

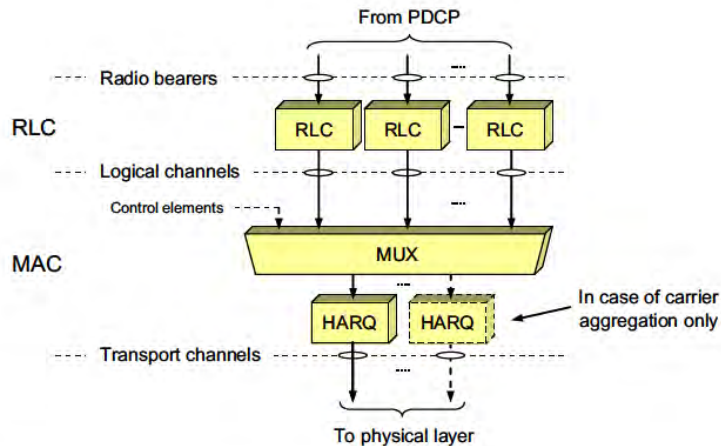
Retransmission Protocols

12

The LTE downlink and uplink transmission schemes have been described from a physical layer perspective in Chapters 10 and 11 respectively. In this chapter, the operation of the two mechanisms responsible for retransmission handling in LTE, the MAC and RLC sublayers, will be described.

Retransmissions of missing or erroneous data units are handled primarily by the hybrid-ARQ mechanism in the MAC layer, complemented by the retransmission functionality of the RLC protocol. The reasons for having a two-level retransmission structure can be found in the trade-off between fast and reliable feedback of the status reports. The hybrid-ARQ mechanism targets very fast retransmissions and, consequently, feedback on success or failure of the decoding attempt is provided to the transmitter after each received transport block. Although it is in principle possible to attain a very low error probability of the hybrid-ARQ feedback, it comes at a cost in transmission power. Keeping the cost reasonable typically results in a feedback error rate of around 1%, which results in a hybrid-ARQ residual error rate of a similar order. Such an error rate is in many cases far too high; high data rates with TCP may require virtually error-free delivery of packets to the TCP protocol layer. As an example, for sustainable data rates exceeding 100 Mbit/s, a packet-loss probability less than 10^{-5} is required [72]. The reason is that TCP assumes packet errors to be due to congestion in the network. Any packet error therefore triggers the TCP congestion-avoidance mechanism with a corresponding decrease in data rate.

Compared to the hybrid-ARQ acknowledgements, the RLC status reports are transmitted relatively infrequently and thus the cost of obtaining a reliability of 10^{-5} or lower is relatively small. Hence, the combination of hybrid-ARQ and RLC attains a good combination of small round-trip time and a modest feedback overhead where the two components complement each other—fast retransmissions due to the hybrid-ARQ mechanism and reliable packet delivery due to the RLC. As the MAC and RLC protocol layers are located in the same network node, a tight interaction between the two protocols is possible. To some extent, the combination of the two can be viewed as *one* retransmission mechanism with *two* feedback channels. However, note that, as discussed in Chapter 8 and illustrated in Figure 12.1, the RLC operates per logical channel, while the hybrid-ARQ operates per transport channel (that is, per component carrier). One hybrid-ARQ entity may therefore retransmit data belonging to multiple logical channels. Similarly, one RLC entity may handle data transmitted across multiple component carriers in the presence of carrier aggregation. Finally, for carrier aggregation, hybrid-ARQ retransmissions must occur on the same component carrier as the

**FIGURE 12.1**

RLC and hybrid-ARQ retransmission mechanisms in LTE

original transmission, as each component carrier has its own independent hybrid-ARQ entity. RLC retransmissions, on the other hand, are not tied to a specific component carrier as, in essence, carrier aggregation is invisible above the MAC layer.

In the following section, the principles behind the hybrid-ARQ and RLC protocols will be discussed in more detail.

12.1 Hybrid ARQ with soft combining

The hybrid-ARQ functionality spans both the physical layer and the MAC layer. Generation of different redundancy versions at the transmitter as well as the soft combining at the receiver are handled by the physical layer, while the hybrid-ARQ protocol is part of the MAC layer. In the presence of carrier aggregation, there is, as already stated, one independent hybrid-ARQ entity per component carrier and terminal. Unless otherwise noted, the description below holds for one component carrier—that is, the description is on a per-component carrier basis.

The basis for the LTE hybrid-ARQ mechanism is a structure with multiple stop-and-wait protocols, each operating on a single transport block. In a stop-and-wait protocol, the transmitter stops and waits for an acknowledgement after each transmitted transport block. This is a simple scheme; the only feedback required is a single bit indicating positive or negative acknowledgement of the transport block. However, since the transmitter stops after

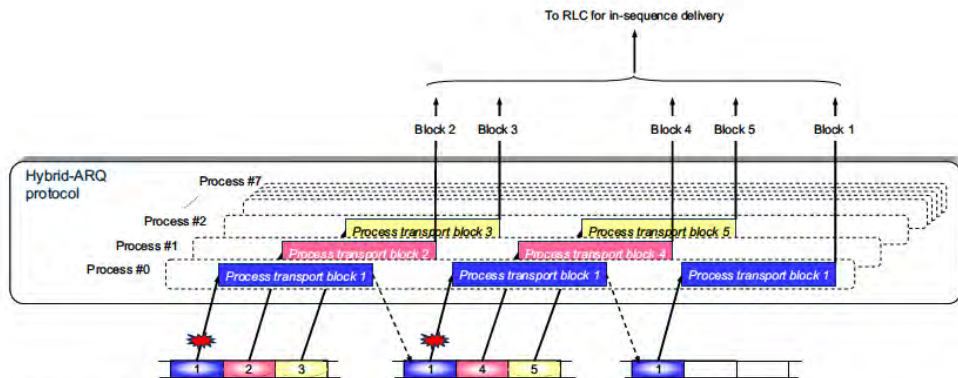


FIGURE 12.2

Multiple parallel hybrid-ARQ processes forming one hybrid-ARQ entity

each transmission, the throughput is also low. LTE therefore applies *multiple* stop-and-wait processes operating in parallel such that, while waiting for acknowledgement from one process, the transmitter can transmit data to another hybrid-ARQ process. This is illustrated in Figure 12.2; while processing the data received in the first hybrid-ARQ process the receiver can continue to receive using the second process, and so on. This structure, multiple hybrid-ARQ processes operating in parallel to form one hybrid-ARQ entity, combines the simplicity of a stop-and-wait protocol while still allowing continuous transmission of data.

There is one hybrid-ARQ entity per terminal (and per component carrier in the case of carrier aggregation; see Figure 12.1). Spatial multiplexing, where two transport blocks can be transmitted in parallel on the same transport channel, as described in Chapter 10, is supported by one hybrid-ARQ entity having two sets of hybrid-ARQ processes with independent hybrid-ARQ acknowledgements. The details for the physical-layer transmission of the downlink and uplink hybrid-ARQ acknowledgements were described in Chapters 10 and 11.

Upon receiving a transport block for a certain hybrid-ARQ process, the receiver makes an attempt to decode the transport block and informs the transmitter about the outcome through a hybrid-ARQ acknowledgement, indicating whether the transport block was correctly decoded or not. The time from reception of data until transmission of the hybrid-ARQ acknowledgement is fixed, hence the transmitter knows from the timing relation which hybrid-ARQ process a received acknowledgement relates to. This is beneficial from an overhead perspective as there is no need to signal the process number along with the acknowledgement.

An important part of the hybrid-ARQ mechanism is the use of *soft combining*, which implies that the receiver combines the received signal from multiple transmission attempts. The principles of hybrid-ARQ with soft combining were outlined in Chapter 6, with

incremental redundancy being the basic scheme in LTE. The receiver needs to know when to perform soft combining prior to decoding and when to clear the soft buffer—that is, the receiver needs to differentiate between the reception of an initial transmission (prior to which the soft buffer should be cleared) and the reception of a retransmission. Similarly, the transmitter must know whether to retransmit erroneously received data or to transmit new data. Therefore, an explicit *new-data indicator* is included for each of the one or two scheduled transport blocks as part of the scheduling information transmitted in the downlink. The new-data indicator is present in both downlink assignments and uplink grants, although the meaning is slightly different for the two.

For downlink data transmission, the new-data indicator is toggled for a new transport block—it is essentially a single-bit sequence number. Upon reception of a downlink scheduling assignment, the terminal checks the new-data indicator to determine whether the current transmission should be soft combined with the received data currently in the soft buffer for the hybrid-ARQ process in question, or if the soft buffer should be cleared.

For uplink data transmission, there is also a new-data indicator transmitted on the downlink PDCCH. In this case, toggling the new-data indicator requests transmission of a new transport block, otherwise the previous transport block for this hybrid-ARQ process should be retransmitted (in which case the eNodeB should perform soft combining).

The use of multiple hybrid-ARQ processes operating in parallel can result in data being delivered from the hybrid-ARQ mechanism out of sequence. For example, transport block 5 in Figure 12.2 was successfully decoded before transport block 1, which required two retransmissions. Out-of-sequence delivery can also occur in the case of carrier aggregation, where transmission of a transport block on one component carrier could be successful while a retransmission is required on another component carrier. To handle out-of-sequence delivery from the hybrid-ARQ protocol, the RLC protocol includes an in-sequence-delivery mechanism, as described in Section 12.2.

Hybrid-ARQ protocols can be characterized as synchronous vs. asynchronous, related to the flexibility in the time domain, as well as adaptive vs. non-adaptive, related to the flexibility in the frequency domain:

- An *asynchronous* hybrid-ARQ protocol implies that retransmissions can occur at any time, whereas a *synchronous* protocol implies that retransmissions occur at a fixed time after the previous transmission (see Figure 12.3). The benefit of a synchronous protocol is that there is no need to explicitly signal the hybrid-ARQ process number as this information can be derived from the subframe number. On the other hand, an asynchronous protocol allows for more flexibility in the scheduling of retransmissions.
- An *adaptive* hybrid-ARQ protocol implies that the frequency location and possibly also the more detailed transmission format can be changed between retransmissions. A *non-adaptive protocol*, in contrast, implies that the retransmission must occur at the same frequency resources and with the same transmission format as the initial transmission.

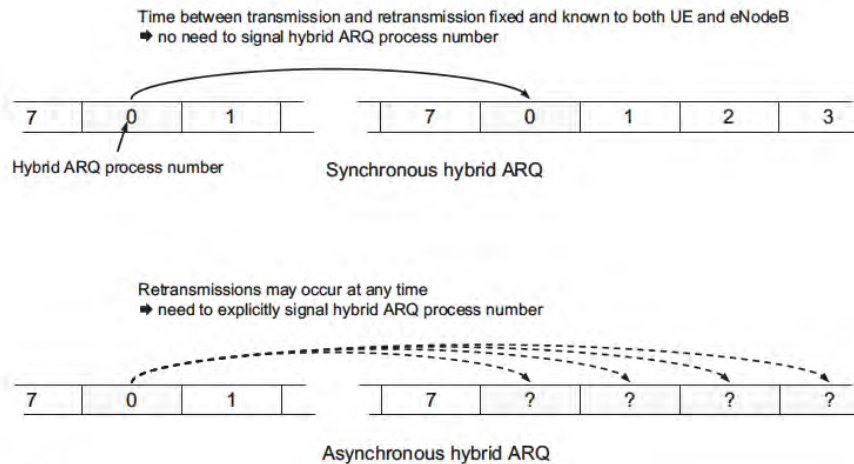


FIGURE 12.3

Synchronous and asynchronous hybrid ARQ

In the case of LTE, asynchronous adaptive hybrid ARQ is used for the downlink. For the uplink, synchronous hybrid ARQ is used. Typically, the retransmissions are non-adaptive, but there is also the possibility to use adaptive retransmissions as a complement.

12.1.1 Downlink hybrid ARQ

In the downlink, retransmissions are scheduled in the same way as new data—that is, they may occur at any time and at an arbitrary frequency location within the downlink cell bandwidth. Hence, the downlink hybrid-ARQ protocol is *asynchronous* and *adaptive*. The support for asynchronous and adaptive hybrid ARQ for the LTE downlink is motivated by the need to avoid collisions with, for example, transmission of system information and MBSFN subframes. Instead of dropping a retransmission that otherwise would collide with MBSFN subframes or transmission of system information, the eNodeB can move the retransmission in time and/or frequency to avoid the overlap in resources.

Support for soft combining is, as described in the introduction, provided through an explicit new-data indicator, toggled for each new transport block. In addition to the new-data indicator, hybrid-ARQ-related downlink control signaling consists of the hybrid-ARQ process number (three bits for FDD, four bits for TDD) and the redundancy version (two bits), both explicitly signaled in the scheduling assignment for each downlink transmission.

Downlink spatial multiplexing implies, as already mentioned, transmission of two transport blocks in parallel on a component carrier. To provide the possibility to retransmit

only one of the transport blocks, which is beneficial as error events for the two transport blocks can be fairly uncorrelated, each transport block has its own separate new-data indicator and redundancy-version indication. However, there is no need to signal the process number separately as each process consists of two subprocesses in the case of spatial multiplexing or, expressed differently, once the process number for the first transport block is known, the process number for the second transport block is given implicitly.

Transmissions on downlink component carriers are acknowledged independently. On each component carrier, transport blocks are acknowledged by transmitting one or two bits on the uplink, as described in Chapter 11. In the absence of spatial multiplexing, there is only a single transport block within a TTI and consequently only a single acknowledgement bit is required in response. However, if the downlink transmission used spatial multiplexing, there are two transport blocks per TTI, each requiring its own hybrid-ARQ acknowledgement bit. The total number of bits required for hybrid-ARQ acknowledgements thus depends on the number of component carriers and the transmission mode for each of the component carriers. As each downlink component carrier is scheduled separately from its own PDCCH, the hybrid-ARQ process numbers are signaled independently for each component carrier.

The terminal should not transmit a hybrid-ARQ acknowledgement in response to reception of system information, paging messages, and other broadcast traffic. Hence, hybrid-ARQ acknowledgements are only sent in the uplink for “normal” unicast transmission.

12.1.2 Uplink hybrid ARQ

Shifting the focus to the uplink, a difference compared to the downlink case is the use of synchronous, non-adaptive operation as the basic principle of the hybrid-ARQ protocol, motivated by the lower overhead compared to an asynchronous, adaptive structure. Hence, uplink retransmissions always occur at an *a priori* known subframe; in the case of FDD operation, uplink retransmissions occur eight subframes after the prior transmission attempt for the same hybrid-ARQ process. The set of resource blocks used for the retransmission on a component carrier is identical to the initial transmission. Thus, the only control signaling required in the downlink for a retransmission is a hybrid-ARQ acknowledgement, transmitted on the PHICH as described in Chapter 10. In the case of a negative acknowledgement on the PHICH, the data is retransmitted.

Spatial multiplexing is handled in the same way as in the downlink—two transport blocks transmitted in parallel on the uplink, each with its own modulation-and-coding scheme and new-data indicator but sharing the same hybrid-ARQ process number. The two transport blocks are individually acknowledged and hence two bits of information are needed in the downlink to acknowledge an uplink transmission using spatial multiplexing. Uplink carrier aggregation is another example of a situation where multiple acknowledgements are needed—one or two acknowledgement bits per uplink component carrier. TDD is a third example, as will be discussed further below, where uplink transmissions in different subframes may need to be acknowledged in the same downlink subframe. Thus, multiple

PHICHs may be needed as each PHICH is capable of transmitting a single bit only. Each of the PHICHs are transmitted on the same downlink component carrier that was used to schedule the initial uplink transmission.

Despite the fact that the basic mode of operation for the uplink is synchronous, non-adaptive hybrid ARQ, there is also the possibility to operate the uplink hybrid ARQ in a synchronous, *adaptive* manner, where the resource-block set and modulation-and-coding scheme for the retransmissions are changed. Although non-adaptive retransmissions are typically used due to the very low overhead in terms of downlink control signaling, adaptive retransmissions are sometimes useful to avoid fragmenting the uplink frequency resource or to avoid collisions with random-access resources. This is illustrated in Figure 12.4. A terminal is scheduled for an initial transmission in subframe n ; a transmission that is not correctly received and consequently a retransmission is required in subframe $n + 8$ (assuming FDD; for TDD the timing depends on the downlink–uplink allocation, as discussed further below). With non-adaptive hybrid ARQ, the retransmissions occupy the same part of the uplink spectrum as the initial transmission. Hence, in this example the spectrum is fragmented, which limits the bandwidth available to another terminal (unless the other terminal is capable of multi-cluster transmission). In subframe $n + 16$, an example of an adaptive retransmission is found; to make room for another terminal to be granted a large part of the uplink spectrum, the retransmission is moved in the frequency domain. It should be noted that the uplink hybrid-ARQ protocol is still synchronous—that is, a retransmission should always occur eight subframes after the previous transmission.

The support for both adaptive and non-adaptive hybrid ARQ is realized by *not* flushing the transmission buffer when receiving a positive hybrid-ARQ acknowledgement on PHICH for a given hybrid-ARQ process. Instead, the actual control of whether data should be retransmitted or not is done by the new-data indicator included in the uplink scheduling grant sent on the PDCCH. The new-data indicator is toggled for each new transport block. If the new-data indicator is toggled, the terminal flushes the transmission buffer and transmits a new data packet. However, if the new-data indicator does not request transmission of a new transport block, the previous transport block is retransmitted. Hence, clearing of the transmission buffer is not handled by the PHICH but by the PDCCH as part of the uplink grant. The negative hybrid-ARQ acknowledgement on the PHICH could instead be seen as a single-bit scheduling grant for retransmissions, where the set of bits to transmit and all the resource

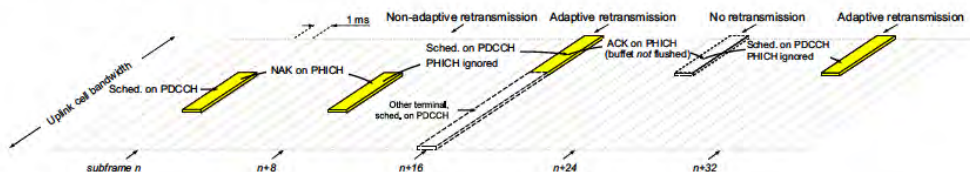


FIGURE 12.4

Non-adaptive and adaptive hybrid-ARQ operation

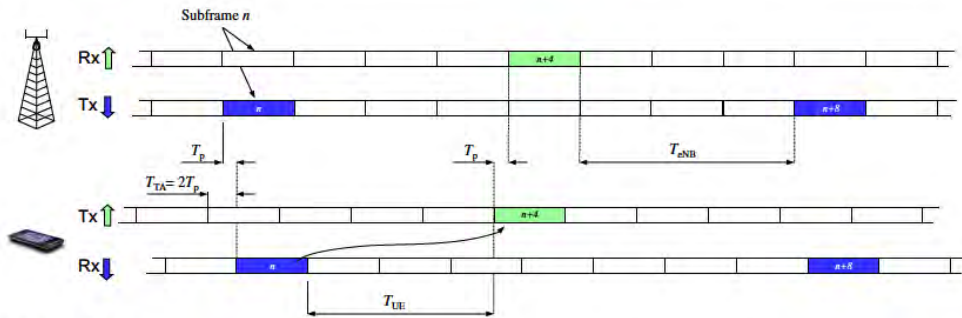


FIGURE 12.5

Timing relation between downlink data in subframe n and uplink hybrid-ARQ acknowledgment in subframe $n + 4$ for FDD

information are known from the previous transmission attempt. In Figure 12.4 an example of postponing a transmission is seen in subframe $n+24$. The terminal has received a positive acknowledgement and therefore does *not* retransmit the data. However, the transmission buffer is not flushed, which later is exploited by an uplink grant requesting retransmission in subframe $n + 32$.

A consequence of the above method of supporting both adaptive and non-adaptive hybrid ARQ is that the PHICH and PDCCH related to the same uplink subframe have the same timing. If this were not the case, the complexity would increase as the terminal would not know whether to obey the PHICH or wait for a PDCCH overriding the PHICH.

As explained earlier, the new-data indicator is explicitly transmitted in the uplink grant. However, unlike the downlink case, the redundancy version is *not* explicitly signaled for each retransmission. With a single-bit acknowledgement on the PHICH, this is not possible. Instead, as the uplink hybrid-ARQ protocol is synchronous, the redundancy version follows a predefined pattern, starting with zero when the initial transmission is scheduled by the PDCCH. Whenever a retransmission is requested by a negative acknowledgement on the PHICH, the next redundancy version in the sequence is used. However, if a retransmission is explicitly scheduled by the PDCCH overriding the PHICH, there is the potential to affect the redundancy version to use. Grants for retransmissions use the same format as ordinary grants (for initial transmissions). One of the information fields in an uplink grant is, as described in Chapter 10, the modulation-and-coding scheme. Of the 32 different combinations this five-bit field can take, three of them are reserved. Those three combinations represent different redundancy versions; hence, if one of these combinations is signaled as part of an uplink grant indicating a retransmission, the corresponding redundancy version is used for the transmission. The transport-block size is already known from the initial transmission as it, by definition, cannot change between retransmission attempts. In Figure 12.4, the initial

transmission in subframe n uses the first redundancy version in sequence as the transport-block size must be indicated for the initial transmission. The retransmission in subframe $n + 8$ uses the next redundancy version in the sequence, while the explicitly scheduled retransmission in subframe $n+16$ can use any redundancy scheme as indicated on the PDCCH.

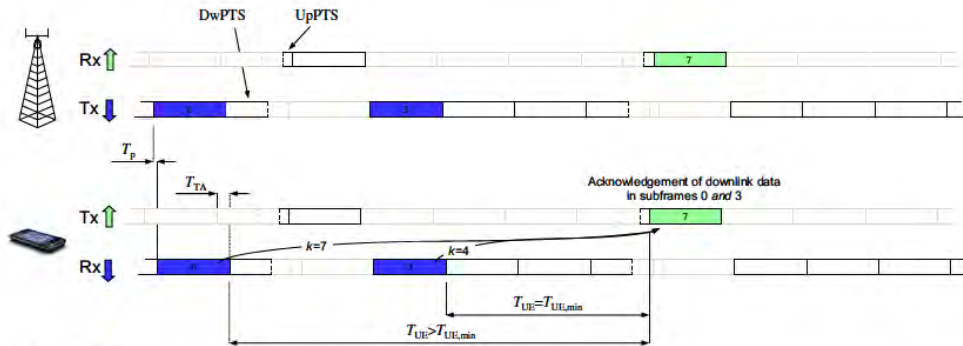
Whether to exploit the possibility to signal an arbitrary redundancy version in a retransmission scheduled by the PDCCH is a trade-off between incremental-redundancy gain and robustness. From an incremental-redundancy perspective, changing the redundancy value between retransmissions is typically beneficial to fully exploit the gain from incremental redundancy. However, as the modulation-and-coding scheme is normally not indicated for uplink retransmissions, as either the single-bit PHICH is used or the modulation-and-coding field is used to explicitly indicate a new redundancy version, it is implicitly assumed that the terminal did not miss the initial scheduling grant. If this is the case, it is necessary to explicitly indicate the modulation-and-coding scheme, which also implies that the first redundancy version in the sequence is used.

12.1.3 Hybrid-ARQ timing

The receiver must know to which hybrid-ARQ process a received acknowledgement is associated. This is handled by the timing of the hybrid-ARQ acknowledgement being used to associate the acknowledgement with a certain hybrid-ARQ process; the timing relation between the reception of data in the downlink and transmission of the hybrid-ARQ acknowledgement in the uplink (and vice versa) is fixed. From a latency perspective, the time between the reception of downlink data at the terminal and transmission of the hybrid-ARQ acknowledgement in the uplink should be as short as possible. At the same time, an unnecessarily short time would increase the demand on the terminal processing capacity and a trade-off between latency and implementation complexity is required. The situation is similar for uplink data transmissions. For LTE, this trade-off led to the decision to have eight hybrid-ARQ processes per component carrier in both uplink and downlink for FDD. For TDD, the number of processes depends on the downlink–uplink allocation, as discussed below.

12.1.3.1 Hybrid-ARQ timing for FDD

Starting with the FDD case, transmission of acknowledgements in the uplink in response to downlink data transmission is illustrated in Figure 12.5. Downlink data on the DL-SCH is transmitted to the terminal in subframe n and received by the terminal, after the propagation delay T_p , in subframe n . The terminal attempts to decode the received signal, possibly after soft combining with a previous transmission attempt, and transmits the hybrid-ARQ acknowledgement in uplink subframe $n + 4$. Note that the start of an uplink subframe at the terminal is offset by T_{TA} relative to the start of the corresponding downlink subframe at the terminal as a result of the timing-advance procedure described in Section 11.6. Upon reception of the hybrid-ARQ acknowledgement, the eNodeB can, if needed, retransmit the

**FIGURE 12.6**

Example of timing relation between downlink data and uplink hybrid-ARQ acknowledgment for TDD (configuration 2)

downlink data in subframe $n + 8$. Thus, eight hybrid-ARQ processes are used—the hybrid-ARQ round-trip time is 8 ms.

The description in the previous paragraph, as well as the illustration in Figure 12.5, describe the timing for downlink data transmission—that is, data on DL-SCH and acknowledgements on PUCCH (or PUSCH). However, the timing of uplink data transmission—data on PUSCH and acknowledgements on PHICH—is identical, namely uplink data transmission in subframe n results in a PHICH transmission in subframe $n + 4$. This is also the same timing relation as for uplink scheduling grants in general (as described in Section 13.3) and allows a scheduling grant on the PDCCH to override the PHICH, as shown in Figure 12.4.

In Figure 12.5 it is seen that the processing time available to the terminal, T_{UE} , depends on the value of the timing advance or, equivalently, on the terminal-to-base-station distance. As a terminal must operate at any distance up to the maximum size supported by the specifications, the terminal must be designed such that it can handle the worst-case scenario. LTE is designed to handle at least 100 km, corresponding to a maximum timing advance of 0.67 ms. Hence, there is approximately 2.3 ms left for the terminal processing, which is considered a reasonable trade-off between the processing requirements imposed on a terminal and the associated delays.

For the eNodeB, the processing time available, denoted as T_{eNB} , is 3 ms and thus of the same order as for the terminal. In the case of downlink data transmission, the eNodeB performs scheduling of any retransmissions during this time and, for uplink data transmission, the time is used for decoding of the received signal. The timing budget is similar for the terminal and the eNodeB, which is motivated by the fact that, although the eNodeB typically

has more processing power than a terminal, it also has to serve multiple terminals and to perform the scheduling operation.

Having the same number of hybrid-ARQ processes in uplink and downlink is beneficial for half-duplex FDD operation, discussed in Chapter 13. By proper scheduling, the uplink hybrid-ARQ transmission from the terminal will coincide with transmission of uplink data, and the acknowledgements related to the reception of uplink data will be transmitted in the same subframe as the downlink data. Thus, using the same number of hybrid-ARQ processes in uplink and downlink results in a 50:50 split between transmission and reception for a half-duplex terminal.

Carrier aggregation¹ in FDD implies that a terminal needs to transmit acknowledgements relating to simultaneous reception of data on multiple DL-SCHs. In LTE, there is at most one PUCCH and it is always transmitted on the primary component carrier. Hence, even for a symmetric carrier aggregation configuration, the possibility to transmit more than two hybrid-ARQ acknowledgement bits in the uplink must be supported. As discussed in Chapter 10, this is handled either by resource selection or by using PUCCH format 3. Acknowledgements in response to UL-SCH reception are transmitted using multiple PHICHs, where the PHICH is transmitted on the same downlink component carrier as the uplink scheduling grant initiating the uplink transmission.

