilled. Furthermore, LBT mechanisms may impact when transmission occur on a certain component carrier, thereby affecting any latency requirements. One possible approach could be to schedule data on unlicensed spectrum only when the buffer status reports indicate that no critical data is awaiting transmission, but this could result in inefficient utilization of unlicensed spectrum. Enhancements to control upon which component carriers a certain radio bearer is mapped can therefore be beneficial. However, this is not part of release 13 but may be considered in future releases.

17.4 ENHANCEMENTS BEYOND RELEASE 13

Uplink transmission in unlicensed spectrum was studied to some extent in release 13, but in order to complete release 13 in time it was decided to focus on downlink transmissions. In release 14, support for uplink transmissions on unlicensed spectrum will be added. Some of the aspects to consider for uplink transmissions have been discussed in the previous sections.

Another enhancement for future releases could be to support LAA also in the dual-connectivity framework. Carrier aggregation, which is the basis for LAA in release 13, in practice requires licensed and unlicensed spectrum to be handled in the same node or across an ideal backhaul as a consequence of the tight timing relations between the component carriers. Dual connectivity with more relaxed timing relations between the two nodes would allow greater deployment flexibility in this respect. For example, an existing macro site in licensed spectrum could be complemented by small nodes handling unlicensed spectrum only even without ideal backhaul between the macro site and the small nodes. One could even envision multiple operators with separate macro networks sharing a common set of nodes in unlicensed spectrum—for example, an indoor deployment in an office building.

Stand-alone operation, where LTE is extended to operate in unlicensed spectrum without support for licensed spectrum, is another possible enhancement. This would broaden the applicability of LTE also to entities not owning any licensed spectrum (e.g., small business, venues, and landlords). Examples of the technical enhancements required include mobility, random access, and system information distribution. Most of these enhancements are also needed for extending LAA to use the dual-connectivity framework as a complement to the carrier-aggregation-based design.