12.1.3.2 Hybrid-ARQ timing for TDD

For TDD operation, the time relation between the reception of data in a certain hybrid-ARQ process and the transmission of the hybrid-ARQ acknowledgement depends on the downlink–uplink allocation. An uplink hybrid-ARQ acknowledgement can only be transmitted in an uplink subframe and a downlink acknowledgement only in a downlink subframe. Furthermore, starting from release 11, different downlink–uplink allocations can be used on component carriers operating in different frequency bands. Therefore, the hybrid-ARQ timing relations for TDD are more intricate than their FDD counterparts.

For simplicity, consider the situation when the same downlink–uplink allocation is used across all component carriers (the case of different allocations will be described later). Since the minimum processing time required in the terminal and the eNodeB remains the same in FDD and TDD, as the same turbo decoders are used and the scheduling decisions are similar, acknowledgements cannot be transmitted earlier than in FDD. Therefore, for TDD the acknowledgement of a transport block in subframe n is transmitted in subframe $n + k$, where $k \geq 4$ and is selected such that $n + k$ is an uplink subframe when the acknowledgement is to be transmitted from the terminal (on PUCCH or PUSCH) and a downlink subframe when the acknowledgement is transmitted from the eNodeB (on PHICH). The value of k depends on the downlink–uplink configuration, as shown in Table 12.1 for both downlink and uplink transmissions. From the table it is seen that, as a consequence of the downlink–uplink

¹In LTE release 10, the number of downlink component carriers was always at least as large as the number of uplink component carriers (see Chapter 9).

Table 12.1 Number of Hybrid-ARQ Processes and Acknowledgement Timing k for Different TDD Configurations																						
Configuration (DL:UL)	Downlink											Uplink										
	Proc.	PDSCH reception in subframe n										Proc.	PUSCH reception in subframe n									
		0	1	2	3	4	5	6	7	8	9		0	1	2	3	4	5	6	7	8	9
0 (2:3)	4	4	6	-	-	-	4	6	-	-	-	6	-	-	4	7	6	-	-	4	7	6
1 (3:2)	7	7	6	-	-	4	7	6	-	-	4	4	-	-	4	6	-	-	-	4	6	-
2 (4:1)	10	7	6	-	4	8	7	6	-	4	8	2	-	-	6	-	-	-	-	6	-	-
3 (7:3)	9	4	11	-	-	-	7	6	6	5	5	3	-	-	6	6	6	-	-	-	-	-
4 (8:2)	12	12	11	-	-	8	7	7	6	5	4	2	-	-	6	6	-	-	-	-	-	-
5 (9:1)	15	12	11	-	9	8	7	6	5	4	13	1	-	-	6	-	-	-	-	-	-	-
6 (5:5)	6	7	7	-	-	-	7	7	-	-	5	6	-	-	4	6	6	-	-	4	7	-

configuration, the value of k is sometimes larger than the FDD value $k = 4$. For example, assuming configuration 2, an uplink transmission on PUSCH received in subframe 2 should be acknowledged on PHICH in subframe $2 + 6 = 8$. Similarly, for the same configuration, downlink transmission on PDSCH in subframe 0 should be acknowledged on PUCCH (or PUSCH) in subframe $0 + 7 = 7$.

From the table it is also seen that the number of hybrid-ARQ processes used for TDD depends on the downlink–uplink configuration, implying that the hybrid-ARQ round-trip time is configuration dependent for TDD (actually, it may even vary between subframes, as seen in Figure 12.6). For the downlink-heavy configurations 2, 3, 4, and 5, the number of downlink hybrid-ARQ processes is larger than for FDD. The reason is the limited number of uplink subframes available, resulting in k -values well beyond 4 for some subframes (Table 12.1).

The PHICH timing in TDD is identical to the timing when receiving an uplink grant, as described in Section 13.3. The reason is the same as in FDD, namely to allow a PDCCH uplink grant to override the PHICH in order to implement adaptive retransmissions, as illustrated in Figure 12.4.

The downlink–uplink allocation for TDD has implications on the number of transport blocks to acknowledge in a single subframe. For FDD there is always a one-to-one relation between uplink and downlink subframes. Hence, in absence of carrier aggregation, a subframe only needs to carry acknowledgements for one subframe in the other direction. For TDD, in contrast, there is not necessarily a one-to-one relation between uplink and downlink subframes. This can be seen in the possible downlink–uplink allocations described in Chapter 9.

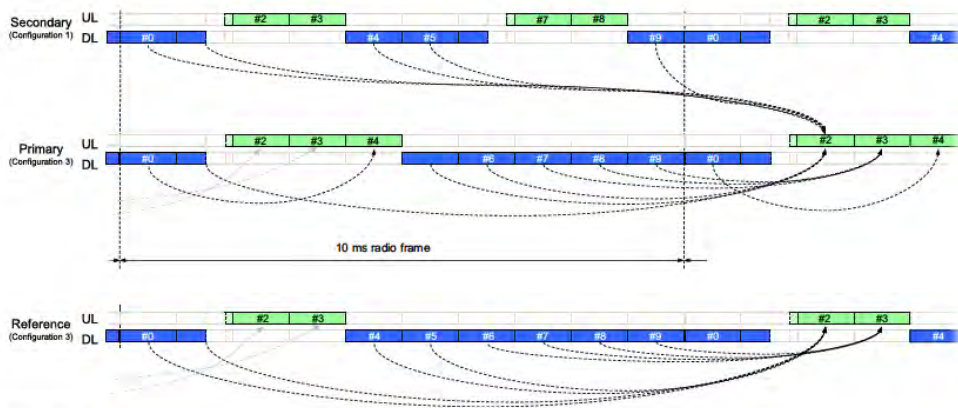


FIGURE 12.7

Aggregation of a primary component carrier with configuration 3 and a secondary component carrier with configuration 1

For uplink transmissions using UL-SCH, each uplink transport block is acknowledged individually by using the PHICH. Hence, in uplink-heavy asymmetries (downlink–uplink configuration 0), the terminal may need to receive two hybrid-ARQ acknowledgements in downlink subframes 0 and 5 in the absence of uplink spatial multiplexing; for spatial multiplexing up to four acknowledgements per component carrier are needed. There are also some subframes where no PHICH will be transmitted. The amount of PHICH groups may therefore vary between subframes for TDD.

For downlink transmissions, there are some configurations where DL-SCH receipt in multiple downlink subframes needs to be acknowledged in a single uplink subframe, as illustrated in [Figure 12.6](#) (the illustration corresponds to the two entries shown in bold in [Table 12.1](#)). Two different mechanisms to handle this are provided in TDD: multiplexing and bundling.

Multiplexing implies that independent acknowledgements for each of the received transport blocks are fed back to the eNodeB. This allows independent retransmission of erroneous transport blocks. However, it also implies that multiple bits need to be transmitted from the terminal, which may limit the uplink coverage. This is the motivation for the bundling mechanism.

Bundling of acknowledgements implies that the outcome of the decoding of downlink transport blocks from multiple downlink subframes can be combined into a single hybrid-ARQ acknowledgement transmitted in the uplink. Only if both of the downlink transmissions in subframes 0 and 3 in the example in [Figure 12.6](#) are correctly decoded will a positive acknowledgement be transmitted in uplink subframe 7.

Combining acknowledgements related to multiple downlink transmissions into a single uplink message assumes that the terminal has not missed any of the scheduling assignments upon which the acknowledgement is based. Assume, as an example, that the eNodeB scheduled the terminal in two (subsequent) subframes, but the terminal missed the PDCCH transmission in the first of the two subframes and successfully decoded the data transmitted in the second subframe. Without any additional mechanism, the terminal will transmit an acknowledgement based on the assumption that it was scheduled in the second subframe only, while the eNodeB will interpret acknowledgement as the terminal successfully received both transmissions. To avoid such errors, the *downlink assignment index* (see [Section 10.4.6](#)) in the scheduling assignment is used. The downlink assignment index in essence informs the terminal about the number of transmissions it should base the combined acknowledgement upon. If there is a mismatch between the assignment index and the number of transmissions the terminal received, the terminal concludes at least one assignment was missed and transmits no hybrid-ARQ acknowledgement, thereby avoiding acknowledging transmissions not received.

Similarly to FDD, carrier aggregation implies that the terminal may need to transmit acknowledgements relating to multiple DL-SCHs in a single uplink subframe. Hybrid ARQ acknowledgements in response to downlink PDSCH reception are transmitted on the PUCCH on the primary component carrier. An acknowledgement can only be transmitted in an uplink subframe on the primary component carrier, and the same mechanisms as used for

transmitting multiple acknowledgements in absence of a carrier, multiplexing or bundling, are used.

In case of identical downlink-uplink allocations across all component carriers, the timing of the acknowledgements, both in uplink and downlink, are the same as in absence of carrier aggregation. However, release 11 introduced the possibility of having *different* allocations on different component carriers which complicates the timing relations further.

In absence of cross-carrier scheduling, the scheduling grants are transmitted on the same component carrier as the corresponding data. As there is no dependency between the scheduling decisions on the different component carriers, downlink transmissions can occur on a secondary component carrier in a subframe that is reserved for uplink transmission on the primary component carrier. If the hybrid ARQ timing would follow the primary component carrier only, there would be no timing relation for the secondary component carrier in some subframes (subframe 4 in Figure 12.7) and, consequently, those subframes would not be possible to use for downlink data transmissions. Therefore, to handle this case, a reference configuration, compatible with the hybrid ARQ timing of both the primary component carrier and all secondary component carriers, is used to derive the hybrid ARQ timing relations for the secondary component carriers. The primary component carrier uses the same timing as in the absence of carrier aggregation. See Figure 12.7 for an example aggregation of primary and secondary component carriers using configurations 3 and 1, respectively, in which case configuration 4 is used as a reference for the timing of any secondary component carrier. The reference configuration for other combinations of configurations can be found in Table 12.2.

In case of cross-carrier scheduling, no reference configuration is needed as the scheduling timing for each component carrier follows the primary component carrier as discussed in Section 13.3. The “problematic” subframe, subframe 4 in Figure 12.7, can therefore never be scheduled, and the hybrid ARQ timing relations can be derived from the primary component carrier alone.

		Secondary Component Carrier						
		0	1	2	3	4	5	6
Primary component carrier	0	0	1	2	3	4	5	6
	1	1	1	2	4	4	5	1
	2	2	2	2	5	5	5	2
	3	3	4	5	3	4	5	3
	4	4	4	5	4	4	5	4
	5	5	5	5	5	5	5	5
	6	6	1	2	3	4	5	6

The timing of the PHICH carrying hybrid ARQ acknowledgements in response to uplink data transmission is also affected by the downlink-uplink configurations. In absence of cross-carrier scheduling, the PHICH timing on each component carrier follows the same principles as in absence of carrier aggregation. In presence of cross-carrier scheduling, the situation is more complicated. For some configurations the PHICH timing is given by the configuration of the scheduling carrier—the carrier on which the (E)PDCCH was transmitted—while for other configurations, the PHICH timing follows the configuration of the scheduled carrier—that is, the carrier upon which the PUSCH was transmitted.

12.2 Radio-link control

The *radio-link control* (RLC) protocol takes data in the form of RLC SDUs from PDCP and delivers them to the corresponding RLC entity in the receiver by using functionality in MAC and physical layers. The relation between RLC and MAC, including multiplexing of multiple logical channels into a single transport channel, is illustrated in Figure 12.8. Multiplexing of several logical channels into a single transport channel is mainly used for priority handling, as described in Section 13.2 in conjunction with downlink and uplink scheduling.

There is one RLC entity per logical channel configured for a terminal, where each RLC entity is responsible for:

- segmentation, concatenation, and reassembly of RLC SDUs;
- RLC retransmission; and
- in-sequence delivery and duplicate detection for the corresponding logical channel.

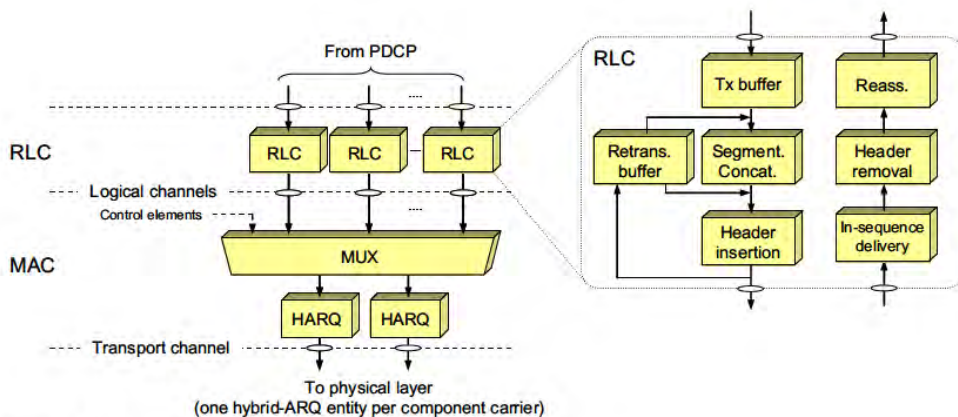


FIGURE 12.8

MAC and RLC structure (single-terminal view)

Other noteworthy features of the RLC are: (1) the handling of varying PDU sizes; and (2) the possibility for close interaction between the hybrid-ARQ and RLC protocols. Finally, the fact that there is one RLC entity per logical channel and one hybrid-ARQ entity per component carrier implies that one RLC entity may interact with multiple hybrid-ARQ entities in case of carrier aggregation.

12.2.1 Segmentation and concatenation, and reassembly of RLC SDUs

The purpose of the segmentation and concatenation mechanism is to generate RLC PDUs of appropriate size from the incoming RLC SDUs. One possibility would be to define a fixed PDU size; a size that would result in a compromise. If the size was too large, it would not be possible to support the lowest data rates. Also, excessive padding would be required in some scenarios. A single small PDU size, however, would result in a high overhead from the header included with each PDU. To avoid these drawbacks, which is especially important given the very large dynamic range of data rates supported by LTE, the RLC PDU size varies dynamically.

Segmentation and concatenation of RLC SDUs into RLC PDUs are illustrated in Figure 12.9. The header includes, among other fields, a sequence number which is used by the reordering and retransmission mechanisms. The reassembly function at the receiver side performs the reverse operation to reassemble the SDUs from the received PDUs.

12.2.2 RLC retransmission

Retransmission of missing PDUs is one of the main functionalities of the RLC. Although most of the errors can be handled by the hybrid-ARQ protocol, there are, as discussed at the beginning of the chapter, benefits of having a second-level retransmission mechanism as a complement. By inspecting the sequence numbers of the received PDUs, missing PDUs can be detected and a retransmission requested from the transmitting side.

Different services have different requirements; for some services (for example, transfer of a large file), error-free delivery of data is important, whereas for other applications (for example, streaming services), a small amount of missing packets is not a problem. The RLC

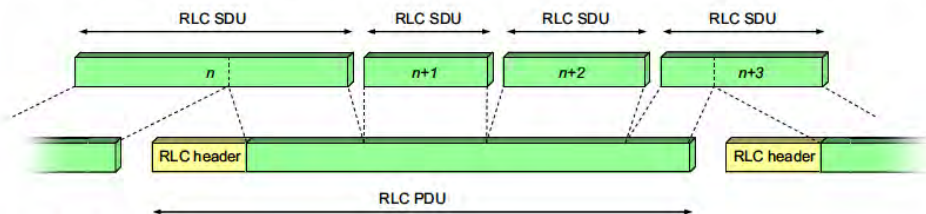


FIGURE 12.9

Generation of RLC PDUs from RLC SDUs

can therefore operate in three different modes, depending on the requirements from the application:

- *Transparent mode* (TM), where the RLC is completely transparent and is essentially bypassed. No retransmissions, no segmentation/reassembly, and no in-sequence delivery take place. This configuration is used for control-plane broadcast channels such as BCCH, CCCH, and PCCH, where the information should reach multiple users. The size of these messages are selected such that all intended terminals are reached with a high probability and hence there is neither need for segmentation to handle varying channel conditions, nor retransmissions to provide error-free data transmission. Furthermore, retransmissions are not feasible for these channels as there is no possibility for the terminal to feed back status reports as no uplink has been established.
- *Unacknowledged mode* (UM) supports segmentation/reassembly and in-sequence delivery, but not retransmissions. This mode is used when error-free delivery is not required, for example voice-over IP, or when retransmissions cannot be requested, for example broadcast transmissions on MTCH and MCCH using MBSFN.
- *Acknowledged mode* (AM) is the main mode of operation for TCP/IP packet data transmission on the DL-SCH. Segmentation/reassembly, in-sequence delivery, and retransmissions of erroneous data are all supported.

In the following, the acknowledged mode is described. The unacknowledged mode is similar, with the exception that no retransmissions are done and each RLC entity is unidirectional.

In acknowledged mode, the RLC entity is bidirectional—that is, data may flow in both directions between the two peer entities. This is necessary as the reception of PDUs needs to be acknowledged back to the entity that transmitted those PDUs. Information about missing PDUs is provided by the receiving end to the transmitting end in the form of so-called *status reports*. Status reports can either be transmitted autonomously by the receiver or requested by the transmitter. To keep track of the PDUs in transit, the transmitter attaches an RLC header to each PDU, including, among other fields, a sequence number.

Both RLC entities maintain two windows—the transmission and reception windows respectively. Only PDUs in the transmission window are eligible for transmission; PDUs with a sequence number below the start of the window have already been acknowledged by the receiving RLC. Similarly, the receiver only accepts PDUs with sequence numbers within the reception window. The receiver also discards any duplicate PDUs as each PDU should be assembled into an SDU only once.

12.2.3 In-sequence delivery

In-sequence delivery implies that data blocks are delivered by the receiver in the same order as they were transmitted. This is an essential part of RLC; the hybrid-ARQ processes operate independently and transport blocks may therefore be delivered out of sequence, as seen in

Figure 12.2. In-sequence delivery implies that SDU n should be delivered prior to SDU $n + 1$. This is an important aspect as several applications require the data to be received in the same order as it was transmitted. TCP can, to some extent, handle IP packets arriving out of sequence, although with some performance impact, while for some streaming applications in-sequence delivery is essential. The basic idea behind in-sequence delivery is to store the received PDUs in a buffer until all PDUs with lower sequence numbers have been delivered. Only when all PDUs with lower sequence numbers have been used for assembling SDUs is the next PDU used. RLC retransmission, provided when operating in acknowledged mode only, operates on the same buffer as the in-sequence delivery mechanism.

12.2.4 RLC operation

The operation of the RLC with respect to retransmissions and in-sequence delivery is perhaps best understood by the simple example in **Figure 12.10**, where two RLC entities are illustrated, one in the transmitting node and one in the receiving node. When operating in acknowledged mode, as assumed below, each RLC entity has both transmitter and receiver functionality, but in this example only one of the directions is discussed as the other direction is identical. In the example, PDUs numbered from n to $n + 4$ are awaiting transmission in the transmission buffer. At time t_0 , PDUs with a sequence number up to and including n have been transmitted and correctly received, but only PDUs up to and including $n - 1$ have been acknowledged by the receiver. As seen in the figure, the transmission window starts from n , the first not-yet-acknowledged PDU, while the reception window starts from $n + 1$, the next PDU expected to be received. Upon reception of PDU n , the PDU is forwarded to the SDU reassembly functionality for further processing.

The transmission of PDUs continues and, at time t_1 , PDUs $n + 1$ and $n + 2$ have been transmitted but, at the receiving end, only PDU $n + 2$ has arrived. One reason for this could be that the missing PDU, $n + 1$, is being retransmitted by the hybrid-ARQ protocol and therefore has not yet been delivered from the hybrid ARQ to the RLC. The transmission window

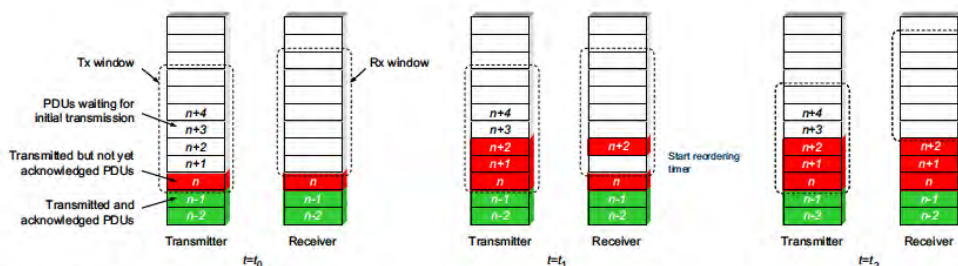


FIGURE 12.10

In-sequence delivery

remains unchanged compared to the previous figure, as none of the PDUs n and higher have been acknowledged by the receiver. Hence, any of these PDUs may need to be retransmitted as the transmitter is not aware of whether they have been received correctly or not. The reception window is not updated when PDU $n + 2$ arrives. The reason is that PDU $n + 2$ cannot be forwarded to SDU assembly as PDU $n + 1$ is missing. Instead, the receiver waits for the missing PDU $n + 1$. Waiting for the missing PDU for an infinite time would stall the queue. Hence, the receiver starts a timer, the *reordering timer*, for the missing PDU. If the PDU is not received before the timer expires, a retransmission is requested. Fortunately, in this example, the missing PDU arrives from the hybrid-ARQ protocol at time t_2 , before the timer expires. The reception window is advanced and the reordering timer is stopped as the missing PDU has arrived. PDUs $n + 1$ and $n + 2$ are delivered for reassembly into SDUs.

Duplicate detection is also the responsibility of the RLC, using the same sequence number as used for reordering. If PDU $n + 2$ arrives again (and is within the reception window), despite it having already been received, it is discarded.

The example above illustrates the basic principle behind in-sequence delivery, which is supported by both acknowledged and unacknowledged modes. However, the acknowledged mode of operation also provides retransmission functionality. To illustrate the principles behind this, consider Figure 12.11, which is a continuation of the example above. At time t_3 , PDUs up to $n + 5$ have been transmitted. Only PDU $n + 5$ has arrived and PDUs $n + 3$ and $n + 4$ are missing. Similar to the case above, this causes the reordering timer to start. However, in this example no PDUs arrive prior to the expiration of the timer. The expiration of the timer at time t_4 triggers the receiver to send a control PDU containing a status report, indicating the missing PDUs, to its peer entity. Control PDUs have higher priority than data PDUs to avoid the status reports being unnecessarily delayed and negatively impacting the retransmission delay. Upon reception of the status report at time t_5 , the transmitter knows that PDUs up to $n + 2$ have been received correctly and the transmission window is advanced. The missing PDUs $n + 3$ and $n + 4$ are retransmitted and, this time, correctly received.

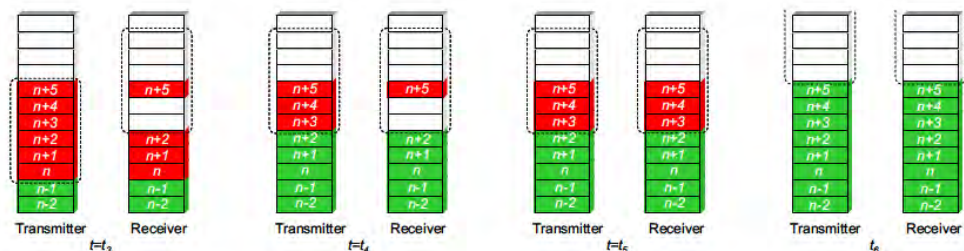


FIGURE 12.11

Retransmission of missing PDUs

The retransmission was triggered by the reception of a status report in this example. However, as the hybrid-ARQ and RLC protocols are located in the same node, tight interaction between the two is possible. The hybrid-ARQ protocol at the transmitting end could therefore inform the RLC at the transmitting end in case the transport block(s) containing PDUs $n + 3$ and $n + 4$ have failed. The RLC can use this to trigger retransmission of missing PDUs without waiting for an explicit RLC status report, thereby reducing the delays associated with RLC retransmissions.

Finally, at time t_6 , all PDUs, including the retransmissions, have been delivered by the transmitter and successfully received. As $n + 5$ was the last PDU in the transmission buffer, the transmitter requests a status report from the receiver by setting a flag in the header of the last RLC data PDU. Upon reception of the PDU with the flag set, the receiver will respond by transmitting the requested status report, acknowledging all PDUs up to and including $n + 5$. Reception of the status report by the transmitter causes all the PDUs to be declared as correctly received and the transmission window is advanced.

Status reports can, as mentioned earlier, be triggered for multiple reasons. However, to control the amount of status reports and to avoid flooding the return link with an excessive number of status reports, it is possible to use a status prohibit timer. With such a timer, status reports cannot be transmitted more often than once per time interval as determined by the timer.

For the initial transmission, it is relatively straightforward to rely on a dynamic PDU size as a means to handle the varying data rates. However, the channel conditions and the amount of resources may also change between RLC retransmissions. To handle these variations, already transmitted PDUs can be (re)segmented for retransmissions. The reordering and retransmission mechanisms described above still apply; a PDU is assumed to be received when all the segments have been received. Status reports and retransmissions operate on individual segments; only the missing segment of a PDU needs to be retransmitted.

This page intentionally left blank

Scheduling and Rate Adaptation

13

The scheduler is a key element and to a large degree determines the overall behavior of the system. Both uplink and downlink transmissions are scheduled and, consequently, there is a downlink and an uplink scheduler in the eNodeB.

The *downlink scheduler* is responsible for dynamically controlling the terminal(s) to transmit to. Each of the scheduled terminals is provided with a *scheduling assignment* consisting of the set of resource blocks upon which the terminal's DL-SCH¹ is transmitted, and the associated transport-format. The basic mode of operation is so-called *dynamic* scheduling, where the eNodeB for each 1 ms TTI conveys scheduling assignments to the selected terminals using the (E)PDCCHs as described in Chapter 10, but there is also a possibility for *semi-persistent* scheduling to reduce the control-signaling overhead. Downlink scheduling assignments and logical channel multiplexing are controlled by the eNodeB, as illustrated in the left part of Figure 13.1.

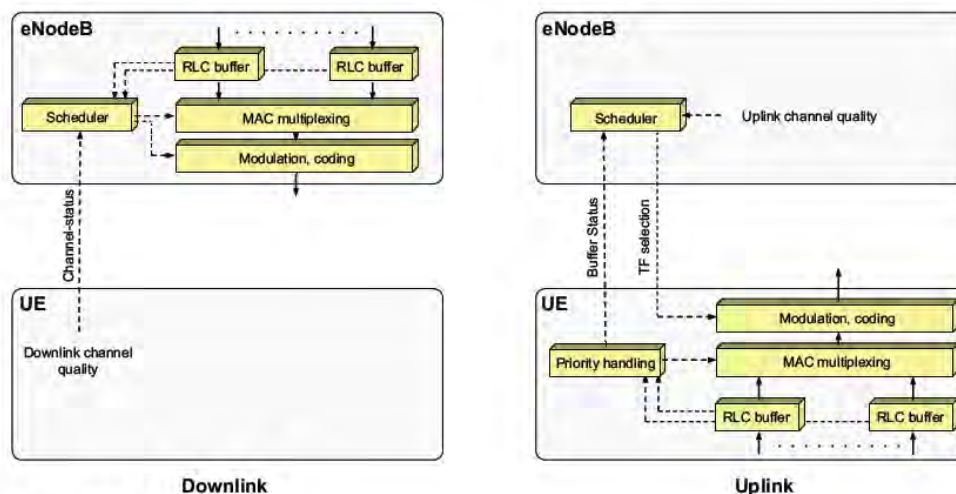


FIGURE 13.1

Transport format selection in downlink (left) and uplink (right)

¹In case of carrier aggregation there is one DL-SCH (or UL-SCH) per component carrier.

The *uplink scheduler* serves a similar purpose, namely to dynamically control which terminals are to transmit on their UL-SCH. Similarly to the downlink case, each scheduled terminal is provided with a *scheduling grant* consisting of the set of resource blocks upon which the terminal should transmit its UL-SCH and the associated transport-format. Also in this case, either dynamic or semi-persistent scheduling can be used. The uplink scheduler is in complete control of the transport format the terminal shall use but, unlike the downlink case, not the logical-channel multiplexing. Instead, the logical-channel multiplexing is controlled by the terminal according to a set of rules. Thus, uplink scheduling is *per terminal* and not per radio bearer. This is illustrated in the right part of Figure 13.1, where the scheduler controls the transport format and the terminal controls the logical channel multiplexing.

Each component carrier is independently scheduled with individual scheduling assignments/grants per component carrier in case of carrier aggregation. Semi-persistent scheduling is only supported on the primary component carriers, motivated by the fact that the main usage is for small payloads not requiring multiple component carriers.

In the following sections, dynamic downlink and uplink scheduling will be described, as well as related functionality such as uplink priority handling, scheduling requests, buffer status and power headroom reporting, semi-persistent scheduling, half-duplex FDD operation, DRX functionality, and reporting of channel-state information.

13.1 Downlink scheduling

The task of the downlink scheduler is to dynamically determine the terminal(s) to transmit to and, for each of these terminals, the set of resource blocks upon which the terminal's DL-SCH should be transmitted. In most cases, a single terminal cannot use the full capacity of the cell, for example due to lack of data. Also, as the channel properties may vary in the frequency domain, it is useful to be able to transmit to different terminals on different parts of the spectrum. Therefore, multiple terminals can be scheduled in parallel in a subframe, in which case there is one DL-SCH per scheduled terminal and component carrier, each dynamically mapped to a (unique) set of frequency resources.

The scheduler is in control of the instantaneous data rate used, and the RLC segmentation and MAC multiplexing will therefore be affected by the scheduling decision. Although formally part of the MAC layer, but to some extent better viewed as a separate entity, the scheduler is thus controlling most of the functions in the eNodeB associated with downlink data transmission:

- *RLC*: Segmentation/concatenation of RLC SDUs is directly related to the instantaneous data rate. For low data rates, it may only be possible to deliver a part of an RLC SDU in a TTI, in which case segmentation is needed. Similarly, for high data rates, multiple RLC SDUs may need to be concatenated to form a sufficiently large transport block.

- *MAC*: Multiplexing of logical channels depends on the priorities between different streams. For example, radio resource control signaling, such as handover commands, typically has a higher priority than streaming data, which in turn has higher priority than a background file transfer. Thus, depending on the data rate and the amount of traffic of different priorities, the multiplexing of different logical channels is affected. Hybrid-ARQ retransmissions also need to be accounted for.
- *LI*: Coding, modulation and, if applicable, the number of transmission layers and the associated precoding matrix are affected by the scheduling decision. The choices of these parameters are mainly determined by the radio conditions and the selected data rate—that is, the transport block size.

The scheduling strategy is implementation-specific and not part of the 3GPP specifications. In principle, any of the schedulers described in Chapter 6 can be applied. However, the overall goal of most schedulers is to take advantage of the channel variations between terminals and preferably to schedule transmissions to a terminal when the channel conditions are advantageous. Most scheduling strategies therefore at least need information about:

- Channel conditions at the terminal.
- Buffer status and priorities of the different data flows.
- The interference situation in neighboring cells (if some form of interference coordination is implemented).

Information about the channel conditions at the terminal can be obtained in several ways. In principle, the eNodeB can use any information available, but typically the CSI reports from the terminal, further described in [Section 13.7](#), are used. However, additional sources of channel knowledge, for example exploiting channel reciprocity to estimate the downlink quality from uplink channel estimates in the case of TDD, can also be exploited by a particular scheduler implementation.

In addition to the channel-state information, the scheduler should take buffer status and priority levels into account. It does not make sense to schedule a terminal with empty transmission buffers. Priorities of the different types of traffic may also vary; RRC signaling may be prioritized over user data. Furthermore, RLC and hybrid-ARQ retransmissions, which are in no way different from other types of data from a scheduler perspective, are typically also given priority over initial transmissions.

Downlink inter-cell interference coordination is also part of the implementation-specific scheduler strategy. A cell may signal to its neighboring cells the intention to transmit with a lower transmission power in the downlink on a set of resource blocks. This information can then be exploited by neighboring cells as a region of low interference where it is advantageous to schedule terminals at the cell edge—terminals that otherwise could not attain high data rates due to the interference level. Inter-cell interference handling is further discussed in Chapter 15.

13.2 Uplink scheduling

The basic function of the *uplink scheduler* is similar to its downlink counterpart, namely to dynamically determine, for each 1 ms interval, which terminals are to transmit and on which uplink resources. As discussed before, the LTE uplink is primarily based on maintaining orthogonality between different uplink transmissions and the shared resource controlled by the eNodeB scheduler is time–frequency resource units. In addition to assigning the time–frequency resources to the terminal, the eNodeB scheduler is also responsible for controlling the transport format the terminal will use for each of the uplink component carriers. This allows the scheduler to tightly control the uplink activity to maximize the resource usage compared to schemes where the terminal autonomously selects the data rate, as autonomous schemes typically require some margin in the scheduling decisions. A consequence of the scheduler being responsible for selection of the transport format is that accurate and detailed knowledge about the terminal situation with respect to buffer status and power availability is more accentuated in LTE, compared to systems where the terminal autonomously controls the transmission parameters.

The basis for uplink scheduling is *scheduling grants*, containing the scheduling decision and providing the terminal information about the resources and the associated transport format to use for transmission of the UL-SCH on one component carrier. Only if the terminal has a valid grant is it allowed to transmit on the corresponding UL-SCH; autonomous transmissions are not possible. Dynamic grants are valid for one subframe—that is, for each subframe in which the terminal is to transmit on the UL-SCH, the scheduler issues a new grant. Uplink component carriers are scheduled independently; if the terminal is to transmit simultaneously on multiple component carriers, multiple scheduling grants are needed.

Similarly to the downlink case, the uplink scheduler can exploit information about channel conditions, and, if some form of interference coordination is employed, the interference situation in neighboring cells. Information about the buffer status in the terminal and its available transmission power is also beneficial to the scheduler. This calls for the reporting mechanisms described below, unlike the downlink case where the scheduler, power amplifier, and transmission buffers all are in the same node. Uplink priority handling is, as already touched upon, another area where uplink and downlink scheduling differs.

Channel-dependent scheduling, which typically is used for the downlink, can be used for the uplink as well. In the uplink, estimates of the channel quality can be obtained from the use of uplink channel sounding, as described in Chapter 11. For scenarios where the overhead from channel sounding is too costly, or when the variations in the channel are too rapid to be tracked (for example at high terminal speeds), uplink diversity can be used instead. The use of frequency hopping as discussed in Chapter 11 is one example of obtaining diversity in the uplink.

Finally, inter-cell interference coordination can be used in the uplink for similar reasons as in the downlink by exchanging information between neighboring cells, as discussed in Chapter 15.

13.2.1 Uplink priority handling

Multiple logical channels of different priorities can be multiplexed into the same transport block using the same MAC multiplexing functionality as in the downlink (described in Chapter 8). However, unlike the downlink case, where the prioritization is under control of the scheduler and up to the implementation, the uplink multiplexing is done according to a set of well-defined rules in the terminal as a scheduling grant applies to a specific uplink carrier of a *terminal*, not to a specific radio bearer within the terminal. Using radio-bearer-specific scheduling grants would increase the control signaling overhead in the downlink and hence per-terminal scheduling is used in LTE.

The simplest multiplexing rule would be to serve logical channels in strict priority order. However, this may result in starvation of lower-priority channels; all resources would be given to the high-priority channel until its transmission buffer is empty. Typically, an operator would instead like to provide at least some throughput for low-priority services as well. Therefore, for each logical channel in an LTE terminal, a *prioritized data rate* is configured in addition to the priority value. The logical channels are then served in decreasing priority order up to their prioritized data rate, which avoids starvation as long as the scheduled data rate is at least as large as the sum of the prioritized data rates. Beyond the prioritized data rates, channels are served in strict priority order until the grant is fully exploited or the buffer is empty. This is illustrated in Figure 13.2.

13.2.2 Scheduling requests

The scheduler needs knowledge about terminals having data to transmit and therefore needs to be scheduled uplink resources. There is no need to provide uplink resources to a terminal with no data to transmit as this would only result in the terminal performing padding to fill up the granted resources. Hence, as a minimum, the scheduler needs to know whether the terminal has data to transmit and should be given a grant. This is known as a *scheduling request*. Scheduling requests are used for terminals not having a valid scheduling grant; terminals that have a valid grant and are transmitting in the uplink provide more detailed information to the eNodeB, as discussed in the next section.

A scheduling request is a simple flag raised by the terminal to request uplink resources from the uplink scheduler. Since the terminal requesting resources by definition has no

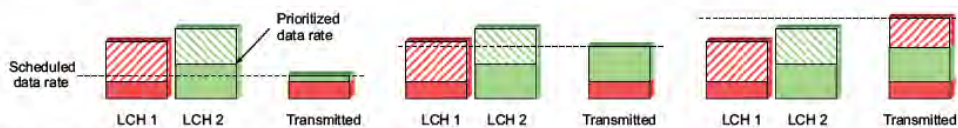
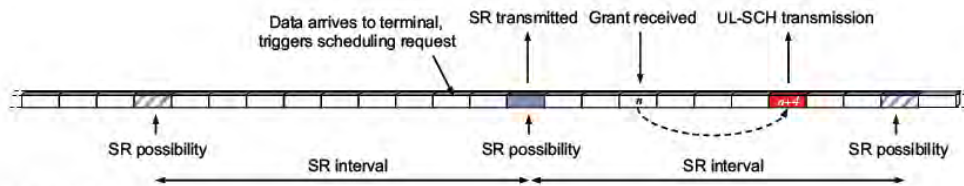


FIGURE 13.2

Prioritization of two logical channels for three different uplink grants

**FIGURE 13.3**

Scheduling request transmission

PUSCH resource, the scheduling request is transmitted on the PUCCH. Each terminal can be assigned a dedicated PUCCH scheduling request resource, occurring every n th subframe, as described in Chapter 11. With a dedicated scheduling-request mechanism, there is no need to provide the identity of the terminal requesting to be scheduled as the identity of the terminal is implicitly known from the resources upon which the request is transmitted. When data with higher priority than already existing in the transmit buffers arrives at the terminal and the terminal has no grant, and hence cannot transmit the data, the terminal transmits a scheduling request at the next possible instant, as illustrated in Figure 13.3. Upon reception of the request, the scheduler can assign a grant to the terminal. If the terminal does not receive a scheduling grant until the next possible scheduling-request instant, then the scheduling request is repeated. In the case of carrier aggregation, the scheduling request is transmitted on the primary component carrier, in line with the general principle of PUCCH transmission on the primary component carrier only. There is only a single scheduling-request bit, irrespective of the number of uplink component carriers the terminal is capable of.

The use of a single bit for the scheduling request is motivated by the desire to keep the uplink overhead small, as a multi-bit scheduling request would come at a higher cost. A consequence of the single-bit scheduling request is the limited knowledge at the eNodeB about the buffer situation at the terminal when receiving such a request. Different scheduler implementations handle this differently. One possibility is to assign a small amount of resources to ensure that the terminal can exploit them efficiently without becoming power limited. Once the terminal has started to transmit on the UL-SCH, more detailed information about the buffer status and power headroom can be provided through the inband MAC control message, as discussed below. Knowledge of the service type may also be used—for example, in the case of voice, the uplink resource to grant is preferably the size of a typical voice-over-IP package. The scheduler may also exploit, for example, path-loss measurements used for mobility and handover decisions to estimate the amount of resources the terminal may efficiently utilize.

An alternative to a dedicated scheduling-request mechanism would be a contention-based design. In such a design, multiple terminals share a common resource and provide their identity as part of the request. This is similar to the design of the random access. The number of bits transmitted from a terminal as part of a request would in this case be larger, with the

correspondingly larger need for resources. In contrast, the resources are shared by multiple users. Basically, contention-based designs are suitable for a situation where there are a large number of terminals in the cell and the traffic intensity, and hence the scheduling intensity, is low. In situations with higher intensities, the collision rate between different terminals simultaneously requesting resources would be too high and lead to an inefficient design.

Since the scheduling-request design for LTE relies on dedicated resources, a terminal that has not been allocated such resources cannot transmit a scheduling request. Instead, terminals without scheduling-request resources configured rely on the random-access mechanism described in Chapter 14. In principle, an LTE terminal can therefore be configured to rely on a contention-based mechanism if this is advantageous in a specific deployment.

13.2.3 Buffer status reports

Terminals that already have a valid grant do not need to request uplink resources. However, to allow the scheduler to determine the amount of resources to grant to each terminal in future subframes, information about the buffer situation, discussed in this section, and the power availability, discussed in the next section, is useful. This information is provided to the scheduler as part of the uplink transmission through MAC control elements (see Chapter 8 for a discussion on MAC control elements and the general structure of a MAC header). The LCID field in one of the MAC subheaders is set to a reserved value indicating the presence of a buffer status report, as illustrated in Figure 13.4.

From a scheduling perspective, buffer information for each logical channel is beneficial, although this could result in a significant overhead. Logical channels are therefore grouped into logical-channel groups and the reporting is done per group. The buffer-size field in a buffer-status report indicates the amount of data awaiting transmission across all logical

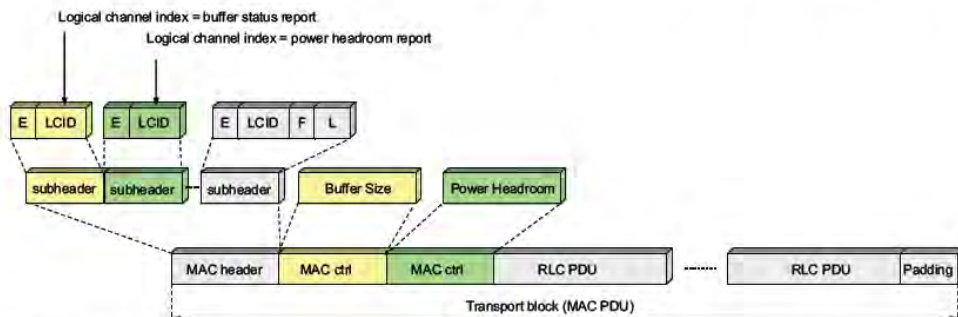


FIGURE 13.4

Signaling of buffer status and power headroom reports

channels in a logical-channel group. A buffer-status report can be triggered for the following reasons:

- Arrival of data with higher priority than currently in the transmission buffer—that is, data in a logical-channel group with higher priority than the one currently being transmitted—as this may impact the scheduling decision.
- Change of serving cell, in which case a buffer-status report is useful to provide the new serving cell with information about the situation in the terminal.
- Periodically as controlled by a timer.
- Instead of padding. If the amount of padding required to match the scheduled transport block size is larger than a buffer-status report, a buffer-status report is inserted. It is better to exploit the available payload for useful scheduling information instead of padding if possible.

13.2.4 Power headroom reports

In addition to buffer status, the amount of transmission power available in each terminal is also relevant for the uplink scheduler. There is little reason to schedule a higher data rate than the available transmission power can support. In the downlink, the available power is immediately known to the scheduler as the power amplifier is located in the same node as the scheduler. For the uplink, the power availability, or *power headroom*, needs to be provided to the eNodeB. Power headroom reports are therefore transmitted from the terminal to the eNodeB in a similar way as the buffer-status reports—that is, only when the terminal is scheduled to transmit on the UL-SCH. A power headroom report can be triggered for the following reasons:

- Periodically as controlled by a timer.
- Change in path loss (the difference between the current power headroom and the last report is larger than a configurable threshold).
- Instead of padding (for the same reason as buffer-status reports).

It is also possible to configure a prohibit timer to control the minimum time between two power-headroom reports and thereby the signaling load on the uplink.

There are two different types of power-headroom reports defined in LTE, *Type 1* and *Type 2*. Type 1 reporting reflects the power headroom assuming PUSCH-only transmission on the carrier, while the Type-2 report, introduced in release 10, assumes combined PUSCH and PUCCH transmission. Type-1 reports are provided for all component carriers, while Type-2 reports are provided for the primary component carrier only (PUCCH can be transmitted on the primary component carrier only).

The Type-1 power headroom valid for a certain subframe and a certain component carrier, assuming that the terminal was really scheduled for PUSCH transmission in that subframe, is given by (in dB)

$$\text{Power Headroom} = P_{\text{CMAX},c} - (P_{0,\text{PUSCH}} + \alpha \cdot PL_{\text{DL}} + 10 \cdot \log_{10}(M) + \Delta_{\text{MCS}} + \delta) \quad (13.1)$$

where the values for M and Δ_{MCS} correspond to the resource assignment and modulation-and-coding scheme used in the subframe to which the power-headroom report corresponds. The quantity δ captures the change in transmission power due to the closed-loop power control as described in Chapter 11. The explicitly configured *maximum per-carrier transmit power* for component carrier c is denoted, $P_{\text{CMAX},c}$. It can be noted that the power headroom is not a measure of the difference between the maximum per-carrier transmit power and the actual carrier transmit power. Rather, it can be seen that the power headroom is a measure of the difference between $P_{\text{CMAX},c}$ and the transmit power that would have been used *assuming that there would have been no upper limit on the transmit power*. Thus, the power headroom can very well be negative, indicating that the per-carrier transmit power was limited by $P_{\text{CMAX},c}$ at the time of the power headroom reporting—that is, the network has scheduled a higher data rate than the terminal can support given the available transmission power. As the network knows what modulation-and-coding scheme and resource size the terminal used for transmission in the subframe to which the power-headroom report corresponds, it can determine the valid combinations of modulation-and-coding scheme and resource size M , assuming that the downlink path loss PL_{DL} and the term δ have not changed substantially.

Type-1 power headroom can also be reported for subframes where there is no actual PUSCH transmission. In such cases, $10 \cdot \log_{10}(M)$ and Δ_{MCS} in the expression above are set to zero. This can be seen as the power headroom assuming a default transmission configuration corresponding to the minimum possible resource assignment ($M = 1$) and the modulation-and-coding scheme associated with $\Delta_{\text{MCS}} = 0$ dB.

Similarly, Type-2 power headroom reporting is defined as the difference between the maximum per-carrier transmit power and the sum of the PUSCH and PUCCH transmit power (equations 11.3 and 11.2, respectively), once again not taking into account any maximum per-carrier power when calculating the PUSCH and PUCCH transmit power.

Along the lines of Type-1 power headroom reporting, the Type-2 power headroom can also be reported for subframes in which no PUSCH and/or PUCCH is transmitted. In that case a virtual PUSCH and PUCCH transmit power is calculated, assuming the smallest possible resource assignment ($M = 1$) and $\Delta_{\text{MCS}} = 0$ dB for PUSCH and $\Delta_{\text{Format}} = 0$ for PUCCH.

13.3 Timing of scheduling assignments/grants

The scheduling decisions, downlink scheduling assignments, and uplink scheduling grants are communicated to each of the scheduled terminals through the downlink L1/L2 control signaling as described in Chapter 10, using one (E)PDCCH per downlink assignment. Each terminal monitors a set of (E)PDCCHs for valid scheduling assignment or grants and, upon detection of a valid assignment or grant, receives PDSCH or transmits PUSCH, respectively.

The terminal needs to know which subframe the scheduling command relates to and, in case of carrier aggregation, to which component carrier.

In the absence of cross-carrier scheduling, downlink scheduling assignments are valid for the component carrier upon which they are transmitted. Similarly, for uplink grants, there is an association between downlink and uplink component carriers such that each uplink component carrier has an associated downlink component carrier. The association is provided as part of the system information. Thus, from the uplink–downlink association, the terminal knows to which uplink component carrier the downlink control information relates to.

In the presence of cross-carrier scheduling, where downlink PDSCH or uplink PUSCH is transmitted on an (associated) component carrier other than that which the (E)PDCCH is transmitted upon, the carrier indicator in the (E)PDCCH provides information about the component carrier used for the PDSCH or PUSCH.

The scheduling timing—the subframe a scheduling assignment or grant received is valid for—is a bit more complex, especially for the TDD uplink, as discussed in the following section.

13.3.1 Downlink scheduling timing

For downlink data transmission, the scheduling assignment is transmitted in the same subframe as the data. Having the scheduling assignment in the same subframe as the corresponding data minimizes the latency in the scheduling process. Also, note that there is no possibility for dynamically scheduling future subframes—data and the scheduling assignment are always in the same subframe. This holds for FDD as well as TDD and both in the presence or absence of carrier aggregation, with the exception of cross-carrier scheduling for TDD with different downlink–uplink allocations.

In case of cross-carrier scheduling for TDD with different downlink–uplink allocations, the downlink scheduling timing for each component carrier follows that of the primary component carrier. Together with the fact that a scheduling assignment cannot point to a downlink subframe in the future, this implies that some downlink subframes on some component carriers cannot be scheduled with cross-carrier scheduling (see Figure 13.5). This

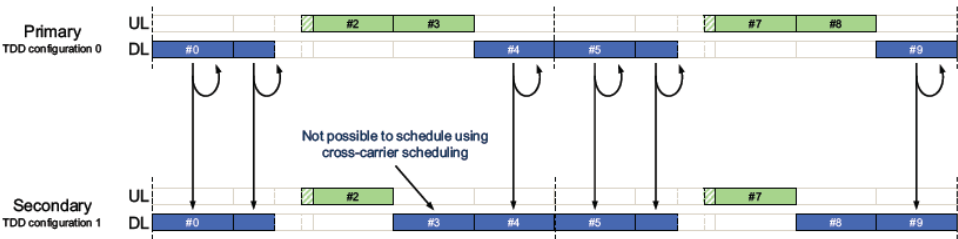


FIGURE 13.5

Example of cross-carrier scheduling with different TDD configurations

is consistent with the fact that, for a TDD terminal not capable of simultaneous transmission and reception, only downlink (uplink) subframes on a secondary component carrier that coincides with a downlink (uplink) subframe on the primary component carrier may be used. Note though that this holds only in case of cross-carrier scheduling *and* TDD with *different* downlink-uplink allocations for the different component carriers—in absence of cross-carrier scheduling, all subframes can be scheduled.

13.3.2 Uplink scheduling timing

The timing for uplink scheduling grants is more intricate than the corresponding downlink scheduling assignments, especially for TDD. The grant cannot relate to the same subframe it was received in as the uplink subframe has already started when the terminal has decoded the grant. The terminal also needs some time to prepare the data to transmit. Therefore, a grant received in subframe n affects the uplink transmission in a later subframe.

For FDD, the grant timing is straightforward. An uplink grant received in a subframe n triggers an uplink transmission in subframe $n + 4$, as illustrated in Figure 13.6. This is the same timing relation as used for uplink retransmission triggered by the PHICH, motivated by the possibility to override the PHICH by a dynamic scheduling grant, as described in Chapter 12.

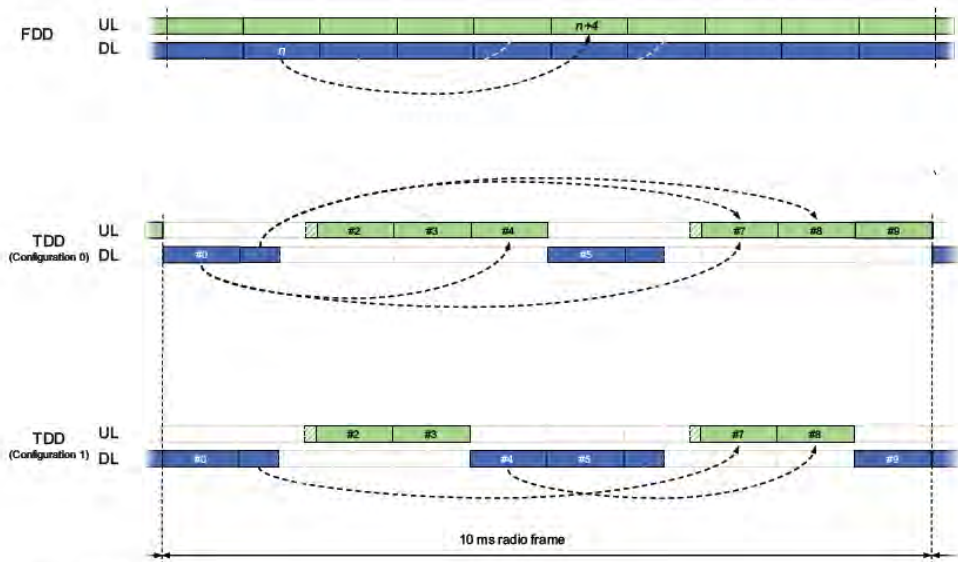


FIGURE 13.6

Timing relation for uplink grants in FDD and TDD configurations 0 and 1

For TDD, the situation is more complicated, especially in case of carrier aggregation with different downlink-uplink allocations on the different component carriers. Initially, consider the case where all component carriers, one or more, have the same downlink-uplink configurations.

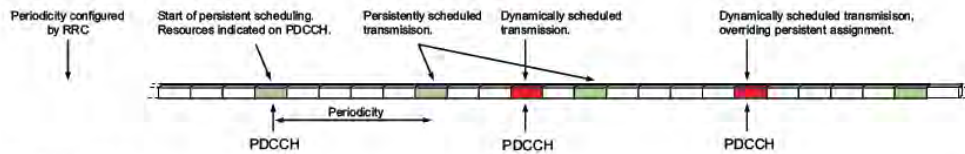
A grant received in subframe n in TDD may not necessarily trigger an uplink transmission in subframe $n + 4$ —that is, the timing relation used in FDD—as subframe $n + 4$ may not be an uplink subframe. Hence, for TDD configurations 1–6 the timing relation is modified such that the uplink transmission occurs in subframe $n + k$, where k is the smallest value larger than or equal to 4 such that subframe $n + k$ is an uplink subframe. This provides at least the same processing time for the terminal as in the FDD case while minimizing the delay from receipt of the uplink grant to the actual transmission. Note that this implies that the time between grant reception and uplink transmission may differ between different subframes. Furthermore, for the downlink-heavy configurations 1–5, another property is that uplink scheduling grants can only be received in some of the downlink subframes.

For TDD configuration 0 there are more uplink subframes than downlink subframes, which calls for the possibility to schedule transmissions in multiple uplink subframes from a single downlink subframe. The same timing relation as for the other TDD configurations is used, but with slight modifications. Recall from Section 10.4.7 that the grant transmitted in the downlink contains an uplink index consisting of two bits. For downlink-uplink configuration 0, the index field specifies which uplink subframe(s) a grant received in a downlink subframe applies to. For example, as illustrated in Figure 13.6, an uplink scheduling grant received in downlink subframe 0 applies to one or both of the uplink subframes 4 and 7, depending on which of the bits in the uplink index are set.

The description above holds also for carrier aggregation with different downlink-uplink allocations in absence of cross-carrier scheduling. In presence of cross-carrier scheduling, the situation is more complicated. For some configurations the timing is given by the configuration of the scheduling carrier—the carrier on which the (E)PDCCH is transmitted—while for other configurations, the timing follows the configuration of the scheduled carrier—that is, the carrier upon which the PUSCH was transmitted.

13.4 Semi-persistent scheduling

The basis for uplink and downlink scheduling is dynamic scheduling, as described in Sections 13.1 and 13.2. Dynamic scheduling with a new scheduling decision taken in each subframe allows for full flexibility in terms of the resources used, and can handle large variations in the amount of data to transmit at the cost of the scheduling decision being sent on an (E)PDCCH in each subframe. In many situations, the overhead in terms of control signaling on the (E)PDCCH is well motivated and relatively small compared to the payload on DL-SCH/UL-SCH. However, some services, most notably voice-over IP, are characterized by regularly occurring transmission of relatively small payloads. To reduce the control signaling overhead for those services, LTE provides semi-persistent scheduling in addition to dynamic scheduling.

**FIGURE 13.7**

Example of semi-persistent scheduling

With semi-persistent scheduling, the terminal is provided with the scheduling decision on the (E)PDCCH, together with an indication that this applies to every n th subframe until further notice. Hence, control signaling is only used once and the overhead is reduced, as illustrated in Figure 13.7. The periodicity of semi-persistently scheduled transmissions—that is, the value of n —is configured by RRC signaling in advance, while activation (and deactivation) is done using the (E)PDCCH using the semi-persistent C-RNTI.² For example, for voice-over IP the scheduler can configure a periodicity of 20 ms for semi-persistent scheduling and, once a talk spurt starts, the semi-persistent pattern is triggered by the (E)PDCCH.

After enabling semi-persistent scheduling, the terminal continues to monitor the set of candidate (E)PDCCHs for uplink and downlink scheduling commands. When a dynamic scheduling command is detected, it takes precedence over the semi-persistent scheduling in that particular subframe, which is useful if the semi-persistently allocated resources occasionally need to be increased. For example, for voice-over IP in parallel with web browsing, it may be useful to override the semi-persistent resource allocation with a larger transport block when downloading the web page.

For the downlink, only initial transmissions use semi-persistent scheduling. Retransmissions are explicitly scheduled using an (E)PDCCH assignment. This follows directly from the use of an asynchronous hybrid-ARQ protocol in the downlink. Uplink retransmissions, in contrast, can either follow the semi-persistently allocated subframes or be dynamically scheduled.

Semi-persistent scheduling is only supported on the primary component carrier and any transmission on a secondary component carrier must be dynamically scheduled. This is reasonable as semi-persistent scheduling is intended for low-rate services for which a single component carrier is sufficient.

13.5 Scheduling for half-duplex FDD

Half-duplex FDD implies that a single terminal cannot receive and transmit at the same time while the eNodeB still operates in full duplex. In LTE, half-duplex FDD is implemented as a scheduler constraint, implying it is up to the scheduler to ensure that a single terminal is not

²Each terminal has two identities, the “normal” C-RNTI for dynamic scheduling and the semi-persistent C-RNTI for semi-persistent scheduling.

scheduled simultaneous in uplink and downlink. Hence, from a terminal perspective, subframes are dynamically used for uplink or downlink. Briefly, the basic principle for half-duplex FDD is that a terminal is receiving in the downlink unless it has been explicitly instructed to transmit in the uplink (either UL-SCH transmission or hybrid-ARQ acknowledgements triggered by a downlink transmission). The timing and structure for control signaling are identical between half- and full-duplex FDD terminals.

An alternative design approach would be to base half-duplex FDD on the TDD control signaling structure and timing, with a semi-static configuration of subframes to either downlink or uplink. However, this would complicate supporting a mixture of half- and full duplex terminals in the same cell as the timing of the control signaling would differ. It would also imply a waste of uplink spectrum resources. All terminals need to be able to receive subframes 0 and 5, as those subframes are used for system information and synchronization signals. Hence, if a fixed uplink–downlink allocation were to be used, no uplink transmissions could take place in those two subframes, resulting in a loss in uplink spectral efficiency of 20%. This is not attractive and led to the choice of implementing half-duplex FDD as a scheduling strategy instead.

An example of half-duplex operation as seen from a terminal perspective is shown in Figure 13.8. In the leftmost part of the figure, the terminal is explicitly scheduled in the uplink and, consequently, cannot receive data in the downlink in the same subframe. The uplink transmission implies the receipt of an acknowledgement on the PHICH four subframes later, as mentioned in Chapter 12, and therefore the terminal cannot be scheduled in the uplink in this subframe. Similarly, when the terminal is scheduled to receive data in the downlink in subframe n , the corresponding hybrid-ARQ acknowledgement needs to be transmitted in the uplink subframe $n + 4$, preventing downlink reception in subframe $n + 4$. The scheduler can exploit this by scheduling downlink data in four consecutive subframes, and uplink transmission in the four next subframes when the terminal needs to transmit hybrid-ARQ acknowledgements in the uplink anyway, and so on. Hence, at most half of the time can be used in the downlink and half in the uplink or, in other words, the asymmetry in half-duplex FDD is 4:4. Efficient support of half-duplex FDD is one of the reasons why the same number of hybrid-ARQ processes was selected in uplink and downlink.

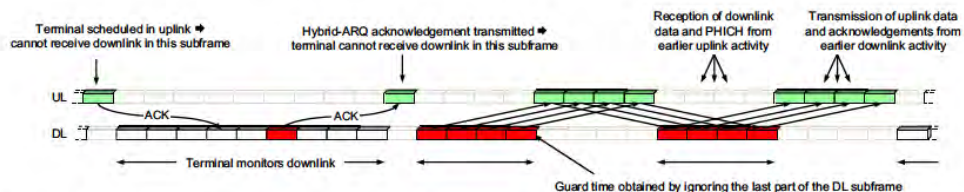


FIGURE 13.8

Example of half-duplex FDD terminal operation

Note that, as the eNodeB is operating in full duplex, regardless of the duplex capability of the terminals, the cell capacity is hardly affected by the presence of half-duplex terminals. Given a sufficient number of terminals with data to transmit/receive, the scheduler can with a high likelihood find a set of terminals to schedule in the uplink and another set to schedule in the downlink in a given subframe.

Similar to TDD, a half-duplex terminal needs some guard time for switching between uplink and downlink. For half-duplex FDD, guard time for the downlink-to-uplink switch is created by allowing the terminal to skip reception of the last OFDM symbols in a downlink subframe immediately preceding an uplink subframe, as described in Chapter 9. Guard time for the uplink-to-downlink switch is handled by setting the appropriate amount of timing advance in the terminals.

13.6 Discontinuous reception (DRX) and component carrier deactivation

Packet-data traffic is often highly bursty, with occasional periods of transmission activity followed by longer periods of silence. From a delay perspective, it is beneficial to monitor the downlink control signaling in each subframe to receive uplink grants or downlink data transmissions and instantaneously react on changes in the traffic behavior. At the same time this comes at a cost in terms of power consumption at the terminal; the receiver circuitry in a typical terminal represents a non-negligible amount of power consumption. To reduce the terminal power consumption, LTE includes mechanisms for *discontinuous reception* (DRX).

The basic mechanism for DRX is a configurable DRX cycle in the terminal. With a DRX cycle configured, the terminal monitors the downlink control signaling only in one subframe per DRX cycle, sleeping with the receiver circuitry switched off in the remaining subframes. This allows for a significant reduction in power consumption: the longer the cycle, the lower the power consumption. Naturally, this implies restrictions to the scheduler as the terminal can be addressed only in the active subframes.

In many situations, if the terminal has been scheduled and active with receiving or transmitting data in one subframe, it is highly likely it will be scheduled again in the near future. One reason could be that it was not possible to transmit all the data in the transmission buffer in one subframe and additional subframes are required. Waiting until the next active subframe according to the DRX cycle, although possible, would result in additional delays. Hence, to reduce the delays, the terminal remains in the active state for a certain configurable time after being scheduled. This is implemented by the terminal (re)starting an inactivity timer every time it is scheduled and remaining awake until the time expires, as illustrated at the top of Figure 13.9.

Retransmissions take place regardless of the DRX cycle. Thus, the terminal receives and transmits hybrid-ARQ acknowledgements as normal in response to data transmission. In the uplink, this also includes retransmissions in the subframes given by the synchronous

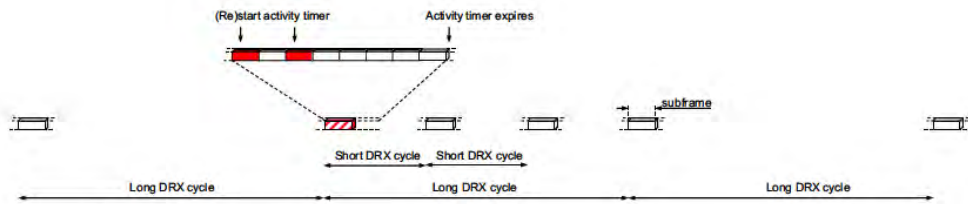
**FIGURE 13.9**

Illustration of DRX operation

hybrid-ARQ timing relation. In the downlink, where asynchronous hybrid ARQ is used, the retransmission time is not fixed in the specifications. To handle this, the terminal monitors the downlink for retransmissions in a configurable time window after the previous transmission.

The above mechanism, a (long) DRX cycle in combination with the terminal remaining awake for some period after being scheduled, is sufficient for most scenarios. However, some services, most notably voice-over IP, are characterized by periods of regular transmission, followed by periods of no or very little activity. To handle these services, a second short DRX cycle can optionally be used in addition to the long cycle described above. Normally, the terminal follows the long DRX cycle, but if it has recently been scheduled, it follows a shorter DRX cycle for some time. Handling voice-over IP in this scenario can be done by setting the short DRX cycle to 20 ms, as the voice codec typically delivers a voice-over-IP packet per 20 ms. The long DRX cycle is then used to handle longer periods of silence between talk spurts.

The DRX mechanism described above is common to all component carriers configured in the terminal. Hence, if the terminal is in DRX it is not receiving on any component carrier, but when it wakes up, all (activated) component carriers will be woken up. Although discontinuous reception greatly reduces the terminal power consumption, it is possible to go one step further in the case of carrier aggregation. From a power-consumption perspective, it is beneficial to receive on as few component carriers as possible. LTE therefore supports deactivation of downlink component carriers. A deactivated component carrier maintains the configuration provided by RRC but cannot be used for reception, neither PDCCH nor PDSCH. When the need arises, a downlink component carrier can be activated rapidly and used for reception within a few subframes. A typical use would be to configure several component carriers, but deactivate all component carriers except the primary one. When a data burst starts, the network could activate several component carriers to maximize the downlink data rate. Once the data burst is delivered, the component carriers could be deactivated again to reduce terminal power consumption.

Activation and deactivation of downlink component carriers are done through MAC control elements. A command to activate a component carrier takes effect eight subframes after receipt—that is, if the MAC control element is received in subframe n , then the additional component carriers are activated starting from subframe $n + 8$. There is also a

timer-based mechanism for deactivation such that a terminal may, after a configurable time with no activity on a certain component carrier, deactivate that component carrier. The primary component carrier is always active as it must be possible for the network to communicate with the terminal.

In the uplink there is no explicit activation of uplink component carriers. However, whenever a downlink component carrier is activated or deactivated, the corresponding uplink component carrier is also activated or deactivated.

13.7 Channel-state information

As mentioned several times, the possibility for downlink channel-dependent scheduling—that is, selecting the downlink transmission configuration and related parameters depending on the instantaneous downlink channel condition, including the interference situation—is a key feature of LTE. An important part of the support for downlink channel-dependent scheduling is *channel-state information* (CSI) provided by terminals to the network, information on which the latter can base its scheduling decisions.

The CSI consist of one or several pieces of information:

- *Rank indication* (RI), providing a recommendation on the transmission rank to use or, expressed differently, the number of layers that should preferably be used for DL-SCH transmission to the terminal.
- *Precoder matrix indication* (PMI), indicating a preferred precoder to use for DL-SCH transmission.
- *Channel-quality indication* (CQI), representing the highest modulation-and-coding scheme that, if used, would mean a DL-SCH transmission using the recommended RI and PMI would be received with a block-error probability of at most 10%.

Together, a combination of the RI, PMI, and CQI forms a CSI report. Exactly what is included in a CSI report depends on the reporting mode the terminal is configured to be in. For example, RI and PMI do not need to be reported unless the terminal is in a spatial-multiplexing transmission mode. However, also given the transmission mode, there are different reporting configurations that typically differ as to what set of resource blocks the report is valid for and whether precoding information is reported or not. The type of information useful to the network also depends on the particular implementation and antenna deployment.

Although referred to as channel-state information, what a terminal delivers to the network are not explicit reports of the downlink channel state. Rather, what the terminal delivers are *suggestions* on the transmission rank and precoding matrix to use, together with an indication of the highest possible modulation-and-coding scheme that the network should not exceed if it intends to keep the block-error probability below 10%. Information about the actual modulation scheme and coding rate used for DL-SCH transmission as well as the set of resource blocks used for the transmission is always included in the downlink scheduling

assignment. Hence, the eNodeB is free to follow the CSI report or to select transmission parameters on its own.

13.7.1 Channel-quality indication

The channel-quality indication, CQI, provides an estimate of the highest modulation-and-coding scheme that, if used with the recommended RI and PMI, would result in a block-error probability for the DL-SCH transmissions of at most 10%. The reason to use CQI as a feedback quantity instead of, for example, the signal-to-noise ratio, is to account for different receiver implementations in the terminal. Also, basing the feedback reports on CQI instead of signal-to-noise ratio simplifies the testing of terminals; a terminal delivering data with more than 10% block-error probability when using the modulation-and-coding scheme indicated by the CQI would fail the test. As will be discussed further below, multiple CQI reports, each representing the channel quality in a certain part of the downlink spectrum, can be part of the channel-state information.

The modulation-and-coding scheme used for DL-SCH transmission can, and often will, differ from the reported CQI as the scheduler needs to account for additional information not available to the terminal when recommending a certain CQI. For example, the set of resource blocks used for the DL-SCH transmission also needs to account for other users. Furthermore, the amount of data awaiting transmission in the eNodeB also needs to be accounted for. There is no need to select a very high data rate, even if the channel conditions would permit this, if there is only a small amount of data to transmit and a small number of resource blocks with robust modulation is sufficient.

13.7.2 Rank indicator and precoder matrix indication

The multi-antenna related part of the CSI consists of the rank indication, RI, and the precoder matrix indication, PMI.

The rank indication is only reported by terminals that are configured in one of the spatial-multiplexing transmission modes. There is at most one RI reported (for a particular CSI process on a particular component carrier, see below for details), valid across the full bandwidth—that is, the RI is frequency non-selective. Note that frequency-selective transmission rank is not possible in LTE as all layers are transmitted on the same set of resource blocks.

The precoder matrix indication provides an indication to the eNodeB of the preferred precoder to use conditioned on the number of layers indicated by the RI. The precoder recommendation may be frequency selective, implying that the terminal may recommend different precoders for different parts of the downlink spectrum, or frequency non-selective.

With regards to the precoder-related recommendations, the network has two choices:

- The network may follow the latest terminal recommendation, in which case the eNodeB only has to confirm (a one-bit indicator in the downlink scheduling assignment)

that the precoder configuration recommended by the terminal is used for the downlink transmission. On receiving such a confirmation, the terminal will use its recommended configuration when demodulating and decoding the corresponding DL-SCH transmission. Since the PMI computed in the terminal can be frequency selective, an eNodeB following the precoding matrix recommended by the terminal may have to apply different precoding matrices for different (sets of) resource blocks.

- The network may select a different precoder, information about which then needs to be explicitly included in the downlink scheduling assignment. The terminal then uses this configuration when demodulating and decoding the DL-SCH. To reduce the amount of downlink signaling, only a single precoding matrix can be signaled in the scheduling assignment, implying that, if the network overrides the recommendation provided by the terminal, the precoding is frequency non-selective. The network may also choose to override the suggested transmission rank only.

The precoder recommended by the terminal is not explicitly signaled but provided as an index into a set of predefined matrices: a so-called *codebook*. The set of matrices from which the terminal selects the preferred precoder depends on the number of antenna ports.

For two or four antenna ports (irrespective of whether CRS or DM-RS is used for demodulation), the codebook is given by the set of precoders for codebook-based precoding (see Section 10.3.2). This is a natural choice for the multi-antenna transmission modes based on cell-specific reference signals (transmission modes 4, 5, and 6) as the network in these modes must use a codebook from this set. For simplicity the same codebook is used for CSI reporting for transmission modes 8, 9, and 10 using demodulation-specific reference signals, even though the network in this case can use any precoders, as the precoder used is transparent to the terminal.

For eight antenna ports, supported in transmission modes 9 and 10 and relying on demodulation-specific reference signals, a somewhat different approach for the codebook design than the two and four antenna cases was taken. All precoders in this case can be factorized as $W = W_1 \cdot W_2$, where the possible entries for W_1 model long-term/wideband aspects such as beam-forming while the possible entries of W_2 address short-term/frequency-selective properties such as polarization properties. One reason for factorizing the codebook is to simplify implementation as the terminal may apply different (time-domain) filtering to the two parts when forming PMI. The resulting codebook is tailored for closely spaced cross-polarized antenna elements (see Figure 13.10), which is a realistic antenna setup, but it also covers the case of eight closely spaced linearly polarized antennas.

Finally, for two, four, and eight antenna ports, the network can restrict the set of matrices from which the terminal can select the recommended precoder, so-called codebook subset restriction, to avoid reporting precoders that are not useful in the antenna setup actually used.

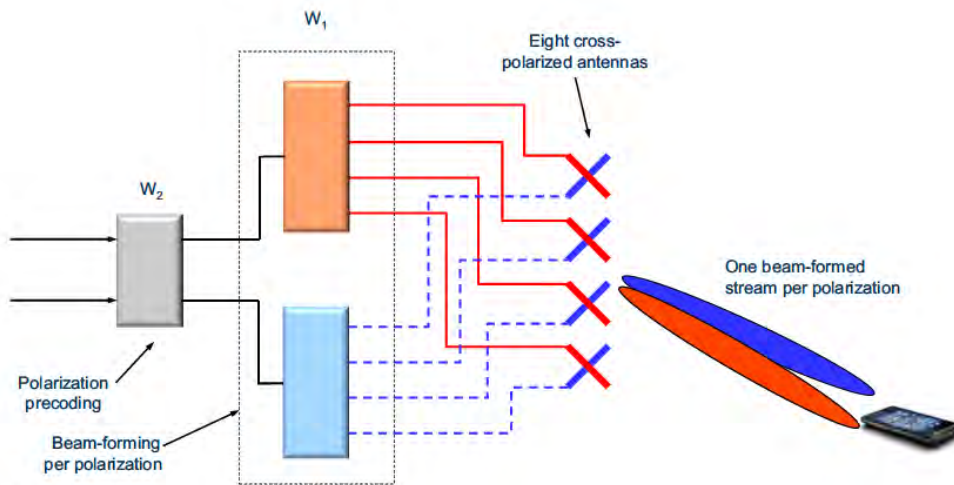


FIGURE 13.10

Illustration of the eight-antenna codebook

13.7.3 Periodic and aperiodic CSI reports

There are two types of CSI reports in LTE, *aperiodic* and *periodic* reports, which are different in terms of how a report is triggered:

- Aperiodic CSI reports are delivered when explicitly requested by the network by means of the channel-state-request flag included in uplink scheduling grants (see Section 10.4.5). An aperiodic CSI report is always delivered using the PUSCH—that is, on a dynamically assigned resource.
- Periodic CSI reports are configured by the network to be delivered with a certain periodicity, possibly as often as once every 2 ms, on a semi-statically configured PUCCH resource. However, similar to hybrid-ARQ acknowledgements normally delivered on PUCCH, channel-state reports are “re-routed” to the PUSCH³ if the terminal has a valid uplink grant.

Aperiodic and periodic reports, despite both providing estimates of the channel and interference conditions at the terminal, are quite different in terms of their detailed contents and usage. In general, aperiodic reports are larger and more detailed than their periodic counterparts. There are several reasons for this. First, the PUSCH, upon which the aperiodic report is transmitted, is capable of a larger payload, and hence a more detailed report, than the

³In release 10 and later, a terminal can be configured for simultaneous PUSCH and PUCCH transmission, in which case the periodic channel-state reports can remain on the PUCCH.

PUCCH used for the periodic reports. Furthermore, as aperiodic reports are transmitted on a per-need basis only, the overhead from these reports is less of an issue compared to periodic reports. Finally, if the network requests a report it is likely that it will transmit a large amount of data to the terminal, which makes the overhead from the report less of an issue compared to a periodic report that is transmitted irrespective of whether the terminal in question will be scheduled in the near future or not. Hence, as the structure and usage of aperiodic and periodic reports is different, they are described separately below, starting with aperiodic reports.

13.7.3.1 Aperiodic CSI reports

Aperiodic CSI reports are transmitted on the PUSCH upon request by the network. Three aperiodic reporting modes, where each mode has several submodes depending on the configuration, are supported in LTE:

- Wideband reports, reflecting the average channel quality across the entire cell bandwidth with a single CQI value. Despite a single average CQI value being provided for the whole bandwidth, the PMI reporting is frequency selective. Frequency-selective reporting is obtained, for reporting purposes only, by dividing the overall downlink bandwidth (of each component carrier) into a number of equally sized *sub-bands*, where each sub-band consists of a set of consecutive resource blocks. The size of a sub-band, ranging from four to eight resource blocks, depends on the cell bandwidth. The PMI is then reported for each sub-band. For transmission modes supporting spatial multiplexing, the CQI and the PMI are calculated assuming the channel rank indicated by the RI, otherwise rank-1 is assumed. Wideband reports are smaller than their frequency-selective counterparts, but do not provide any information about the frequency domain.
- UE-selected reports, where the terminal selects the best M sub-bands and reports, in addition to the indices of the selected sub-bands, one CQI reflecting the average channel quality over the selected M sub-bands together with one wideband CQI reflecting the channel quality across the full downlink carrier bandwidth. This type of report thus provides frequency-domain information about the channel conditions. The sub-band size, ranging from two to four resource blocks, and the value of M , ranging from 1 to 6, depends on the downlink carrier bandwidth. Depending on the sub-mode configured, the PMI and RI are also provided as part of this type of report.
- Configured reports, where the terminal reports one wideband CQI reflecting the channel quality across the full downlink carrier bandwidth and one CQI per sub-band. The sub-band size depends on the downlink carrier bandwidth and is in the range of four to eight resource blocks. Depending on the sub-mode configured, the PMI and RI are also provided as part of this type of report.

The different aperiodic reporting modes are summarized in [Table 13.1](#).

The preceding discussion also holds for carrier aggregation, although with some modifications and enhancements as channel-state reports for multiple downlink component

Table 13.1 Possible A periodic Reporting Modes for Different Transmission Modes								
Transmission Mode	Reporting Mode							
	Wideband CQI			Frequency-Selective CQI				
				UE-Selected Sub-bands			Conf. Sub-bands	
	1-0: No PMI	1-1: Wideband PMI	1-2: Selective PMI	2-0: No PMI	2-1: Wideband PMI	2-2:Selective PMI	3-0: No PMI	3-1: Wideband PMI
1 Single antenna, CRS				•			•	
2 Transmit diversity				•			•	
3 Open-loop spatial mux				•			•	
4 Closed-loop spatial mux			•			•		•
5 Multi-user MIMO								•
6 Codebook-based beam-form			•			•		•
7 Single-layer trans., DM-RS				•			•	
8 Dual-layer trans., DM-RS			•	•		•	•	•
9 Multi-layer trans., DM-RS			•	•		•	•	•
10 Multi-layer trans., DM-RS			•	•		•	•	•

carriers are typically needed. For aperiodic reporting, the two-bit⁴ CSI request in the downlink control signaling allows for three different sets of CSI reports to be requested (the fourth bit combination represents no CSI request). Of these three alternatives, one is used to trigger a CSI report for the downlink component carrier associated with the uplink component carrier the scheduling grant relates to. The remaining alternatives point to one of two configurable combinations of component carriers for which the CSI report should be generated. Thus, as an example, for a terminal capable of two downlink component carriers, aperiodic reports can, with the proper configuration, be requested for the primary component carrier, the secondary component carrier, or both.

13.7.3.2 Periodic CSI reports

Periodic reports are configured by the network to be delivered with a certain periodicity. The periodic reports are transmitted on the PUCCH (unless the terminal has a simultaneous grant for PUSCH transmission and is not using simultaneous PUSCH and PUCCH transmission). The limited payload possible on PUCCH, compared to PUSCH, implies that the different types of information in a periodic report may not be possible to transmit in a single subframe. Therefore, some of the reporting modes will transmit one or several of the wideband CQI, the wideband CQI including PMI, the RI, and the CQI for the UE-selected sub-bands at different time instants. Furthermore, the RI can typically be reported less often, compared to the reporting of PMI and CQI, reflecting the fact that the suitable number of layers typically varies on a slower basis, compared to the channel variations that impact the choice of precoder matrix and modulation rate, and coding scheme.

Two periodic reporting modes, again with different submodes possible, are supported in LTE:

- Wideband reports, reflecting the average channel quality across the entire cell bandwidth with a single CQI value. If PMI reporting is enabled, a single PMI valid across the full bandwidth is reported.
- UE-selected reports. Although named in the same way as for aperiodic reports, the principle for UE-selected periodic reports is different. The total bandwidth (of a component carrier) is divided into one to four *bandwidth parts*, with the number of bandwidth parts obtained from the cell bandwidth. For each bandwidth part, the terminal selects the best sub-band within that part. The sub-band size ranges from four to eight resource blocks. Since the supported payload size of the PUCCH is limited, the reporting cycles through the bandwidth parts and in one subframe reports the wideband CQI and PMI (if enabled) for that bandwidth part, as well as the best sub-band and the CQI for that sub-band. The RI (if enabled) is reported in a separate subframe.

⁴Recall from Chapter 10 that the CSI request field expanded from one to two bits in the case of carrier aggregation.

The different periodic reporting modes are summarized in Table 13.2. Note that all PMI reporting, if enabled, is of wideband type. There is no support for frequency-selective PMI in periodic reporting, as the amount of bits would result in a too large overhead.

A typical use of periodic and aperiodic reporting could be to configure lightweight periodic CSI reporting PUCCH, for example to provide feedback of the wideband CQI and no PMI information (mode 1-0). Upon arrival of data to transmit in the downlink to a specific terminal, aperiodic reports could be requested as needed, for example with frequency-selective CQI and PMI (mode 3-1).

For carrier aggregation, the basic principle is to configure the reporting cycles such that the CSI reports for the different component carriers are not transmitted simultaneously on PUCCH. Thus, CSI reports for different component carriers are transmitted in different subframes.

13.7.4 CSI estimation

Channel-state reports, irrespective of whether they are aperiodic or periodic, need measurements of the channel properties as well as the interference level.

Measuring the channel gain is relatively straightforward and from the first release it is well specified which subframe the CSI report relates to. The reference signal which the channel-gain estimate is based upon, CRS or CSI-RS, depends on the transmission mode. For transmission modes already supported in release 8/9, the cell-specific reference signals are used, while for transmission modes 9 and 10, introduced in releases 10 and 11, respectively, the CSI-RS is used.

Measuring the interference level, which is required in order to form a relevant CSI, is more cumbersome, and the measurement is greatly affected by the transmission activity in neighboring cells. LTE releases 10 and earlier do not specify how to measure the interference level, and leave the details for the implementation. However, in practice, interference is measured as the noise on the cell-specific reference signals—that is, the residual after subtracting the reference signal from the received signal in the appropriate resource elements is used as an estimate of the interference level. At low loads, this approach unfortunately often results in overestimating the interference level as the measurements are dominated by CRS transmissions in neighboring cells (assuming the same CRS positions in the neighboring cells), irrespective of the actual load in those cells. Furthermore, the terminal may also choose to average the interference level across multiple subframes,⁵ further adding to the uncertainty on how the interference is measured by the terminal.

To address these shortcomings and to better support various CoMP schemes, transmission mode 10, introduced in release 11, provides tools for the network to control on which resource elements the interference is measured. The basis is a so-called *CSI Interference Measurement*

⁵In release 10, there was a possibility to limit the interference averaging to different subsets of subframes in order to improve the support for heterogeneous deployments (see Chapter 15).

Table 13.2 Possible Periodic Reporting Modes for Different Transmission Modes								
Transmission Mode	Reporting Mode							
	Wideband CQI			Frequency-Selective CQI				
				UE-Selected Sub-bands			Conf. Sub-bands	
	1-0: No PMI	1-1: Wideband PMI	1-2: Selective PMI	2-0: No PMI	2-1: Wideband PMI	2-2: Selective PMI	3-0: No PMI	3-1: Wideband PMI
1 Single antenna, CRS	•			•				
2 Transmit diversity	•			•				
3 Open-loop spatial mux	•			•				
4 Closed-loop spatial mux		•			•			
5 Multi-user MIMO		•			•			
6 Codebook-based beam-form		•			•			
7 Single-layer trans., DM-RS	•			•				
8 Dual-layer trans., DM-RS	•	•		•	•			
9 Multi-layer trans., DM-RS	•	•		•	•			
10 Multi-layer trans., DM-RS	•	•		•	•			

(CSI-IM) configuration, where a CSI-IM configuration is the set of resource elements in one subframe the terminal should use for measuring interference. The received power in the resource elements corresponding to the CSI-IM configuration is used as an estimate of the interference (and noise). The subframe in which the interference should be measured is also specified, thereby avoiding terminal-specific, and to the network unknown, interference averaging.

Configuring a CSI-IM is done in a similar manner as a CSI-RS and the same set of configurations is available (see Chapter 10). In practice, a CSI-IM resource would typically correspond to a CSI-RS resource in which nothing is transmitted from the cell or, in the general case, from a certain transmission point. Thus, in practice, the CSI-IM resource will typically be covered by the set of zero-power CSI-RS resources configured for the terminal. However, CSI-IM and zero-power CSI-RS serve different purposes. CSI-IM is defined in order to specify a set of resource elements on which a terminal should measure the interference level, while zero-power CSI-RS is defined in order to specify a set of resource elements avoided by the PDSCH mapping.

Since the CSI-IM does not collide with the CRS in neighboring cells but rather the PDSCH (assuming a synchronized network), the interference measurement better reflects the transmission activity in neighboring cells, leading to a more accurate interference estimate at low loads. Hence, with the channel conditions estimated from the CSI-RS and the interference situation estimated from the CSI-IM, the network has detailed control of the interference situation the CSI report reflects.

In some scenarios, the eNodeB benefits from *multiple* CSI reports, derived under different interference hypotheses. Therefore, release 11 provides support for up to four *CSI processes* in a terminal, where a CSI process is defined by one CSI-RS configuration and one CSI-IM configuration. CSI is then reported separately for each process.⁶ A more in-depth discussion on the usage of CSI processes to support CoMP is found in Chapter 15.

⁶It is possible to configure *rank inheritance*, in which case the rank reported by one CSI process is inherited from another CSI process.

Access Procedures

14

The previous chapters have described the LTE uplink and downlink transmission schemes. However, prior to transmission of data, the terminal needs to connect to the network. This chapter describes the procedures necessary for a terminal to be able to access an LTE-based network.

14.1 Acquisition and cell search

Before an LTE terminal can communicate with an LTE network it has to do the following:

- Find and acquire synchronization to a cell within the network.
- Receive and decode the information, also referred to as the *cell system information*, needed to communicate with and operate properly within the cell.

The first of these steps, often simply referred to as *cell search*, is discussed in this section. The next section then discusses, in more detail, the means by which the network provides the cell system information.

Once the system information has been correctly decoded, the terminal can access the cell by means of the random-access procedure as described in [Section 14.3](#).

14.1.1 Overview of LTE cell search

A terminal does not only need to carry out cell search at power-up—that is, when initially accessing the system. Rather, to support mobility, terminals need to continuously search for, synchronize to, and estimate the reception quality of neighboring cells. The reception quality of the neighboring cells, in relation to the reception quality of the current cell, is then evaluated to conclude if a handover (for terminals in RRC_CONNECTED) or cell reselection (for terminals in RRC_IDLE) should be carried out.

LTE cell search consists of the following basic steps:

- Acquisition of frequency and symbol synchronization to a cell.
- Acquisition of frame timing of the cell—that is, determination of the start of the downlink frame.
- Determination of the physical-layer cell identity of the cell.

As already mentioned, for example, in Chapter 10, there are 504 different physical-layer cell identities defined for LTE. The set of physical-layer cell identities is further divided into 168 cell-identity groups, with three cell identities within each group.

To assist the cell search, two special signals are transmitted on each downlink component carrier, the *Primary Synchronization Signal* (PSS) and the *Secondary Synchronization Signal* (SSS). Although having the same detailed structure, the time-domain positions of the synchronization signals within the frame differ somewhat depending on whether the cell is operating in FDD or TDD:

- In the case of FDD (upper part of Figure 14.1), the PSS is transmitted within the last symbol of the first slot of subframes 0 and 5, while the SSS is transmitted within the second to last symbol of the same slot—that is, just prior to the PSS.
- In the case of TDD (lower part of Figure 14.1), the PSS is transmitted within the third symbol of subframes 1 and 6—that is, within the DwPTS—while the SSS is transmitted in the last symbol of subframes 0 and 5—that is, three symbols ahead of the PSS.

It should be noted that the difference in PSS/SSS time-domain structure between FDD and TDD allows for the terminal to detect the duplex mode of the acquired carrier if this is not known in advance.

Within one cell, the two PSSs within a frame are identical. Furthermore, the PSS of a cell can take three different values depending on the physical-layer cell identity of the cell. More specifically, the three cell identities within a cell-identity group always correspond to different PSS. Thus, once the terminal has detected and identified the PSS of the cell, it has found the following:

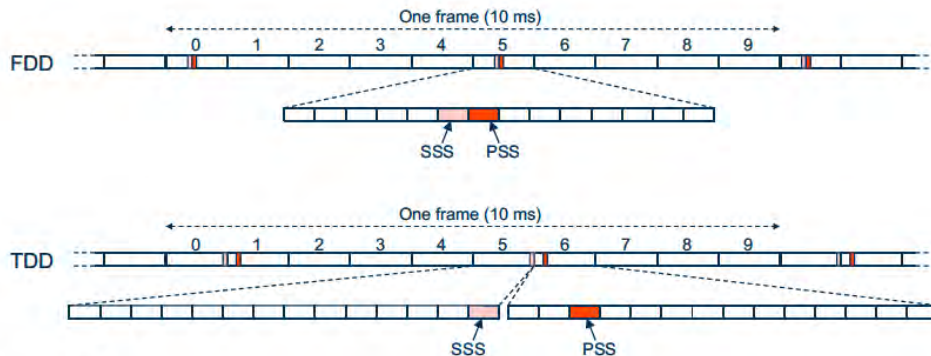
- Five-millisecond timing of the cell and thus also the position of the SSS which has a fixed offset relative to the PSS.¹
- The cell identity within the cell-identity group. However, the terminal has not yet determined the cell-identity group itself—that is, the number of possible cell identities has been reduced from 504 to 168.

Thus, from the SSS, the position of which is known once the PSS has been detected, the terminal should find the following:

- Frame timing (two different alternatives given the found position of the PSS).
- The cell-identity group (168 alternatives).

Furthermore, it should be possible for a terminal to do this by the reception of one single SSS. The reason is that, for example, in the case when the terminal is searching for cells on other carriers, the search window may not be sufficiently large to cover more than one SSS.

¹This assumes that the terminal knows if it has acquired an FDD or a TDD carrier. Otherwise the terminal needs to try two different hypotheses regarding the SSS position relative to the PSS, thereby also indirectly detecting the duplex mode of the acquired carrier.

**FIGURE 14.1**

Time-domain positions of PSS and SSS in case of FDD and TDD

To enable this, each SSS can take 168 different values corresponding to the 168 different cell-identity groups. Furthermore, the set of values valid for the two SSSs within a frame (SSS1 in subframe 0 and SSS2 in subframe 5) are different, implying that, from the detection of a single SSS, the terminal can determine whether SSS1 or SSS2 has been detected and thus determine frame timing.

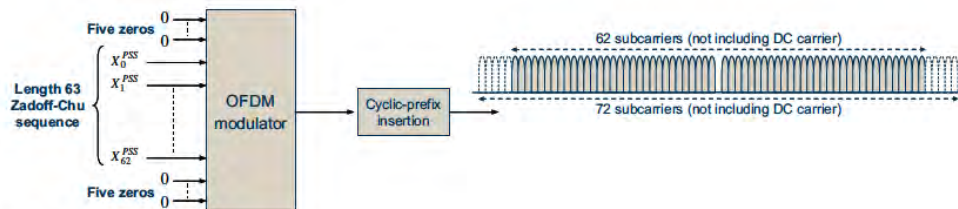
Once the terminal has acquired frame timing and the physical-layer cell identity, it has identified the cell-specific reference signal. The behavior is slightly different depending on whether it is an initial cell search or cell search for the purpose of neighboring cell measurements:

- In the case of initial cell search—the terminal state is in RRC_IDLE mode—the reference signal will be used for channel estimation and subsequent decoding of the BCH transport channel to obtain the most basic set of system information.
- In the case of mobility measurements—the terminal is in RRC_CONNECTED mode—the terminal will measure the received power of the reference signal. If the measurement fulfills a configurable condition, it will trigger sending of a reference signal received power (RSRP) measurement report to the network. Based on the measurement report, the network will conclude whether a handover should take place. The RSRP reports can also be used for component carrier management,² for example whether an additional component carrier should be configured or if the primary component carrier should be reconfigured.

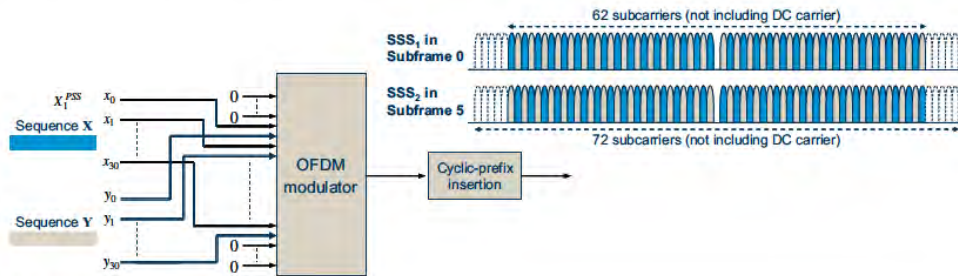
14.1.2 PSS structure

On a more detailed level, the three PSSs are three length-63 Zadoff–Chu (ZC) sequences (see Section 11.2) extended with five zeros at the edges and mapped to the center 73 subcarriers

²As discussed in Chapter 9, the specifications use the terms primary and secondary cells instead of primary and secondary component carriers.

**FIGURE 14.2**

Definition and structure of PSS

**FIGURE 14.3**

Definition and structure of SSS

(center six resource blocks) as illustrated in Figure 14.2. It should be noted that the center subcarrier is actually not transmitted as it coincides with the DC subcarrier. Thus, only 62 elements of the length-63 ZC sequences are actually transmitted (element X_{32}^{PSS} is not transmitted).

The PSS thus occupies 72 resource elements (not including the DC carrier) in subframes 0 and 5 (FDD) and subframes 1 and 6 (TDD). These resource elements are then not available for transmission of DL-SCH.

14.1.3 SSS structure

Similar to PSS, the SSS occupies the center 72 resource elements (not including the DC carrier) in subframes 0 and 5 (for both FDD and TDD). As described above, the SSS should be designed so that:

- The two SSSs (SSS1 in subframe 0 and SSS2 in subframe 5) take their values from sets of 168 possible values corresponding to the 168 different cell-identity groups.
- The set of values applicable for SSS2 is different from the set of values applicable for SSS1 to allow for frame-timing detection from the reception of a single SSS.

The structure of the two SSSs is illustrated in Figure 14.3. SSS1 is based on the frequency interleaving of two length-31 m-sequences X and Y, each of which can take 31 different values (actually 31 different shifts of the same m-sequence). Within a cell, SSS2 is based on exactly the same two sequences as SSS1. However, the two sequences have been swapped in the frequency domain, as outlined in Figure 14.3. The set of valid combinations of X and Y for SSS1 has then been selected so that a swapping of the two sequences in the frequency domain is not a valid combination for SSS1. Thus, the above requirements are fulfilled:

- The set of valid combinations of X and Y for SSS1 (as well as for SSS2) are 168, allowing for detection of the physical-layer cell identity.
- As the sequences X and Y are swapped between SSS1 and SSS2, frame timing can be found.

14.2 System information

By means of the basic cell-search procedure described in Section 14.1, a terminal synchronizes to a cell, acquires the physical-layer identity of the cell, and detects the cell frame timing. Once this has been achieved, the terminal has to acquire the cell system information. This is information that is repeatedly broadcast by the network and which needs to be acquired by terminals in order for them to be able to access and, in general, operate properly within the network and within a specific cell. The system information includes, among other things, information about the downlink and uplink cell bandwidths, the uplink/downlink configuration in the case of TDD, detailed parameters related to random-access transmission, etc.

In LTE, system information is delivered by two different mechanisms relying on two different transport channels:

- A limited amount of system information, corresponding to the so-called Master-Information Block (MIB), is transmitted using the BCH.
- The main part of the system information, corresponding to different so-called System-Information Blocks (SIBs), is transmitted using the downlink shared channel (DL-SCH).

It should be noted that system information in both the MIB and the SIBs corresponds to the BCCH logical channel. Thus, as also illustrated in Figure 8.7, BCCH can be mapped to both BCH and DL-SCH depending on the exact BCCH information.

14.2.1 Master information block and BCH transmission

As mentioned above, the MIB transmitted using BCH consists of a very limited amount of system information, mainly such information that is absolutely needed for a terminal to be

able to read the remaining system information provided using DL-SCH. More specifically, the MIB includes the following information:

- Information about the downlink cell bandwidth. Three bits are available within the MIB to indicate the downlink bandwidth. Thus, up to eight different bandwidths, measured in number of resource blocks, can be defined for each frequency band.
- Information about the PHICH configuration of the cell. As mentioned in Section 10.4.2, the terminal must know the PHICH configuration to be able to receive the L1/L2 control signaling on PDCCH. The PDCCH information, in turn, is needed to acquire the remaining part of the system information which is carried on the DL-SCH (see further below). Thus, information about the PHICH configuration (three bits) is included in the MIB—that is, transmitted using BCH, which can be received and decoded without first receiving any PDCCH.
- The System Frame Number (SFN) or, more exactly, all bits except the two least significant bits of the SFN are included in the MIB. As described below, the terminal can indirectly acquire the two least significant bits of the SFN from the BCH decoding.

BCH physical-layer processing, such as channel coding and resource mapping, differs quite substantially from the corresponding processing and mapping for DL-SCH outlined in Chapter 10.

As can be seen in Figure 14.4, one BCH transport block, corresponding to the MIB, is transmitted every 40 ms. The BCH Transmissions Time Interval (TTI) thus equals 40 ms.

The BCH relies on a 16-bit CRC, in contrast to a 24-bit CRC used for all other downlink transport channels. The reason for the shorter BCH CRC is to reduce the relative CRC overhead, having the very small BCH transport-block size in mind.

BCH channel coding is based on the same rate-1/3 tail-biting convolutional code as is used for the PDCCH control channel. The reason for using convolutional coding for BCH, rather than the Turbo code used for all other transport channels, is the small size of the BCH transport block. With such small blocks, tail-biting convolutional coding actually outperforms Turbo coding. The channel coding is followed by rate matching, in practice repetition of the coded bits, and bit-level scrambling. QPSK modulation is then applied to the coded and scrambled BCH transport block.

BCH multi-antenna transmission is limited to transmit diversity—that is, SFBC in the case of two antenna ports and combined SFBC/FSTD in the case of four antenna ports. Actually, as mentioned in Chapter 10, if two antenna ports are available within the cell, SFBC must be used for BCH. Similarly, if four antenna ports are available, combined SFBC/FSTD must be used. Thus, by blindly detecting what transmit-diversity scheme is used for BCH, a terminal can indirectly determine the number of cell-specific antenna ports within the cell and also the transmit-diversity scheme used for the L1/L2 control signaling.

As can also be seen from Figure 14.4, the coded BCH transport block is mapped to the first subframe of each frame in four consecutive frames. However, as can be seen in Figure 14.5 and in contrast to other downlink transport channels, the BCH is not mapped on

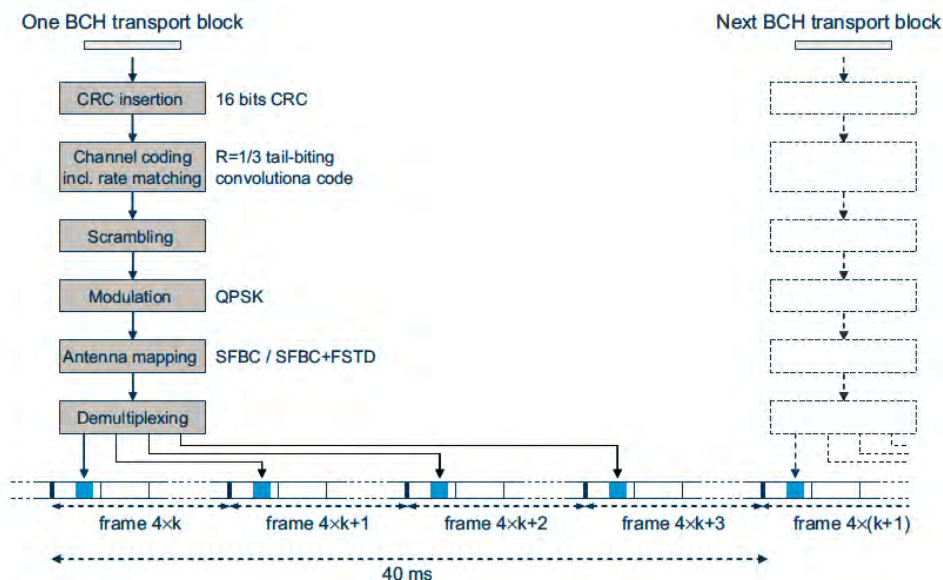


FIGURE 14.4

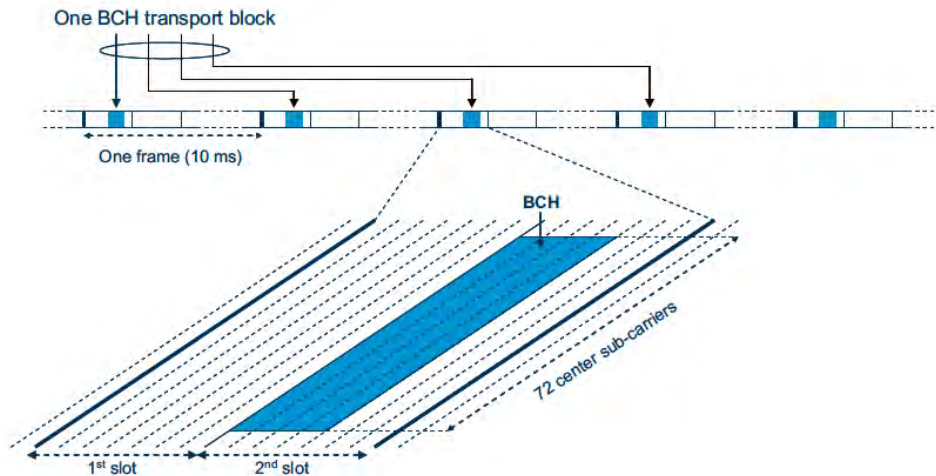
Channel coding and subframe mapping for the BCH transport channel

a resource-block basis. Instead, the BCH is transmitted within the first four OFDM symbols of the second slot of subframe 0 and only over the 72 center subcarriers.³ Thus, in the case of FDD, BCH follows immediately after the PSS and SSS in subframe 0. The corresponding resource elements are then not available for DL-SCH transmission.

The reason for limiting the BCH transmission to the 72 center subcarriers, regardless of the cell bandwidth, is that a terminal may not know the downlink cell bandwidth when receiving BCH. Thus, when first receiving BCH of a cell, the terminal can assume a cell bandwidth equal to the minimum possible downlink bandwidth—that is, six resource blocks corresponding to 72 subcarriers. From the decoded MIB, the terminal is then informed about the actual downlink cell bandwidth and can adjust the receiver bandwidth accordingly.

The total number of resource elements to which the coded BCH is mapped is very large compared to the size of the BCH transport block, implying extensive repetition coding or, equivalently, massive processing gain for the BCH transmission. Such large processing gain is needed as it should be possible to receive and correctly decode the BCH also by terminals in neighboring cells, implying potentially very low receiver Signal-to-Interference-and-Noise

³Not including the DC carrier.

**FIGURE 14.5**

Detailed resource mapping for the BCH transport channel

Ratio (SINR) when decoding the BCH. At the same time, many terminals will receive BCH in much better channel conditions. Such terminals then do not need to receive the full set of four subframes over which a BCH transport block is transmitted to acquire sufficient energy for correct decoding of the transport block. Instead, already by receiving only a few or perhaps only a single subframe, the BCH transport block may be decodable.

From the initial cell search, the terminal has found only the cell frame timing. Thus, when receiving BCH, the terminal does not know to what set of four subframes a certain BCH transport block is mapped. Instead, a terminal must try to decode the BCH at four possible timing positions. Depending on which decoding is successful, indicated by a correct CRC check, the terminal can implicitly determine 40 ms timing or, equivalently, the two least significant bits of the SFN.⁴ This is the reason why these bits do not need to be explicitly included in the MIB.

14.2.2 System-information blocks

As already mentioned, the MIB on the BCH only includes a very limited part of the system information. The main part of the system information is instead included in different System-Information Blocks (SIBs) that are transmitted using the DL-SCH. The presence of

⁴BCH scrambling is defined with 40 ms periodicity, hence even if the terminal successfully decodes the BCH after observing only a single transmission instant, it can determine the 40 ms timing.

system information on DL-SCH in a subframe is indicated by the transmission of a corresponding PDCCH marked with a special System-Information RNTI (SI-RNTI). Similar to the PDCCH providing the scheduling assignment for “normal” DL-SCH transmission, this PDCCH also indicates the transport format and physical resource (set of resource blocks) used for the system-information transmission.

LTE defines a number of different SIBs characterized by the type of information that is included within them:

- SIB1 includes information mainly related to whether a terminal is allowed to camp on the cell. This includes information about the operator/operators of the cell, if there are restrictions with regards to what users may access the cell, etc. SIB1 also includes information about the allocation of subframes to uplink/downlink and configuration of the special subframe in the case of TDD. Finally, SIB1 includes information about the time-domain scheduling of the remaining SIBs (SIB2 and beyond).
- SIB2 includes information that terminals need in order to be able to access the cell. This includes information about the uplink cell bandwidth, random-access parameters, and parameters related to uplink power control.
- SIB3 mainly includes information related to cell reselection.
- SIB4–SIB8 include neighboring-cell-related information, including information related to neighboring cells on the same carrier, neighboring cells on different carriers, and neighboring non-LTE cells, such as WCDMA/HSPA, GSM, and CDMA2000 cells.
- SIB9 contains the name of the home-eNodeB.
- SIB10–SIB12 contain public warning messages, for example earthquake information.
- SIB13 contains information necessary for MBMS reception (see also Chapter 17).
- SIB14 is used to provide enhanced access barring information, controlling the possibilities for terminals to access the cell.
- SIB15 contains information necessary for MBMS reception on neighboring carrier frequencies.
- SIB16 contains information related to GPS time and Coordinated Universal Time (UTC).

Not all the SIBs need to be present. For example, SIB9 is not relevant for an operator-deployed node and SIB13 is not necessary if MBMS is not provided in the cell.

Similar to the MIB, the SIBs are broadcasted repeatedly. How often a certain SIB needs to be transmitted depends on how quickly terminals need to acquire the corresponding system information when entering the cell. In general, a lower-order SIB is more time critical and is thus transmitted more often compared to a higher-order SIB. SIB1 is transmitted every 80 ms, whereas the transmission period for the higher-order SIBs is flexible and can be different for different networks.

The SIBs represent the basic system information to be transmitted. The different SIBs are then mapped to different System-Information messages (SIs), which correspond to the actual transport blocks to be transmitted on DL-SCH. SIB1 is always mapped, by itself, on to the first system-information message SI-1,⁵ whereas the remaining SIBs may be group-wise multiplexed on to the same SI subject to the following constraints:

- The SIBs mapped to the same SI must have the same transmission period. Thus, as an example, two SIBs with a transmission period of 320 ms can be mapped to the same SI, whereas an SIB with a transmission period of 160 ms must be mapped to a different SI.
- The total number of information bits that is mapped to a single SI must not exceed what is possible to transmit within a transport block.

It should be noted that the transmission period for a given SIB might be different in different networks. For example, different operators may have different requirements concerning the period when different types of neighboring-cell information needs to be transmitted. Furthermore, the amount of information that can fit into a single transport block very much depends on the exact deployment situation, such as cell bandwidth, cell size, and so on.

Thus, in general, the SIB-to-SI mapping for SIBs beyond SIB1 is flexible and may be different for different networks or even within a network. An example of SIB-to-SI mapping is illustrated in Figure 14.6. In this case, SIB2 is mapped to SI-2 with a transmission period of 160 ms. SIB3 and SIB4 are multiplexed into SI-3 with a transmission period of 320 ms, whereas SIB5, which also requires a transmission period of 320 ms, is mapped to a separate SI (SI-4). Finally, SIB6, SIB7, and SIB8 are multiplexed into SI-5 with a transmission period of 640 ms. Information about the detailed SIB-to-SI mapping, as well as the transmission period of the different SIs, is provided in SIB1.

Regarding the more detailed transmission of the different system-information messages, there is a difference between the transmission of SI-1, corresponding to SIB1, and the transmission of the remaining SIs.

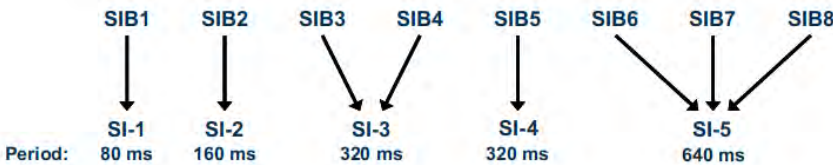


FIGURE 14.6

Example of mapping of SIBs to SIs

⁵Strictly speaking, as SIB1 is not multiplexed with any other SIBs, it is not even said to be mapped to an SI. Rather, SIB1 in itself directly corresponds to the transport block.

The transmission of SI-1 has only a limited flexibility. More specifically, SI-1 is always transmitted within subframe 5. However, the bandwidth or, in general, the set of resource blocks over which SI-1 is transmitted, as well as other aspects of the transport format, may vary and is signaled on the associated PDCCH.

For the remaining SIs, the scheduling on DL-SCH is more flexible in the sense that each SI can, in principle, be transmitted in any subframe within time windows with well-defined starting points and durations. The starting point and duration of the time window of each SI are provided in SIB-1. It should be noted that an SI does not need to be transmitted on consecutive subframes within the time window, as is illustrated in Figure 14.7. Within the time window, the presence of system information in a subframe is indicated by the SI-RNTI on PDCCH, which also provides the frequency-domain scheduling as well as other parameters related to the system-information transmission.

Different SIs have different non-overlapping time windows. Thus, a terminal knows what SI is being received without the need for any specific identifier for each SI.

In case of a relatively small SI and a relatively large system bandwidth, a single subframe may be sufficient for the transmission of the SI. In other cases, multiple subframes may be needed for the transmission of a single SI. In the latter case, instead of segmenting each SI into sufficiently small blocks that are separately channel coded and transmitted in separate subframes, the complete SI is channel coded and mapped to multiple, not necessarily consecutive, subframes.

Similar to the case of the BCH, terminals that are experiencing good channel conditions may then be able to decode the complete SI after receiving only a subset of the subframes to which the coded SI is mapped, while terminals in bad positions need to receive more subframes for proper decoding of the SI. This approach has two benefits:

- Similar to BCH decoding, terminals in good positions need to receive fewer subframes, implying the possibility for reduced terminal power consumption.
- The use of larger code blocks in combination with Turbo coding leads to improved channel-coding gain.

Strictly speaking the single transport block containing the SI is not transmitted over multiple subframes. Rather, the subsequent SI transmissions are seen as autonomous

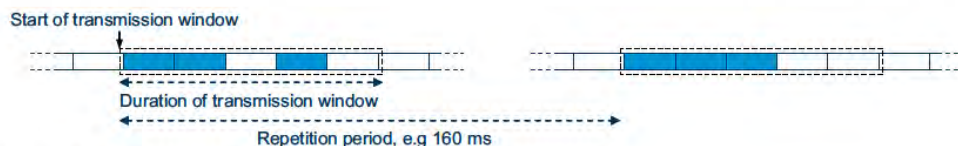


FIGURE 14.7

Transmission window for the transmission of an SI

hybrid-ARQ retransmissions of the first SI transmission—that is, retransmissions taking place without any explicit feedback signaling provided on the uplink.

For terminals capable of carrier aggregation, the system information for the primary component carrier is obtained as described above. For secondary component carriers, the terminal does not need to read the system-information blocks but assumes that the information obtained for the primary component carrier also holds for the secondary component carriers. System information specific for the secondary component carrier is provided through dedicated RRC signaling as part of the procedure to configure an additional secondary component carrier. Using dedicated signaling instead of reading the system information on the secondary component carrier enables faster activation of secondary component carriers as the terminal otherwise would have to wait until the relevant system information had been transmitted.

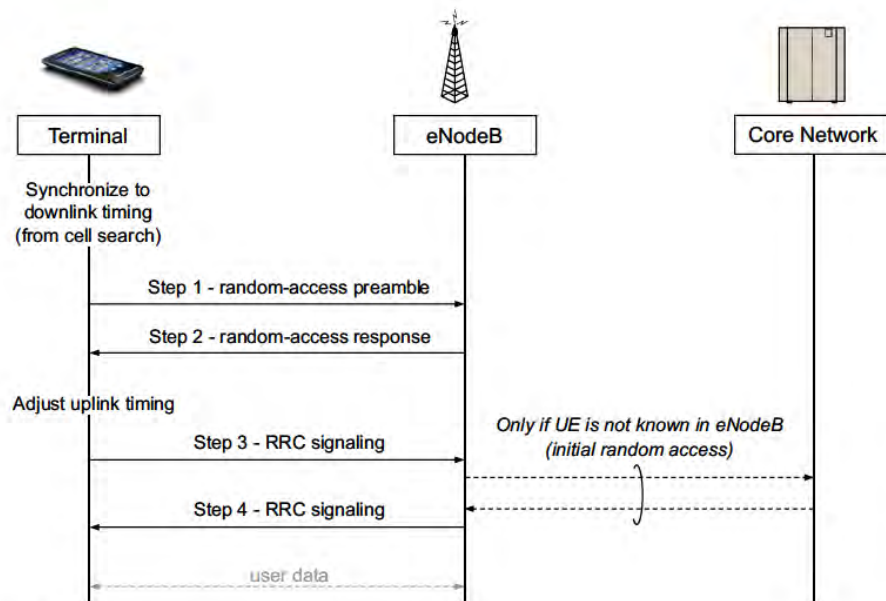
14.3 Random access

A fundamental requirement for any cellular system is the possibility for the terminal to request a connection setup, commonly referred to as random access. In LTE, random access is used for several purposes, including:

- for initial access when establishing a radio link (moving from RRC_IDLE to RRC_CONNECTED; see Chapter 8 for a discussion on different terminal states);
- to re-establish a radio link after radio-link failure;
- for handover when uplink synchronization needs to be established to the new cell;
- to establish uplink synchronization if uplink or downlink data arrives when the terminal is in RRC_CONNECTED and the uplink is not synchronized;
- for the purpose of positioning using positioning methods based on uplink measurements;
- as a scheduling request if no dedicated scheduling-request resources have been configured on PUCCH (see Chapter 13 for a discussion on uplink scheduling procedures).

Acquisition of uplink timing is a main objective for all the cases above; when establishing an initial radio link (that is, when moving from RRC_IDLE to RRC_CONNECTED), the random-access procedure also serves the purpose of assigning a unique identity, the C-RNTI, to the terminal.

Either a contention-based or a contention-free scheme can be used, depending on the purpose. Contention-based random access can be used for all previously discussed purposes, while contention-free random access can only be used for re-establishing uplink synchronization upon downlink data arrival, uplink synchronization of secondary component carriers,

**FIGURE 14.8**

Overview of the random-access procedure

handover, and positioning. The basis for the random access is the four-step procedure illustrated in Figure 14.8, with the following steps:

1. The transmission of a random-access preamble, allowing the eNodeB to estimate the transmission timing of the terminal. Uplink synchronization is necessary as the terminal otherwise cannot transmit any uplink data.
2. The network transmits a timing advance command to adjust the terminal transmit timing, based on the timing estimate obtained in the first step. In addition to establishing uplink synchronization, the second step also assigns uplink resources to the terminal to be used in the third step in the random-access procedure.
3. The transmission of the mobile-terminal identity to the network using the UL-SCH similar to normal scheduled data. The exact content of this signaling depends on the state of the terminal, in particular whether it is previously known to the network or not.
4. The final step consists of transmission of a contention-resolution message from the network to the terminal on the DL-SCH. This step also resolves any contention due to multiple terminals trying to access the system using the same random-access resource.

Only the first step uses physical-layer processing specifically designed for random access. The subsequent three steps utilize the same physical-layer processing as used for normal uplink and downlink data transmission. In the following, each of these steps is described in more detail. Only the first two steps of the preceding procedure are used for contention-free random access as there is no need for contention resolution in a contention-free scheme.

Both the terminal and the network can initiate a random-access attempt. In the latter case, RRC signaling or a so-called PDCCH order is used. A PDCCH order is a specific message transmitted on the PDCCH, containing information about when to initiate the random-access procedure and, in case of a contention-free random access, the preamble to use. PDCCH orders are primarily intended as a tool for the network to re-establish uplink synchronization but can be used for other purposes as well. A terminal may perform random access on its primary component carrier only⁶, except for establishing uplink timing alignment for a secondary timing advance group.

14.3.1 Step 1: Random-access preamble transmission

The first step in the random-access procedure is the transmission of a random-access preamble. The main purpose of the preamble transmission is to indicate to the base station the presence of a random-access attempt and to allow the base station to estimate the delay between the eNodeB and the terminal. The delay estimate will be used in the second step to adjust the uplink timing.

The time–frequency resource on which the random-access preamble is transmitted is known as the Physical Random-Access Channel (PRACH). The network broadcasts information to all terminals in which time–frequency resource random-access preamble transmission is allowed (that is, the PRACH resources), in SIB-2. As part of the first step of the random-access procedure, the terminal selects one preamble to transmit on the PRACH.

In each cell, there are 64 preamble sequences available. Two subsets of the 64 sequences are defined as illustrated in Figure 14.9, where the set of sequences in each subset is signaled as part of the system information. When performing a (contention-based) random-access

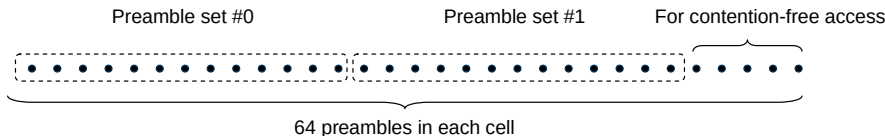


FIGURE 14.9

Preamble subsets

⁶The primary component carrier is terminal-specific as already discussed; hence, from an eNodeB perspective, random access may occur on multiple component carriers.

attempt, the terminal selects at random one sequence in one of the subsets. As long as no other terminal is performing a random-access attempt using the same sequence at the same time instant, no collisions will occur and the attempt will, with a high likelihood, be detected by the eNodeB.

The subset to select the preamble sequence from is given by the amount of data the terminal would like to (and from a power perspective can) transmit on the UL-SCH in the third random-access step. Hence, from the preamble the terminal used, the eNodeB will get some guidance on the amount of uplink resources to be granted to the terminal.

If the terminal has been requested to perform a contention-free random access, for example for handover to a new cell, the preamble to use is explicitly indicated from the eNodeB. To avoid collisions, the eNodeB should preferably select the contention-free preamble from sequences outside the two subsets used for contention-based random access.

14.3.1.1 PRACH time–frequency resources

In the frequency domain, the PRACH resource, illustrated in Figure 14.10, has a bandwidth corresponding to six resource blocks (1.08 MHz). This nicely matches the smallest uplink cell bandwidth of six resource blocks in which LTE can operate. Hence, the same random-access preamble structure can be used, regardless of the transmission bandwidth in the cell.

In the time domain, the length of the preamble region depends on the configured preamble, as will be discussed further below. The basic random-access resource is 1 ms in duration, but it is also possible to configure longer preambles. Also, note that the eNodeB uplink scheduler in principle can reserve an arbitrary long-random-access region by simply avoiding scheduling terminals in multiple subsequent subframes.

Typically, the eNodeB avoids scheduling any uplink transmissions in the time–frequency resources used for random access, resulting in the random-access preamble being orthogonal to user data. This avoids interference between UL-SCH transmissions and random-access attempts from different terminals. However, from a specification perspective, nothing prevents the uplink scheduler from scheduling transmissions in the random-access region. Hybrid-ARQ retransmissions are examples of this; synchronous non-adaptive hybrid-ARQ

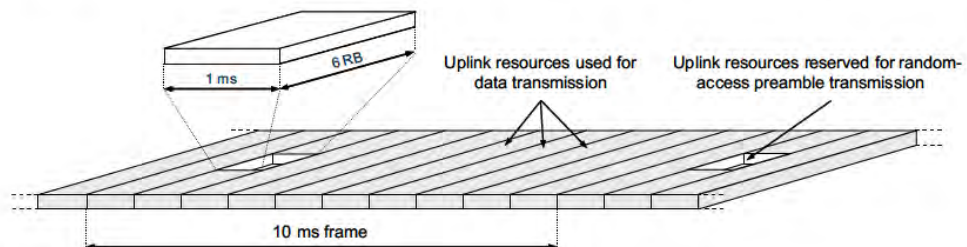


FIGURE 14.10

Principal illustration of random-access preamble transmission

retransmissions may overlap with the random-access region and it is up to the implementation to handle this, either by moving the retransmissions in the frequency domain as discussed in Chapter 12 or by handling the interference at the eNodeB receiver.

For FDD, there is at most one random-access region per subframe—that is, multiple random-access attempts are not multiplexed in the frequency domain. From a delay perspective, it is better to spread out the random-access opportunities in the time domain to minimize the average waiting time before a random-access attempt can be initialized.

For TDD, multiple random-access regions can be configured in a single subframe. The reason is the smaller number of uplink subframes per radio frame in TDD. To maintain the same random-access capacity as in FDD, frequency-domain multiplexing is sometimes necessary. The number of random-access regions is configurable and can vary from one per 20 ms to one per 1 ms for FDD; for TDD up to six attempts per 10 ms radio frame can be configured.

14.3.1.2 Preamble structure and sequence selection

The preamble consists of two parts:

- Preamble sequence.
- Cyclic prefix.

Furthermore, the preamble transmission uses a guard period to handle the timing uncertainty. Prior to starting the random-access procedure, the terminal has obtained downlink synchronization from the cell-search procedure. However, as uplink synchronization has not yet been established prior to random access, there is an uncertainty in the uplink timing⁷ as the location of the terminal in the cell is not known. The uplink timing uncertainty is proportional to the cell size and amounts to $6.7 \mu\text{s}/\text{km}$. To account for the timing uncertainty and to avoid interference with subsequent subframes not used for random access, a guard time is used as part of the preamble transmission—that is, the length of the actual preamble is shorter than 1 ms.

Including a cyclic prefix as part of the preamble is beneficial as it allows for frequency-domain processing at the base station (discussed further below), which can be advantageous from a complexity perspective. Preferably, the length of the cyclic prefix is approximately equal to the length of the guard period. With a preamble sequence length of approximately 0.8 ms, there is 0.1 ms cyclic prefix and 0.1 ms guard time. This allows for cell sizes up to 15 km and is the typical random-access configuration, configuration 0 in Figure 14.11. To handle larger cells, where the timing uncertainty is larger, preamble configurations 1–3 can be used. Some of these configurations also support a longer preamble sequence to increase the preamble energy at the detector, which can be beneficial in larger cells. The preamble configuration used in a cell is signaled as part of the system information.

⁷The start of an uplink frame at the terminal is defined relative to the start of a downlink frame received at the terminal.

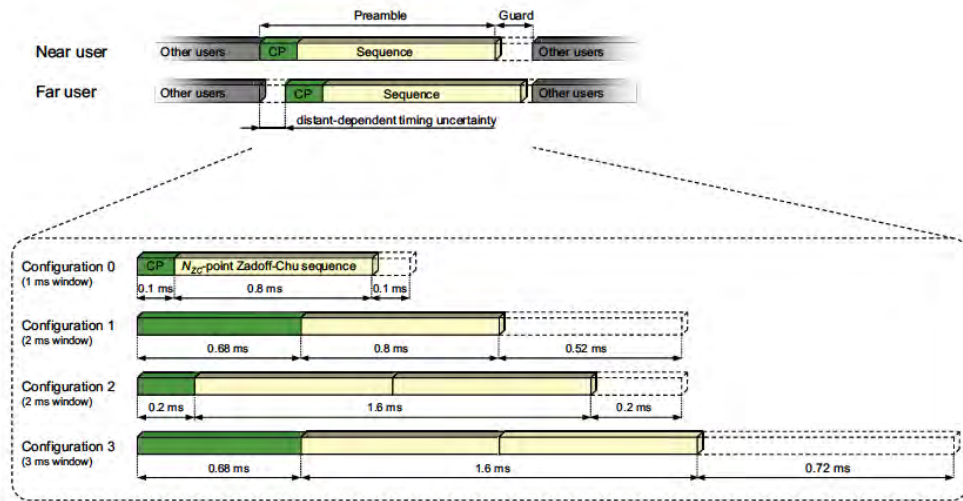


FIGURE 14.11

Different preamble formats

Finally, note that guard times larger than those in Figure 14.11 can easily be created by not scheduling any uplink transmissions in the subframe following the random-access resource.

The preamble formats in Figure 14.11 are applicable to both FDD and TDD. However, for TDD, there is an additional fourth preamble configuration for random access. In this configuration, the random-access preamble is transmitted in the UpPTS field of the special subframe instead of in a normal subframe. Since this field is at most two OFDM symbols long, the preamble and the possible guard time are substantially shorter than the preamble formats described above. Hence, format 4 is applicable to very small cells only. The location of the UpPTS, next to the downlink-to-uplink switch for TDD, also implies that the interference from distant base stations may interfere with this short random-access format, which limits its usage to small cells and certain deployment scenarios.

14.3.1.3 PRACH power setting

The basis for setting the transmission power of the random-access preamble is a downlink path-loss estimate obtained from measuring the cell-specific reference signals on the primary downlink component carrier. From this path-loss estimate, the initial PRACH transmission power is obtained by adding a configurable offset.

The LTE random-access mechanism allows power ramping where the actual PRACH transmission power is increased for each unsuccessful random-access attempt. For the first attempt, the PRACH transmission power is set to the initial PRACH power. In most cases, this

is sufficient for the random-access attempts to be successful. However, if the random-access attempt fails (random-access failures are detected at the second of four random-access steps, as described in the following sections), the PRACH transmission power for the next attempt is increased by a configurable step size to increase the likelihood of the next attempt being successful.

Since the random-access preamble is orthogonal to the user data, the need for power ramping to control intra-cell interference is smaller than in other systems with non-orthogonal random access and in many cases the transmission power is set such that the first random-access attempt with a high likelihood is successful. This is beneficial from a delay perspective.

14.3.1.4 Preamble sequence generation

The preamble sequences are generated from cyclic shifts of root Zadoff–Chu sequences [74]. Zadoff–Chu sequences are also used for creating the uplink reference signals as described in Chapter 11, where the structure of those sequences is described. From each root Zadoff–Chu sequence $X_{\text{ZC}}^{(u)}(k)$, $\lfloor N_{\text{ZC}}/N_{\text{CS}} \rfloor$ cyclically shifted⁸ sequences are obtained by cyclic shifts of N_{CS} each, where N_{ZC} is the length of the root Zadoff–Chu sequence. The generation of the random-access preamble is illustrated in Figure 14.12. Although the figure illustrates generation in the time domain, frequency-domain generation can equally well be used in an implementation.

Cyclically shifted Zadoff–Chu sequences possess several attractive properties. The amplitude of the sequences is constant, which ensures efficient power amplifier utilization and maintains the low PAR properties of the single-carrier uplink. The sequences also have ideal cyclic auto-correlation, which is important for obtaining an accurate timing estimation at the eNodeB. Finally, the cross-correlation between different preambles based on cyclic shifts of the same Zadoff–Chu root sequence is zero at the receiver, as long as the cyclic shift N_{CS} used when generating the preambles is larger than the maximum round-trip propagation time in the cell plus the maximum delay spread of the channel. Therefore, due to the ideal

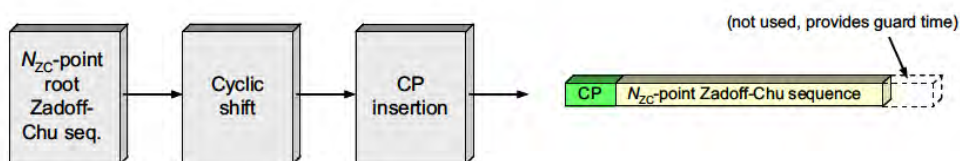


FIGURE 14.12

Random-access preamble generation

⁸The cyclic shift is in the time domain. Similar to the uplink reference signals and control signaling, this can equivalently be described as a phase rotation in the frequency domain.

cross-correlation property, there is no intra-cell interference from multiple random-access attempts using preambles derived from the same Zadoff–Chu root sequence.

To handle different cell sizes, the cyclic shift N_{CS} is signaled as part of the system information. Thus, in smaller cells, a small cyclic shift can be configured, resulting in a larger number of cyclically shifted sequences being generated from each root sequence. For cell sizes below 1.5 km, all 64 preambles can be generated from a single root sequence. In larger cells, a larger cyclic shift needs to be configured and to generate the 64 preamble sequences, multiple root Zadoff–Chu sequences must be used in the cell. Although the larger number of root sequences is not a problem in itself, the zero cross-correlation property only holds between shifts of the same root sequence and from an interference perspective it is therefore beneficial to use as few root sequences as possible.

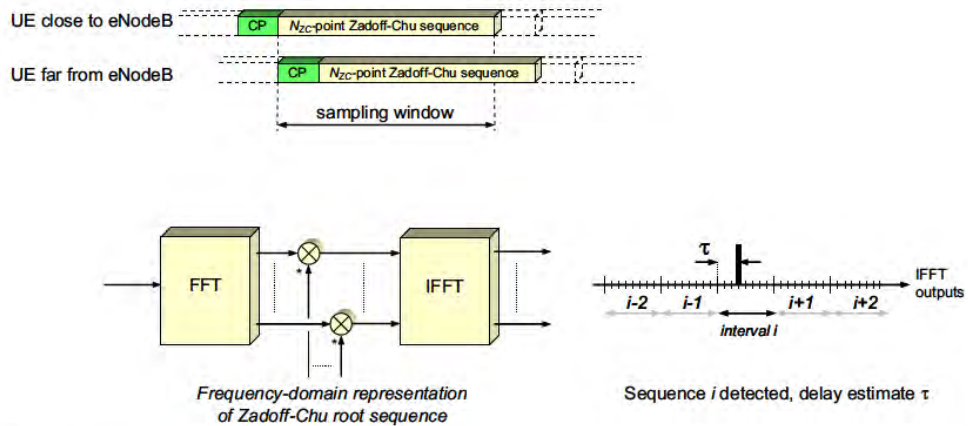
Reception of the random-access preamble is discussed further below, but in principle it is based on correlation of the received signal with the root Zadoff–Chu sequences. One disadvantage of Zadoff–Chu sequences is the difficulties in separating a frequency offset from the distance-dependent delay. A frequency offset results in an additional correlation peak in the time domain, a correlation peak that corresponds to a spurious terminal-to-base-station distance. In addition, the true correlation peak is attenuated. At low-frequency offsets, this effect is small and the detection performance is hardly affected. However, at high Doppler frequencies, the spurious correlation peak can be larger than the true peak. This results in erroneous detection; the correct preamble may not be detected or the delay estimate may be incorrect.

To avoid the ambiguities from spurious correlation peaks, the set of preamble sequences generated from each root sequence can be restricted. Restrictions imply that only some of the sequences that can be generated from a root sequence are used to define random-access preambles. Whether restrictions should be applied or not to the preamble generation is signaled as part of the system information. The location of the spurious correlation peak relative to the “true” peak depends on the root sequence and hence different restrictions have to be applied to different root sequences. The restrictions to apply are broadcasted as part of the system information in the cell.

14.3.1.5 Preamble detection

The base-station processing is implementation specific, but due to the cyclic prefix included in the preamble, low-complexity frequency-domain processing is possible. An example hereof is shown in Figure 14.13. Samples taken in a time-domain window are collected and converted into the frequency-domain representation using an FFT. The window length is 0.8 ms, which is equal to the length of the Zadoff–Chu sequence without a cyclic prefix. This allows handling timing uncertainties up to 0.1 ms and matches the guard time defined for the basic preamble configuration.

The output of the FFT, representing the received signal in the frequency domain, is multiplied by the complex-conjugate frequency-domain representation of the root Zadoff–Chu sequence and the result is fed through an IFFT. By observing the IFFT outputs, it is

**FIGURE 14.13**

Random-access preamble detection in the frequency domain

possible to detect which of the shifts of the root Zadoff-Chu sequence has been transmitted and its delay. Basically, a peak of the IFFT output in interval i corresponds to the i^{th} cyclically shifted sequence and the delay is given by the position of the peak within the interval. This frequency-domain implementation is computationally efficient and allows simultaneous detection of multiple random-access attempts using different cyclically shifted sequences generated from the same root Zadoff-Chu sequence. In the case of multiple attempts there will simply be a peak in each of the corresponding intervals.

14.3.2 Step 2: Random-access response

In response to the detected random-access attempt, the eNodeB will, as the second step of the random-access procedure, transmit a message on the DL-SCH, containing:

- The index of the random-access preamble sequences the network detected and for which the response is valid.
- The timing correction calculated by the random-access preamble receiver.
- A scheduling grant, indicating resources the terminal will use for the transmission of the message in the third step.
- A temporary identity, the TC-RNTI, used for further communication between the terminal and the network.

If the network detects multiple random-access attempts (from different terminals), the individual response messages of multiple terminals can be combined in a single transmission. Therefore, the response message is scheduled on the DL-SCH and indicated on a PDCCH

using an identity reserved for random-access response, the RA-RNTI.⁹ The usage of the RA-RNTI is also necessary, as a terminal may not have a unique identity in the form of a C-RNTI allocated. All terminals that have transmitted a preamble monitor the L1/L2 control channels for random-access response within a configurable time window. The timing of the response message is not fixed in the specification in order to be able to respond to many simultaneous accesses. It also provides some flexibility in the base-station implementation. If the terminal does not detect a random-access response within the time window, the attempt will be declared as failed and the procedure will repeat from the first step again, possibly with an increased preamble transmission power.

As long as the terminals that performed random access in the same resource used different preambles, no collision will occur and from the downlink signaling it is clear to which terminal(s) the information is related. However, there is a certain probability of contention—that is, multiple terminals using the same random-access preamble at the same time. In this case, multiple terminals will react upon the same downlink response message and a collision occurs. Resolving these collisions is part of the subsequent steps, as discussed below. Contention is also one of the reasons why hybrid ARQ is not used for transmission of the random-access response. A terminal receiving a random-access response intended for another terminal will have incorrect uplink timing. If hybrid ARQ were used, the timing of the hybrid-ARQ acknowledgement for such a terminal would be incorrect and may disturb uplink control signaling from other users.

Upon reception of the random-access response in the second step, the terminal will adjust its uplink transmission timing and continue to the third step. If contention-free random access using a dedicated preamble is used, then this is the last step of the random-access procedure as there is no need to handle contention in this case. Furthermore, the terminal already has a unique identity allocated in the form of a C-RNTI.

14.3.3 Step 3: Terminal identification

After the second step, the uplink of the terminal is time synchronized. However, before user data can be transmitted to/from the terminal, a unique identity within the cell, the C-RNTI, must be assigned to the terminal (unless the terminal already has a C-RNTI assigned). Depending on the terminal state, there may also be a need for additional message exchange for setting up the connection.

In the third step, the terminal transmits the necessary messages to the eNodeB using the UL-SCH resources assigned in the random-access response in the second step. Transmitting the uplink message in the same manner as scheduled uplink data instead of attaching it to the preamble in the first step is beneficial for several reasons. First, the amount of information transmitted in the absence of uplink synchronization should be minimized, as the need for a

⁹There are actually several RA-RNTIs defined. Which RA-RNTI a terminal is listening to is given by the time and frequency resource upon which the random-access preamble was transmitted.

large guard time makes such transmissions relatively costly. Secondly, the use of the “normal” uplink transmission scheme for message transmission allows the grant size and modulation scheme to be adjusted to, for example, different radio conditions. Finally, it allows for hybrid ARQ with soft combining for the uplink message. The latter is an important aspect, especially in coverage-limited scenarios, as it allows for the use of one or several retransmissions to collect sufficient energy for the uplink signaling to ensure a sufficiently high probability of successful transmission. Note that RLC retransmissions are not used for the uplink RRC signaling in step 3.

An important part of the uplink message is the inclusion of a terminal identity, as this identity is used as part of the contention-resolution mechanism in the fourth step. If the terminal is in the RRC_CONNECTED state—that is, connected to a known cell and therefore has a C-RNTI assigned—this C-RNTI is used as the terminal identity in the uplink message.¹⁰ Otherwise, a core-network terminal identifier is used and the eNodeB needs to involve the core network prior to responding to the uplink message in step 3.

Terminal-specific scrambling is used for transmission on UL-SCH, as described in Chapter 11. However, as the terminal may not yet have been allocated its final identity, the scrambling cannot be based on the C-RNTI. Instead, a temporary identity is used (TC-RNTI).

14.3.4 Step 4: Contention resolution

The last step in the random-access procedure consists of a downlink message for contention resolution. Note that, from the second step, multiple terminals performing simultaneous random-access attempts using the same preamble sequence in the first step listen to the same response message in the second step, and therefore have the same temporary identifier. Hence, the fourth step in the random-access procedure is a contention-resolution step to ensure that a terminal does not incorrectly use another terminal’s identity. The contention resolution mechanism differs somewhat depending on whether the terminal already has a valid identity in the form of a C-RNTI or not. Note that the network knows from the uplink message received in step three whether the terminal has a valid C-RNTI or not.

If the terminal already had a C-RNTI assigned, contention resolution is handled by addressing the terminal on the PDCCH using the C-RNTI. Upon detection of its C-RNTI on the PDCCH the terminal will declare the random access attempt successful, and there is no need for contention-resolution-related information on the DL-SCH. Since the C-RNTI is unique to one terminal, unintended terminals will ignore this PDCCH transmission.

If the terminal does not have a valid C-RNTI, the contention resolution message is addressed using the TC-RNTI, and the associated DL-SCH contains the contention-resolution message. The terminal will compare the identity in the message with the identity transmitted in the third step. Only a terminal which observes a match between the identity received in the

¹⁰The terminal identity is included as a MAC control element on the UL-SCH.

fourth step and the identity transmitted as part of the third step will declare the random-access procedure successful and promote the TC-RNTI from the second step to the C-RNTI. Since uplink synchronization has already been established, hybrid ARQ is applied to the downlink signaling in this step, and terminals with a match between the identity they transmitted in the third step and the message received in the fourth step will transmit a hybrid-ARQ acknowledgement in the uplink.

Terminals that do not detect PDCCH transmission with their C-RNTI or do not find a match between the identity received in the fourth step and the respective identity transmitted as part of the third step are considered to have failed the random-access procedure and need to restart the procedure from the first step. No hybrid-ARQ feedback is transmitted from these terminals. Furthermore, a terminal that has not received the downlink message in step 4 within a certain time from the transmission of the uplink message in step 3 will declare the random-access procedure as failed and need to restart from the first step.

14.4 Paging

Paging is used for network-initiated connection setup when the terminal is in RRC_IDLE. In LTE, the same mechanism as for “normal” downlink data transmission on the DL-SCH is used and the mobile terminal monitors the L1/L2 control signaling for downlink scheduling assignments related to paging. Since the location of the terminal typically is not known on a cell level, the paging message is typically transmitted across multiple cells in the so-called tracking area (the tracking area is controlled by the MME; see [9] for a discussion on tracking areas).

An efficient paging procedure should allow the terminal to sleep with no receiver processing most of the time and to briefly wake up at predefined time intervals to monitor paging information from the network. Therefore, a paging cycle is defined, allowing the terminal to sleep most of the time and only briefly wake up to monitor the L1/L2 control signaling. If the terminal detects a group identity used for paging (the P-RNTI) when it wakes up, it will process the corresponding downlink paging message transmitted on the PCH. The paging message includes the identity of the terminal(s) being paged, and a terminal not finding its identity will discard the received information and sleep according to the DRX cycle. As the uplink timing is unknown during the DRX cycles, no hybrid-ARQ acknowledgements can be transmitted and consequently hybrid ARQ with soft combining is not used for paging messages.

The network configures in which subframes a terminal should wake up and listen for paging. Typically, the configuration is cell specific, although there is a possibility to complement the setting by terminal-specific configuration. In which frame a given terminal should wake up and search for the P-RNTI on a PDCCH is determined by an equation taking as input the identity of the terminal as well as a cell-specific and (optionally) a terminal-specific paging cycle. The identity used is the so-called IMSI, an identity coupled to the

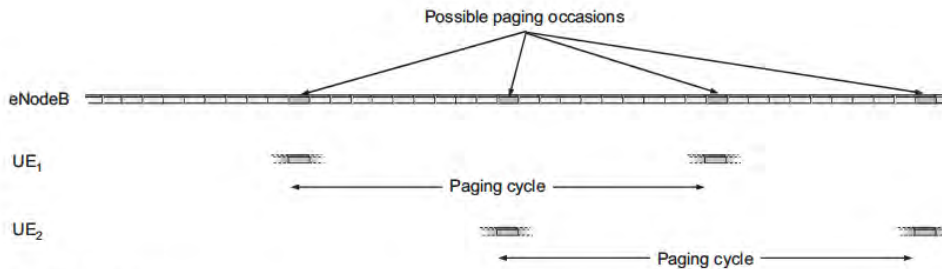


FIGURE 14.14

Illustration of paging cycles

		Number of Paging Subframes per Paging Cycle								
		1/32	1/16	1/8	1/4	1/2	1	2	4	
Paging subframes in a paging frame	FDD	9	9	9	9	9	9	4, 9	0, 4, 5, 9	
	TDD	0	0	0	0	0	0	0, 5	0, 1, 5, 6	

subscription, as an idle mode terminal does not have a C-RNTI allocated, and the paging cycle for a terminal can range from once per 256 up to once per 32 frames. The subframe within a frame to monitor for paging is also derived from the IMSI. Since different terminals have different IMSI, they will compute different paging instances. Hence, from a network perspective, paging may be transmitted more often than once per 32 frames, although not all terminals can be paged at all paging occasions as they are distributed across the possible paging instances, as shown in Figure 14.14.

Paging messages can only be transmitted in some subframes, ranging from one subframe per 32 frames up to a very high paging capacity with paging in four subframes in every frame. The configurations are shown in Table 14.1. Note that, from a network perspective, the cost of a short paging cycle is minimal as resources not used for paging can be used for normal data transmission and are not wasted. However, from a terminal perspective, a short paging cycle increases the power consumption as the terminal needs to wake up frequently to monitor the paging instants.

In addition to initiating connection to terminals being in RRC_IDLE, paging can also be used to inform terminals in RRC_IDLE as well as RRC_CONNECTED about changes of system information. A terminal being paged for this reason knows that the system information will change and therefore needs to acquire the update system information as described in Section 14.2.

Multi-Point Coordination and Transmission

The principle of *spatial reuse* lies at the core of any radio-communication system. Spatial reuse implies that the same communication resource, in the LTE case the same time-frequency resource, is simultaneously used for communication at different, spatially separated locations.

Inherently, transmissions carried out on the same time-frequency resource will cause interference to each other. To limit this interference, early cellular technologies relied on a static frequency separation between neighbor cells. As an example, in a *reuse-3* deployment the overall set of available frequency resources is divided into three groups. As illustrated in Figure 15.1, only frequencies of one of these groups are then used within a given cell, with frequencies of other groups being used in the most neighboring cells. Interference from transmissions in the most neighboring cells is then avoided, and each communication link will experience a relatively high signal-to-interference ratio (SIR) regardless of the terminal position within the coverage area.

However, for modern mobile-broadband technologies such as LTE, which should be able to provide very high end-user data rates when the channel conditions so allow, this is not a good approach. Being hard-limited to only a fraction of the overall available spectrum at a given transmission point would reduce the maximum achievable transmission bandwidth that can be used at the transmission point, with a corresponding reduction in the maximum achievable data rates as a consequence.¹

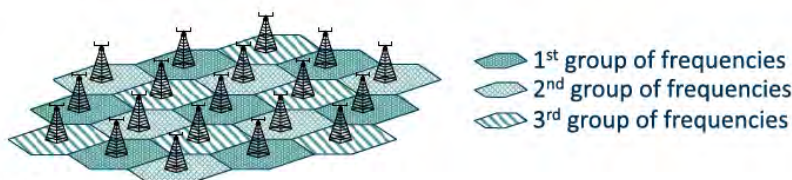


FIGURE 15.1

Reuse-3 deployment

¹Instead of “cell” we here use the more general term “(network) transmission point.” For a homogeneous deployment, which is the focus of this chapter, one can typically assume that each “transmission point” corresponds to a “cell.” For the case of heterogeneous deployments, which is the topic of the next chapter, the distinction will in some cases be more important.

Even more important, in a mobile broadband system dominated by highly dynamic packet-data traffic there is frequently no data available for transmission at a given transmission point. Having statically assigned a part of the overall available spectrum to that transmission point with no possibility to use the corresponding frequency resources to provide higher instantaneous transmission capacity at neighboring transmission points would imply an inefficient use of the available spectrum. Rather, in order to maximize system efficiency, as well as to enable as high as possible end-user data rates, a mobile broadband system should be deployed such that, fundamentally, all frequency resources are available for use at each transmission point.

At the same time, for the specific case of transmission to a terminal close to the border between the coverage areas of two transmission points (see Figure 15.2), end-user quality as well as overall efficiency would be further improved if interference from the neighboring transmission point could be avoided.

Thus, even if all frequency resources should fundamentally be available for use at each transmission point, coordination across the transmission points can be beneficial. Such coordination could, for example, imply avoiding transmission, or transmitting with lower power or in a different direction (beam-forming), on a certain time-frequency resource in order to reduce the interference to terminals served by other, neighboring transmission points if such a terminal would otherwise experience severe interference.

In certain cases, one could even consider using both transmission points for transmission to the same terminal. This would not only avoid interference from the neighboring transmission point, but would also boost the overall signal power available for transmission to the terminal.

The preceding discussion implicitly assumed downlink transmission, with terminal reception being interfered by downlink transmissions from other network transmission points. However, the concept of coordination between network points as a means to better control the interference levels is applicable also to the uplink, although the interference situation in this case is somewhat different.

For the uplink, the interference level experienced by a certain link does not depend on where the transmitting terminal is located but rather on the location of the *interfering* terminals, with interfering terminals closer to the border between two, in this case, network



FIGURE 15.2

Downlink interference to terminal close to the border between two transmission points

**FIGURE 15.3**

Uplink interference from a terminal close to the border between two reception points

reception *points* causing more interference to the neighboring reception point (see Figure 15.3). Still, the fundamental goal of uplink coordination is the same as for the downlink—that is, to avoid the most severe interference situations.

One can envision two main deployment scenarios when considering coordination between network points:

- Coordination within a *homogeneous deployment*, primarily between nodes in a macro deployment.
- Coordination within a *heterogeneous deployment*, for example between a macro node and under-laid lower-power nodes.

The focus of this chapter will be on the first type of coordination—coordination within homogeneous deployments. Heterogeneous deployments will create additional challenges and corresponding need for coordination between transmission points. This will be further discussed in the next chapter as part of a more general discussion on heterogeneous deployments.

One factor that will impact the possibilities for coordination between network points is the available backhaul connectivity and especially its associated latency. Highly dynamic coordination requires low-latency connectivity between the points to be coordinated. One case when this is definitely available is when the points to be coordinated correspond to sectors of the same site (“intra-site coordination”). However, very-low-latency connectivity may be available also in the case of geographically separated transmission points, especially if there are direct physical links (for example, optical or wireless links), between the points. In other cases, only not-so-low-latency inter-node connectivity (for example, connectivity with latency in the order of several tens of milliseconds or more), may be available, in which case one is limited to less dynamic coordination.

The 3GPP activities related to coordination between network points for LTE can be divided into two phases:

- Release-8 activities on “*inter-cell interference coordination*” (ICIC), primarily focusing on inter-eNB (X2) signaling to assist such coordination.
- Releases-10/11 activities on multi-point coordination/transmission targeting more dynamic coordination and focusing on new radio-interface features and terminal capabilities to enable/improve such coordination.

15.1 Release-8 inter-cell interference coordination

The potential gains of coordination between transmission/reception points were extensively discussed in the early phase of LTE standardization with focus on coordination within homogeneous macro deployments. More specifically, the focus was on defining X2 signaling that could be used to enhance such coordination between cells corresponding to different eNodeBs.² As the X2 interface is typically associated with not-so-low latency, the focus of the release-8 activities was on relatively slow coordination.

In the case of scheduling located at a higher-level node above the eNodeB, coordination between cells of different eNodeB would, at least conceptually, be straightforward as it could be carried out at the higher-level node. However, in the LTE radio-network architecture there is no higher-level node defined, and scheduling is assumed to be carried out at the eNodeB. Thus, the best that can be done from an LTE specification point-of-view is to introduce messages that convey information about the local scheduling strategy/status between neighboring eNodeBs. An eNodeB can then use the information provided by neighboring eNodeBs as input to its own scheduling process. It is important to understand though that the LTE specifications do not specify how an eNodeB should react to this information. Rather, this is up to scheduler implementation.

To assist uplink interference coordination, two X2 messages were defined as part of LTE release 8: the *High Interference Indicator* (HII) and the *Overload Indicator* (OI) (see Figure 15.4).

The *High Interference Indicator* provides information about the set of resource blocks within which an eNodeB has high sensitivity to interference. Although nothing is explicitly specified on how an eNodeB should react to the HII (or any other ICIC-related X2 signaling)

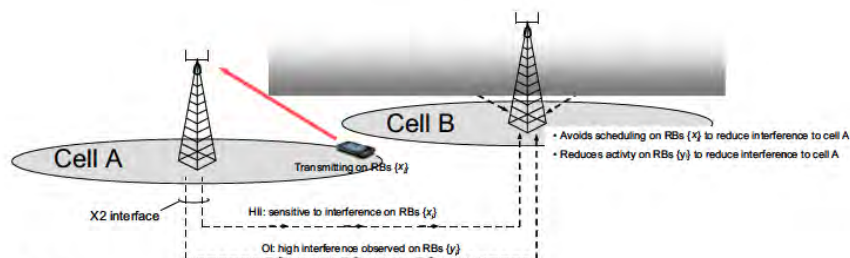


FIGURE 15.4

Illustration of uplink ICIC based on the HII and OI X2 signaling

²As the focus was on X2—that is, inter-eNodeB—signaling the transmission/reception points relevant for the coordination would inherently correspond to different cells.

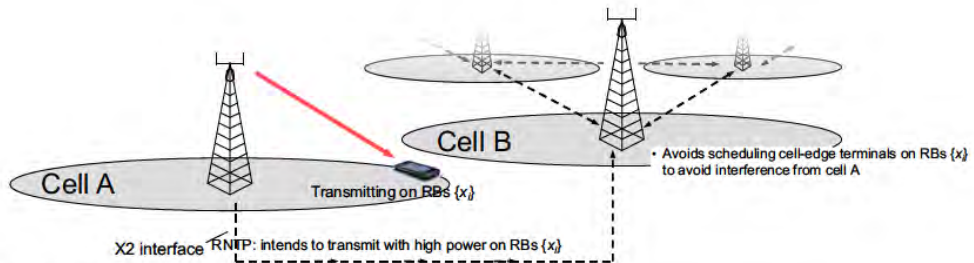
**FIGURE 15.5**

Illustration of downlink ICIC based on RNTP X2 signaling

received from a neighboring eNodeB, a reasonable action for the receiving eNodeB would be to try to avoid scheduling its own cell-edge terminals on the same resource blocks, thereby reducing the uplink interference to cell-edge transmissions in its own cell as well as in the cell from which the HII was received. The HII can thus be seen as a *proactive* tool for ICIC, trying to prevent the occurrence of *too-low-SIR* situations.

In contrast to the HII, the *Overload Indicator* (OI) is a *reactive* ICIC tool, essentially indicating, at three levels (Low/Medium/High), the uplink interference experienced by a cell on its different resource blocks. A neighboring eNodeB receiving the OI could then change its scheduling behavior to improve the interference situation for the eNodeB issuing the OI.

For the downlink, the *Relative Narrowband Transmit Power* (RNTP) was defined to support ICIC operation (see Figure 15.5). The RNTP is similar to the HII in the sense that it provides information, for each resource block, whether or not the relative transmit power of that resource block is to exceed a certain level. Similar to the HII, a neighboring cell can use the information provided by the received RNTP when scheduling its own terminals, especially terminals on the cell edge that are more likely to be interfered by the neighboring cell.

15.2 Releases-10/11 multi-point coordination/transmission

During the work on LTE release 10, the possibility for more dynamic coordination between network points was discussed under the term *Coordinated Multi Point* (CoMP) transmission/reception. Although initially discussed as part of the 3GPP activities on LTE release 10, the main features related to CoMP were introduced into the LTE specifications as part of release 11.

A main difference between the LTE release-8 ICIC activities described previously and the releases-10/11 CoMP activities is that the latter focused on radio-interface features and terminal functionality to assist different coordination means. At the same time, there were no discussions on specific inter-eNodeB signaling to support CoMP. Rather, there was an

assumption that low-latency backhaul was available for the coordination, in practice limiting release-11 CoMP features to either sectors of the same site or network points connected by direct low-latency links. There was also an implicit assumption that the different network points involved in the coordination were tightly synchronized and time aligned with each other.

The different approaches to CoMP considered for the LTE downlink can be divided into two main groups:

- Schemes where transmission is carried out from a specific transmission point but where the scheduling and link adaptation may be coordinated between transmission points. We will refer to this as *multi-point coordination*.
- Schemes where transmission to a terminal may be carried out from different transmission points (*multi-point transmission*). The transmission can then either switch dynamically between the different transmission points or be carried out jointly from multiple points.

A similar distinction can be made for the uplink transmission direction in which case one would distinguish between (uplink) multi-point coordination where the uplink scheduling is coordinated between different reception points, and *multi-point reception* where reception may be carried out at multiple points. It can now be pointed out that, at least from a radio-interface point-of-view, uplink multi-point coordination/reception is very much a network implementation issue with very little impact on the terminal and very little visibility in the radio-interface specifications. The discussions in [Sections 15.2.1 and 15.2.2](#) will focus on coordination in the downlink transmission direction. [Section 15.2.3](#) will briefly discuss some aspects of uplink multi-point coordination/reception.

15.2.1 Multi-point coordination

As described above, multi-point coordination implies that transmission is carried out from a specific transmission point but functions such as link adaptation and/or scheduling are coordinated between multiple points.

15.2.1.1 Coordinated link adaptation

Link adaptation—that is, the dynamic selection of data rate based on estimates/predictions of the instantaneous channel conditions to be experienced by a transmission—is one of the basic mechanisms for good system performance in LTE. Good link adaptation relies on the availability of good predictions of the interference level to be experienced by the transmission. However, in case of highly dynamic traffic conditions, the traffic activity of neighbor transmission points may vary rapidly. As a consequence, the interference level may also vary rapidly and in an (apparently) unpredictable way.

Coordinated link adaptation is about using information about transmission decisions of neighboring transmission points in the link-adaptation process—that is, in the decision with

what data rate to transmit on a given resource. Note that this implies a multi-step process in the scheduling and link adaptation at transmission points:

1. For a given subframe, transmission points carry out transmission decisions. In the simplest case this may be decisions on whether or not to transmit data on a certain set of time-frequency resources—that is, a certain set of resource blocks within the subframe. In a more general case it may also include, for example, decisions on transmission power and/or beam-forming decisions—in what “direction” to transmit—for the given set of resources.
2. Information about the transmission decisions are shared between neighboring transmission points.
3. Transmission points use the information about transmission decisions of neighboring transmission points as input to the link-adaption decision for the transmission(s) to take place in the given subframe.

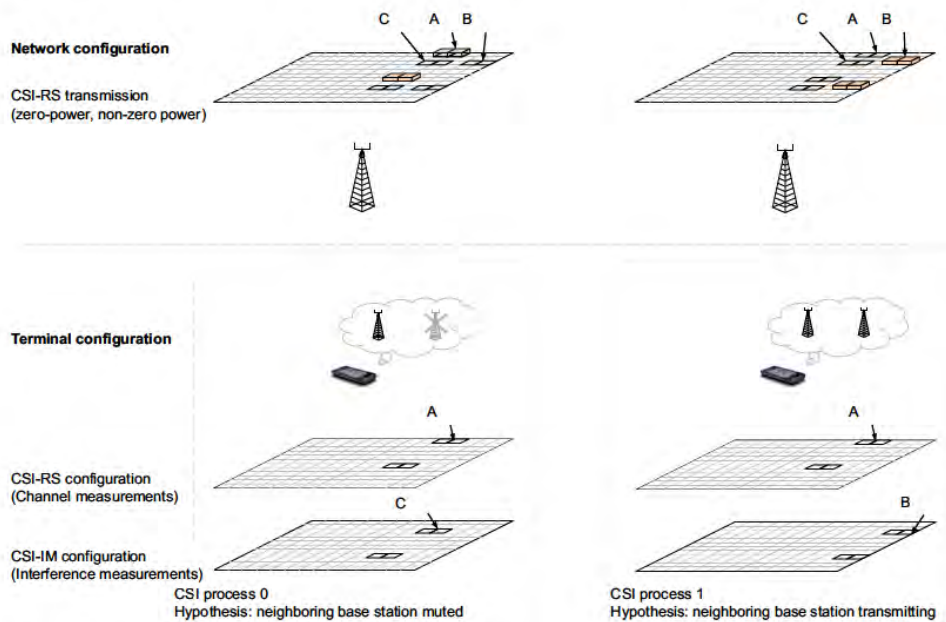
In LTE, link adaptation is carried out at the network side. However, as described in Chapter 13, the network typically bases the link-adaptation decisions on CSI reports provided by the terminals. To enable coordinated link adaptation—that is, to allow for the network to take information about the transmission decisions made by neighboring transmissions into account in the rate selection—the terminal should provide multiple CSI reports corresponding to different *hypotheses* regarding the transmission decisions of neighboring transmission points. These CSI reports can then be used together with information about the actual transmission decisions of neighboring transmission points in the link adaptation.

In order for a terminal to be able to provide CSI reports corresponding to different hypotheses regarding the transmission decisions of neighboring transmission points, it should be configured with *multiple CSI processes*. As described in Chapter 13, each such process would correspond to a set of CSI-RS, one for each antenna port, and one CSI-IM resource for interference estimation. In order to support coordinated link adaptation, the set of CSI-RS should be the same for all processes and reflect the channel of the different antenna port(s) for the transmission point from which transmission is to be carried out. In contrast, the CSI-IM resources of the different CSI processes should be different and configured in such a way that they reflect the interference to be expected for different hypotheses regarding the transmission decisions of neighboring transmission points.

As an example, Figure 15.6 illustrates the case of coordinated link adaptation between two transmission points. The figure also illustrates three different CSI-RS resources on which there are either transmission, corresponding to ordinary (non-zero-power) CSI-RS, or no transmission, corresponding to zero-power CSI-RS, for the two transmission points.

For a terminal associated with the left side transmission point, two CSI processes are configured:

- Process 0 with CSI-RS corresponding to resource A and CSI-IM corresponding to resource C (configured as zero-power CSI-RS at the neighboring transmission point).

**FIGURE 15.6**

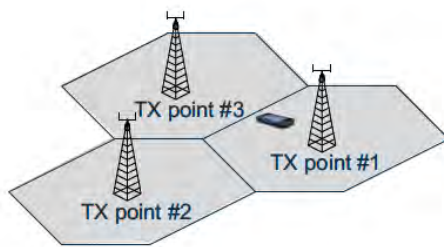
Example of using multiple CSI processes

CSI-reporting by this CSI process will thus reflect the channel state under the hypothesis that there is no transmission from the neighboring transmission point.

- Process 1 with CSI-RS corresponding to resource A (same as for process 0) and CSI-IM corresponding to resource B (configured as non-zero-power CSI-RS at the neighboring transmission point). CSI reported by this process will thus reflect the channel state under the hypothesis that there is transmission from the neighboring transmission point.

The CSI reports delivered by the terminal for the different CSI processes would thus correspond to the different hypotheses regarding the transmission decision of the neighboring transmission point. Based on information regarding the expected transmission from the neighboring transmission point, the network can select the appropriate CSI report and use that in the link-adaptation decision.

Coordinated link adaptation can also be carried out between more than two transmission points. As an example, consider a case where one would like to carry out coordinated link adaptation between three different transmission points (Figure 15.7). In this case there is a



Resource	TX point #1	TX point #2	TX point #3
A	ZP CSI-RS	ZP CSI-RS	ZP CSI-RS
B	ZP CSI-RS	ZP CSI-RS	CSI-RS
C	ZP CSI-RS	CSI-RS	ZP CSI-RS
D	ZP CSI-RS	CSI-RS	CSI-RS
E	CSI-RS	ZP CSI-RS	ZP CSI-RS
F	CSI-RS	ZP CSI-RS	CSI-RS
G	CSI-RS	CSI-RS	ZP CSI-RS

FIGURE 15.7

CSI-RS/IM structure to support coordinated link adaption between three transmission points

need for a total of seven CSI-RS resources, labeled A to G, configured as non-zero-power and zero-power CSI-RS at the different transmission points, according to Figure 15.7.

A terminal associated with transmission point 1 should, in this case, be configured with four CSI processes, the CSI-IM resources of which would correspond to resources A to D in Figure 15.7. Measurements on these four CSI-IM resources would provide interference predictions that would correspond to different hypotheses regarding the transmission decisions of neighboring points. More specifically:

- Measurements on the CSI-IM resource corresponding to resource A would provide an interference prediction corresponding to the hypothesis that there is no transmission from neither transmission point 2 nor transmission point 3 at the time of transmission.
- Measurements on the CSI-IM resource corresponding to resource B would provide an interference prediction corresponding to the hypothesis that there is transmission from transmission point 2 but not from transmission point 3.
- Measurements on the CSI-IM resource corresponding to resource C would provide an interference prediction corresponding to the hypothesis that there is transmission from transmission point 3 but not from transmission point 2.
- Finally, measurements on the CSI-IM resource corresponding to resource D would provide an interference prediction corresponding to the hypothesis that there is transmission from both transmission point 2 and transmission point 3.

Similarly, a terminal associated with transmission point 2 would be configured with CSI processes, the CSI-IM of which would correspond to resource A, B, E, and F in Figure 15.7. In this case, as an example, measurements on the CSI-IM resource corresponding to resource E would provide an interference prediction corresponding to the hypothesis that there is transmission from transmission point 1 but not from transmission point 3.

Likewise, a terminal associated with transmission point 3 would be configured with CSI processes, the CSI-IM of which would correspond to resources A, C, E, and G in Figure 15.7.

15.2.1.2 Coordinated scheduling

Dynamic link adaption as described above is about using information about the transmission decisions made by neighboring transmission points in the link-adaptation decision—that is, in the selection of transmission rate to be used by a transmission point. Dynamic link adaptation is applicable and useful regardless of whether or not the actual transmission decisions are coordinated between the transmission points.

Coordinated scheduling is about coordinating the actual transmission decision(s) between transmission points. Thus, while coordinated link adaptation is about sharing of information between transmission points for better predictions of the interference levels, coordinated scheduling is about sharing of information and coordination between transmission points to reduce and control the actual interference levels.

In its most simple case, coordinated scheduling is about dynamically preventing transmission at a certain time-frequency resource in order to reduce the interference to be experienced by a terminal served by a neighboring transmission point. In the LTE CoMP discussions, this has been referred to as *Dynamic Point Blanking*. In the more general case it can also involve dynamically adjusting the transmit power (*Coordinated Power Control*) or dynamically adjusting the transmission direction (*Coordinated Beam-Forming*) for a specific set of resources.

In order to enable dynamic point blanking the network should be able to estimate/predict the impact to a terminal, in terms of expected channel quality, of transmissions from a neighboring transmission point and also be able to predict how much the channel quality would improve if transmissions from the neighboring transmission point would not take place. To enable this, multiple CSI processes configured in the same way as for coordinated link adaptation, discussed earlier, may be used. The different CSI processes provide different CSI reports reflecting different hypotheses regarding the transmission at the neighboring transmission points. By comparing these CSI reports, the network can estimate how much would be gained from blanking relevant time/frequency resources at a neighboring transmission point.

As an example, consider a terminal associated with transmission point 1 in the example scenario of Figure 15.7. If there would be a large difference in the CQI—that is, the recommended data rate—of the CSI reports corresponding to resources B and D, this would be an indication that the terminal would be severely interfered by transmissions from transmission point 2 and it would be relevant to consider blanking of relevant time-frequency resources at that transmission point to improve the experienced channel quality and, as a consequence, the data rate that can be supported, for the terminal associated with transmission point 1.

On the other hand, if it would be a very small difference in the CQI of the CSI reports corresponding to resources B and D, this would be an indication that the terminal is not severely interfered by transmissions from transmission point 2 and, at least in terms of channel quality for this specific terminal, it is not beneficial to apply blanking to transmission point 2.

15.2.2 Multi-point transmission

Multi-point coordination as described previously implies that transmissions carried out from neighboring transmission points are coordinated in terms of scheduling (if/when to transmit) and/or link adaption (with what rate to transmit). However, the transmission to a given terminal is still assumed to be carried out from one specific transmission point. In contrast, in case of *multi-point transmission*, the transmission to a given terminal can be carried out from different transmission points, either so that the point of transmission can change dynamically, referred to as *Dynamic Point Selection*, or so that the transmission can be carried out jointly from multiple transmission points, referred to as *Joint Transmission*.

15.2.2.1 Dynamic point selection

As mentioned earlier, dynamic point selection implies transmission from a single transmission point but where the point of transmission can be changed dynamically, as illustrated in Figure 15.8.

In the LTE context, dynamic point selection, and actually all CoMP schemes, are assumed to be based on the use of transmission mode 10. Thus, in case of dynamic point selection PDSCH transmission relies on DM-RS for channel estimation. As a consequence, the terminal does not need to be aware of the change of transmission point. What the terminal will see is simply a PDSCH transmission, the instantaneous channel of which may change abruptly as the point-of-transmission is changed. In essence, from a terminal point-of-view, the situation would be identical to that of beam-forming based on non-codebook-precoding.³

To assist in the dynamic selection of transmission point, a terminal should provide CSI reports corresponding to multiple transmission points. Similar to coordinated link adaptation and coordinated scheduling this may be achieved by configuring the terminal with multiple CSI processes.

As just described, in the case of coordinated link adaptation and coordinated scheduling the different CSI processes should correspond to the same transmission point—that is, the set of CSI-RS should be the same for the different processes. At the same time, the CSI-IM



FIGURE 15.8

Dynamic point selection between two transmission points

³There would be a difference in what the terminal can assume in terms of quasi-co-location, as further discussed in the following.

resources should be different for the different processes, allowing for the interference measurements and thus also the CSI reports to reflect different hypotheses regarding the transmission decisions of the neighboring transmission points.

In contrast, to support dynamic point selection the different CSI processes should provide CSI reports corresponding to different transmission points. Thus, the set of CSI-RS of the different processes should be different and correspond to CSI-RS transmitted by the different transmission points between which the dynamic point selection is carried out.

In addition to CSI reporting, the other main specification impact of dynamic point selection relates to PDSCH mapping and what a terminal can assume in terms of quasi-co-location relation between different reference signals.

In the normal case, PDSCH mapping to the resource blocks assigned for transmission avoids the resource elements used for CRS transmission within the serving cell of the terminal. However, in case of dynamic point selection the PDSCH transmission to a terminal may be carried out from a transmission point associated with a different cell than the serving cell. If this cell has a different CRS structure, in terms of number of CRS and/or CRS frequency shift, and the PDSCH mapping remained according to the CRS structure of the serving cell, CRS transmissions from the actual point of transmission would be severely interfered by the PDSCH transmission. Rather, in case of dynamic point selection, one would like the PDSCH mapping to dynamically match the CRS structure of the actual point-of-transmission.

A similar situation may arise for L1/L2 control signaling. The size of the control region, and thus the starting point for PDSCH transmission, of a certain cell may vary dynamically with information about the size of the control region provided to the terminal by means of the PCFICH (see Section 10.4). However, in case of dynamic point selection, if the size of the control region of the actual point-of-transmission differs from that of the serving cell and the PDSCH mapping remained according to the PCFICH of the serving cell, the L1/L2 control signaling of the actual point-of-transmission would run the risk of being severely interfered by the PDSCH transmission. Thus, similar to the case of CRS one would like the PDSCH mapping to dynamically match the size of the control region of the actual point-of-transmission.

Furthermore, as described in Chapter 10, a terminal may be configured with a set of zero-power CSI-RS resources. From a terminal point-of-view the zero-power CSI-RS simply defines a set of resource elements to which PDSCH is not mapped, typically because these resources elements are used for other purposes, for example, as CSI-RS for other terminals or as CSI-IM resources. However, if the PDSCH transmission is carried out from a different point one would typically like the PDSCH mapping to avoid a different set of resource elements as this transmission point would typically use different resource elements for CSI-RS and CSI-IM.

To handle these related problems in a unified way, transmission mode 10 allows for the dynamic reconfiguration of the PDSCH mapping by means of a *PDSCH-mapping-and-quasi-co-location indicator* provided as part of the downlink scheduling assignment, more specifically as part of DCI format 2D as described in Section 10.4.

A terminal can be provided with up to four different *PDSCH mapping and quasi-co-location configurations*.⁴ Each such configuration specifies:

- a specific CRS configuration in terms of number of CRS and CRS frequency shift;
- a specific PDSCH starting point;
- a specific MBSFN subframe configuration;⁵
- a specific zero-power CSI-RS configuration.

The *PDSCH-mapping-and-quasi-co-location* indicator provided in the scheduling assignment then explicitly indicates which one of the up to four different configurations the terminal should assume for the PDSCH mapping for the corresponding subframe.

It should be noted that the PDSCH starting point indicator does not guarantee a perfect match with the size of the control region of the actual point-of-transmission as the size of the control region can vary dynamically. The PDSCH starting point indication should be set to a sufficiently large value to guarantee that the PDSCH transmission does not overlap with the control region. As the size of the control region never exceeds three OFDM symbols (corresponding to control signaling in OFDM symbols 0, 1, and 2), the most straightforward way of achieving this would be to set the PDSCH starting point to 3.⁶ However, lower values can be used, allowing for a larger PDSCH payload, if one knows that, in a certain cell, the control region will always be limited to a lower value.

As the name suggests, the *PDSCH-mapping-and-quasi-co-location* configurations and the corresponding indicator in the scheduling assignment also provide information about what the terminal can assume in terms of quasi-co-location relation between antenna ports. As discussed in Chapter 10, for transmission mode 1 to 9, a terminal can assume that the antenna ports corresponding to CRS of the serving cell, DM-RS, and the CSI-RS configured for the terminal are all jointly quasi-co-located. However, in case of dynamic point selection, a terminal may be configured with different sets of CSI-RS by means of different CSI processes. In practice these sets of CSI-RS correspond to different transmission points. Eventually, the PDSCH together with its corresponding DM-RS will be transmitted from one of these transmission points. The antenna ports used for the PDSCH transmission (antenna ports 7 up to 14) will then, in practice, be quasi-co-located with the set of CSI-RS corresponding to that specific transmission point. To provide this information to the terminal, which is not explicitly aware of from what transmission point the PDSCH transmission takes place, each *PDSCH mapping and quasi-co-location* configuration also indicates a specific set of CSI-RS for which the terminal can assume quasi-co-location with the DM-RS for the specific subframe.

⁴Note the difference between the *PDSCH-mapping-and-quasi-co-location configuration* provided by means of higher layer signaling and the *PDSCH-mapping-and-quasi-co-location indicator* provided by the scheduling assignment.

⁵The MBSFN subframe configuration is related to the CRS configuration as it impacts the presence of CRS reference symbols in the data part of the subframe.

⁶In case of the smallest LTE cell bandwidths, the size of the control region could be up to four OFDM symbols.



FIGURE 15.9

Joint transmission from two transmission points to the same terminal

15.2.2.2 Joint transmission

While dynamic point selection implies transmission from a single transmission point but where the point of transmission can be changed dynamically, joint transmission implies the possibility for simultaneous transmission from multiple transmission points to the same terminal (see Figure 15.9).

In case of joint transmission one can distinguish between two cases:

- Coherent joint transmission.
- Non-coherent joint transmission.

In case of coherent joint transmission it is assumed that the network has knowledge about the detailed channels to the terminal from the two or more points involved in the joint transmission and selects transmission weights accordingly, for example, to focus the energy at the position of the terminal. Thus, coherent joint transmission can be seen as a kind of beam-forming for which the antennas taking part in the beam-forming are not co-located, but correspond to different transmission points.

There is currently no support in the LTE specifications for the terminal to report this kind of detailed channel knowledge for multiple transmission points, and thus there is currently no explicit support for coherent joint transmission.

In contrast, for *non-coherent* joint transmission it is assumed that the network does not make use of any such detailed channel knowledge in the joint transmission. Thus, the only gain of non-coherent joint transmission is that the power of multiple transmission points is used for transmission to the same terminal—in practice a power gain. The benefit of this depends on whether or not the power of the second transmission point can be of better use for the transmission to other terminals and also to what extent the extra transmission will cause harmful interference to other transmissions. In practice, non-coherent joint transmission is only beneficial at low-load situations where there is no other terminal available for which the second transmission point can be used and where the additional interference from the second transmission does not really cause any harm.

It should be noted that, in case of joint transmission, PDSCH may be jointly transmitted from points corresponding to two cells with, for example, different CRS configurations or different MBSFN configurations. In such a case one would like for the PDSCH mapping to

match the configurations of both cells. However, currently each *PDSCH-mapping-and-quasi-co-location* configuration only corresponds to a single CRS configuration.

15.2.3 Uplink multi-point coordination/reception

Sections 15.2.1 and 15.2.2 focused on *downlink* multi-point coordination/transmission. However, as already mentioned, the same basic principles are also possible for transmissions in the uplink direction (*uplink CoMP*). More specifically:

- Dynamic coordination of uplink transmissions in order to control uplink interference and achieve improved uplink system performance (uplink multi-point *coordination*).
- Reception of uplink transmissions at multiple points (uplink multi-point *reception* or uplink *joint reception*).

However, in contrast to downlink multi-point coordination/transmission, uplink multi-point coordination/reception has very little impact on the radio-interface specifications. Especially, any channel-state information needed for uplink scheduling coordination would be directly derived at the network (reception) side and would not require any specific terminal feedback.

Also, a terminal does not need to be aware at what point its uplink transmission is received as long as any downlink transmission corresponding to the uplink transmission (such as hybrid-ARQ feedback) is transmitted in a, for the terminal, expected way. In practice this means that even if the uplink is received at a point different from the transmission point associated with the serving cell, feedback such as hybrid-ARQ acknowledgements must still be transmitted from the transmission point of the serving cells. This would require the availability of low-latency connectivity between the reception and transmission points to ensure, for example, that the hybrid-ARQ timing relations are retained. As already mentioned, the 3GPP releases-10/11 CoMP discussions focused on the situation of low-latency connectivity between transmission/reception points.

Some aspects of the radio-interface design have taken the possibility for multi-point reception into account. For example, uplink multi-point reception was the main reason for introducing support for terminal-specific assignment of uplink reference-signal sequences, as described in Section 11.2.

This page intentionally left blank

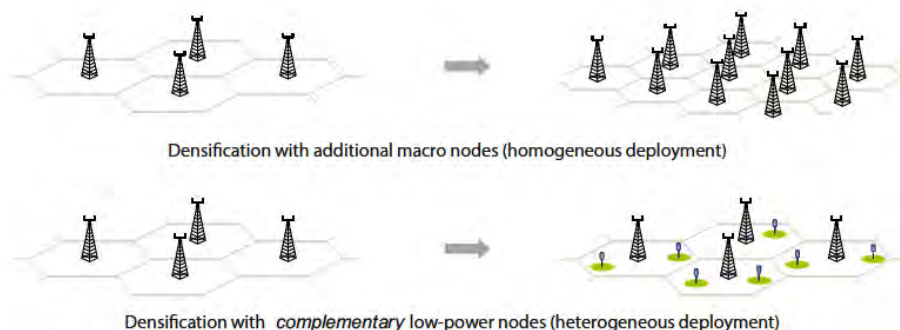
Heterogeneous Deployments

16

The continuous increase in traffic within mobile-broadband systems and an equally continuous increase in terms of the data rates requested by end-users will impact how cellular networks are deployed in the future. In general, providing very high system capacity (traffic per area unit) and very high per-user data rates will require a densification of the radio-access network, that is, the deployment of additional network nodes (or transmission/reception points). By increasing the number of nodes, the traffic per area unit can be increased without requiring a corresponding increase in the traffic that needs to be supported per network node. Also, by increasing the number of network nodes, the base-station-to-terminal distances will, in general, be shorter, implying a link-budget improvement and a corresponding improvement in achievable data rates.

A uniform densification of the macro-cell layer—that is, reducing the coverage area of each cell and increasing the total number of macro-cell sites¹—as illustrated at the top of Figure 16.1 is a path that has already been taken by many operators. As an example, in many major cities the distance between macro-cell sites is less than a few hundred meters in many cases.

An alternative or complement to a uniform densification of the macro layer is to deploy additional lower-power nodes under the coverage area of the macro layer, as illustrated at the

**FIGURE 16.1**

Homogeneous vs. heterogeneous densification

¹A macro node is herein defined as a high-power node with its antennas typically located above rooftop level.

bottom of Figure 16.1. In scenarios where the users are highly clustered, for example gathered around a square or in a shopping mall, such a strategy can be advantageous. The result is a *heterogeneous* or *multi-layered* network deployment, where the underlaid layer of low-power pico nodes² does not need to provide full-area coverage. Rather, low-power nodes can be deployed to increase capacity and achievable data rates where needed. Outside of the pico-layer coverage, terminals would access the network by means of the overlaid macro layer. Although this chapter assumes LTE being used in both the macro and pico layers, another possibility is to use WLAN based on the 802.11 family of standards to complement the macro layer. The LTE specifications provide support for WLAN interworking, including seamless as well as non-seamless mobility, at the core network level.

The idea of multiple deployment layers is in itself not new; “hierarchical cell structures” have been used since the mid-’90s. However, with extensive use of mobile broadband, the interest in heterogeneous deployments as a mean to increase capacity and end-user data rates has increased significantly.

It is important to point out that this is a deployment strategy, not a technology component, and as such is already possible in the first release of LTE (as well as in other technologies). Nevertheless, releases 10 and 11 provide additional features improving the support for heterogeneous deployments, in particular in the area of handling inter-layer interference.

16.1 Interference scenarios in heterogeneous deployments

One distinctive property of a heterogeneous deployment is the large difference in transmit power between the overlaid macro layer and the underlaid pico layer. Depending on the scenario, this may result in significantly more complex-interference scenarios compared to a homogeneous network, more specifically interference between the layers. *Inter-layer interference handling* is therefore a crucial aspect in most heterogeneous deployments.

If different frequency resources, in particular different frequency bands, are used for the different layers, the inter-layer interference is less of an issue. Frequency separation is also the traditional way of handling inter-layer interference in, for example, GSM, where different carrier frequencies are used in the different layers. However, for a wideband radio-access technology such as LTE, using different carrier frequencies for different layers may lead to an undesirable spectrum fragmentation, as discussed in the previous chapter. As an example, for an operator having access to 20 MHz of spectrum, a static frequency separation between two layers would imply that the total available spectrum had to be divided, with less than 20 MHz of spectrum being available in each layer. This could reduce the maximum achievable data rates within each layer. Also, assigning a substantial part of the overall available spectrum to a layer during periods of relatively low traffic may lead to inefficient

²The term “pico node” is herein used to denote a low-power node, typically with the antennas located below rooftop level.

spectrum utilization. Thus, with a wideband high-data-rate system such as LTE, it should preferably be possible to deploy a multi-layered network with the same spectrum being available in all layers. This is in line with the motivation for single-frequency reuse in Chapter 15. Nevertheless, separate spectrum allocation for the two layers is a relevant scenario, especially if new spectrum becomes available at very high frequencies less suitable for wide area coverage. Furthermore, frequency-separated deployments imply that the duplex scheme can be chosen independently between the layers, for example using FDD in the wide-area macro layer and TDD in the local-area pico layer.

Simultaneous use of the same spectrum in different layers implies inter-layer interference. The characteristics of the inter-layer interference depend on the transmission power in the respective layer, as well as in the node-association strategy used.

Traditionally, node association or cell association—that is, determining which network point the terminal should be connected to—is based on terminal measurements of the received power of some downlink signal, more specifically the cell-specific reference signals in the case of LTE. Based on the terminal reporting those measurements to the network, the network decides whether a handover should take place or not. This is a simple and robust approach. In homogeneous deployments with all transmission points having the same power, downlink measurements reflect the uplink path loss, and downlink-optimized network-point association is also reasonable from an uplink perspective. However, in a heterogeneous deployment, this approach can be challenged due to the large difference in transmission power between the layers. In principle, the best uplink reception point is not necessarily the best downlink transmission point, implying that uplink and downlink points ideally should be determined separately. For example, downlink point selection could be based on received signal strength, while uplink point selection preferably is based on the lowest path loss. This is illustrated in Figure 16.2 where the “cell border” is different for uplink and downlink. However, since there are tight and time-critical dependencies between the uplink and downlink—for example in the form of hybrid-ARQ acknowledgements transmitted in the downlink as a result of receiving uplink transmissions—the two links are in practice terminated in the same node.³ Node association is therefore a compromise between the best choice from an uplink and downlink perspective.

From a single link perspective, associating the terminal with the transmission point with the highest received power implies that the terminal often is connected to a higher-power macro point even if the path loss to a pico point is significantly smaller. This will not be optimal from an uplink coverage and capacity point of view. It should also be noted that, even in terms of downlink system efficiency, it may not be optimal to select the transmission point with the highest received power in a heterogeneous network deployment. Although transmissions from the macro point are received with higher power than from a pico point, this is at least partly due to the higher transmit power of the macro point. In that case, transmission

³The downlink transmission point and the uplink reception point could be geographically separated if remote antennas are used (see Section 16.5).

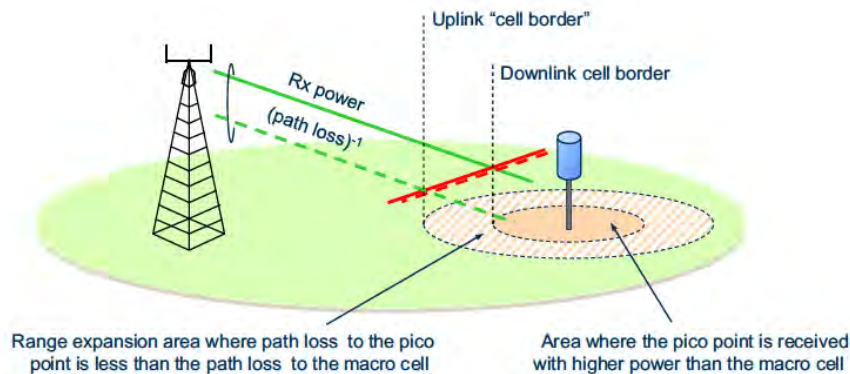
**FIGURE 16.2**

Illustration of high-interference area in a heterogeneous network deployment with range expansion

from the macro point is associated with a higher “cost” in terms of interference to other cells. Expressed alternatively, a transmission from the macro point will prohibit the use of the same physical resource in *any* of the underlaid pico points.

Alternatively, at the other extreme, node association could be based solely on estimates of the (uplink) path loss. In practice this can be achieved by applying an offset to the received-power measurements used in conventional cell association, an offset that would compensate for the difference in transmit power between different transmission points. Such an offset is supported by LTE already from the first release and possible to configure on a per-terminal basis. By using an offset in the node-association strategy, the area in which a pico point is selected is extended, as illustrated in Figure 16.2. This is sometimes referred to as *range expansion*.

Selecting the network point to which the path loss is the smallest—that is, applying a large degree of range expansion—would maximize the uplink received power/SINR, thus maximizing the achievable uplink data rates. Alternatively, for a given target received power, the terminal transmit-power, and thus the interference to other cells, would be reduced, leading to higher overall uplink system efficiency. Also, it could allow for the same downlink physical resource to also be used by other pico points, thereby also improving downlink system efficiency.

However, due to the difference in transmit power between the transmission points of the different deployment layers, downlink transmissions from the macro point will be received with substantially higher power in the range expansion area (illustrated by the dashed region in Figure 16.2) than the actual desired downlink transmission from the pico point. Within this area, there is thus potential for severe downlink inter-cell interference from the macro point to terminals receiving transmissions from a pico point—that is, the signal-to-interference ratio

at the terminal for any signals transmitted from the pico point is below 0 dB. The interference has both a static load-independent component stemming from the cell-specific reference signals (CRS), synchronization signals (PSS, SSS) and system information (PBCH), and a dynamic load-dependent component stemming from data transmissions (PDSCH) and control signaling (PCFICH, PHICH, PDCCH, EPDCCH).

The interference from PDSCH transmissions in the macro layer to lower-power PDSCH transmissions from a pico point can be relatively straightforwardly handled by scheduling coordination between the nodes according to the same principles as inter-cell interference coordination described in Section 15.1. As an example, an overlaid macro point could simply avoid high-power PDSCH transmission in resource blocks in which a terminal in the range expansion region of a pico point is to receive downlink data transmission. Such coordination can be more or less dynamic depending on to what extent and on what time scale the overlaid and underlaid nodes can be coordinated. The same coordination as for the PDSCH could also be applied to the EPDCCH. It should also be noted that, for a pico network point located on the border between two macro cells, it may be necessary to coordinate scheduling between the pico point and both macro cells.

Less obvious is how to handle interference due to the macro-node transmissions that cannot use dynamic inter-cell interference coordination, such as the L1/L2 control signaling (PDCCH, PCFICH, and PHICH). Within a layer, for example between two macro nodes, interference between such transmissions is not a critical issue as LTE, including its control channels, has been designed to allow for one-cell frequency reuse and a corresponding SIR down to and even below -5 dB. This inherent robustness allows for moderate range expansion, in the order of a couple of dB. In many scenarios this amount of range expansion is adequate and further increasing it would not improve performance, while in other scenarios a larger amount of range expansion may be useful. Using a large amount of range expansion can result in a signal-to-interference ratio that is too low for the control channels to operate correctly and calling for means to handle this interference situation. Note that the usefulness of range expansion is highly scenario dependent. One simplistic example where range expansion may not be useful is illustrated in the right part of Figure 16.3, where the building walls provide isolation between the two cell layers.

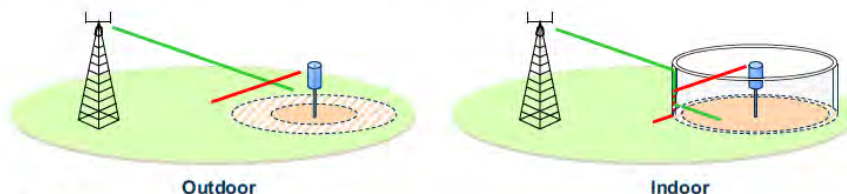


FIGURE 16.3

Illustration of range expansion in different scenarios

In the following sections, four different approaches to heterogeneous deployments are described:

- Release-8 functionality, using features available in the first release of LTE to support a medium amount of range expansion. No inter-cell time synchronization or coordination is assumed.
- Frequency-domain partitioning, where an extensive amount of range expansion is supported through interference handling in the frequency domain, for example, by using carrier aggregation.
- Time-domain partitioning, where an extensive amount of range expansion is supported through interference handling in the time domain.
- Shared cell, using CoMP techniques from Chapter 15 to support a large amount of range expansion.

In the first three approaches above, each transmission point defines a unique cell—that is, it has a unique cell identity and transmits all signals associated with a cell such as cell-specific reference signals and system information. The last approach differs in this respect as a transmission point not necessarily defines a unique cell. Instead, *multiple* geographically separated transmission points may belong to the same cell.

Finally, note that all schemes except the first assumes inter-layer coordination and time-synchronization across (neighboring) transmission points.

16.2 Heterogeneous deployments using rel-8 functionality

Heterogeneous deployments are already possible from the first release of LTE using release-8 functionality. In this case, the transmission points define unique cells, and point association, or cell selection, is typically based on the downlink received power as in the homogeneous case. Despite being simple—for example, there is no need for inter-cell time synchronization or inter-cell coordination—a fair amount of range expansion, up to several dBs, is easily achieved in this scheme by adjusting the cell selection offset. The amount of macro interference naturally limits the amount of range expansion possible, but the amount of range expansion possible is sufficient for many scenarios. Additional tools available in release 8 to obtain a fair amount of range expansion include PDCCH power boosting in the pico cell, fractional loading of the PDCCH in the overlaid macro cell to reduce interference, and adjusting the PDCCH operating point in terms of PDCCH error probability.

It is also worth pointing out that, in many scenarios, most of the gains are obtained by simply deploying the pico nodes with no or only a small amount of range expansion. However, in some specific scenarios, a larger amount of range expansion may be useful, calling for some of the schemes discussed below.

16.3 Frequency-domain partitioning

Frequency-domain partitioning attempts to reduce interference by using different parts of the frequency spectrum in the different layers. The transmission points define unique cells, and measurements of the received downlink power are used as the basis of point association or, equivalently, cell selection.

The simplest case is a static split, using different and non-overlapping pieces of spectrum in the macro and pico layers as illustrated in Figure 16.4. Although simple, such a scheme suffers from not being able to dynamically reallocate resources between the layers to follow instantaneous traffic variations.

A more dynamic approach for handling the downlink inter-layer interference in the range-expansion zone in case of a large amount of range expansion is to use carrier aggregation in combination with cross-carrier scheduling as outlined in Section 10.4.8. The basic idea is to split the overall spectrum into two parts through the use of two downlink carriers, f_1 and f_2 , as illustrated in Figure 16.4, but without the loss of flexibility resulting from a static split.

In terms of data (PDSCH) transmission, both carriers are available in both layers and interference between the layers is handled by “conventional” inter-cell interference coordination (see Section 15.1). As already mentioned, such interference coordination can be more

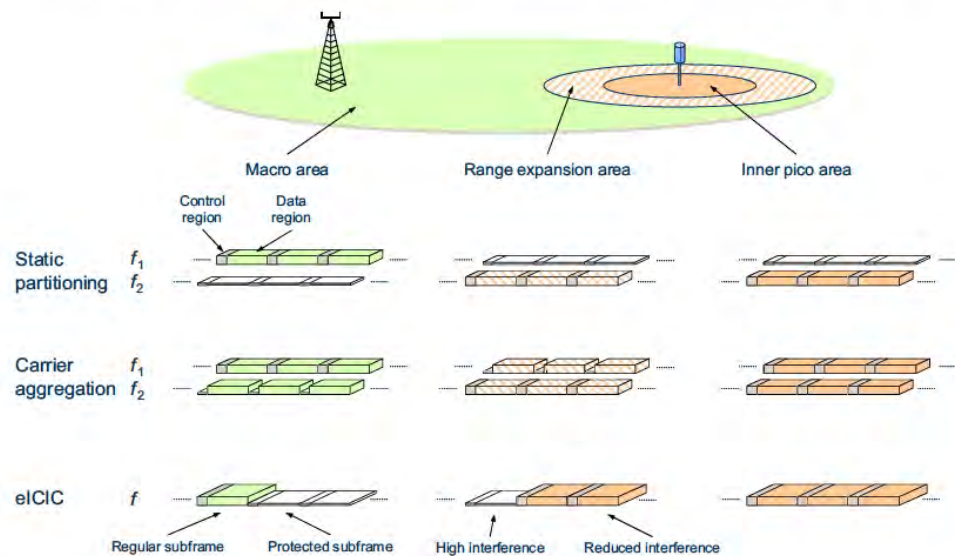


FIGURE 16.4

Frequency-domain and time-domain partitioning

or less dynamic depending on the time scale on which the layers can be coordinated. Also, the possibility for carrier aggregation allows for both carriers—that is, the total available spectrum—to be assigned for transmission to a single terminal. Thus, at least for carrier-aggregation-capable terminals, there is no spectrum fragmentation in terms of data (PDSCH) transmission. Legacy terminals, on the other hand, will experience peak rates from a single carrier only. This may be an issue for an operator with a large fleet of legacy terminals.

On the other hand, in terms of L1/L2 control signaling (PCFICH, PHICH, PDCCH), there is at least partly a more semi-static frequency separation between the layers. More specifically, the macro layer should avoid high-power transmission within the control region on carrier f_1 . Assuming a time-synchronized network, interference to the control region of the pico layer is reduced on this carrier and the pico cells can use the carrier for control signaling to terminals in the range expansion region. Due to the possibility for cross-carrier scheduling, DL-SCH transmission can still be scheduled on both carriers as well as an aggregation of these subject to dynamic inter-layer interference coordination, even if the macro cell only transmits control signaling on carrier f_1 . The same is true for a pico cell; even if the pico cell can only use carrier f_2 for transmission of scheduling assignments to terminals in the range-expansion zone, DL-SCH transmissions can still be scheduled on both carriers.

It should be noted that, for terminals in the inner part of a pico cell, carrier f_1 could also be used for L1/L2 control signaling. Similarly, macro cells could also use carrier f_2 for control signaling, assuming a reduced transmit power is used. Thus, the macro cell could use carrier f_2 for lower-power control signaling, for example, for terminals close to the corresponding macro transmission point.

In the preceding discussion PDCCH has been assumed, but the EPDCCH can equally well be used. In principle, the EPDCCH can be subject to the same inter-cell interference coordination scheme as the PDSCH. This could be used to support large amounts of range expansion without carrier aggregation. In this case, the macro and pico layers simply use different sets of physical resource-block pairs, coordinated in a more or less dynamic manner. However, note that the EPDCCH is limited to the terminal-specific search spaces only. Consequently, the PDCCH is needed for scheduling of, for example, system information.

Finally, note that the (cell-specific reference) signals used by the terminal to maintain synchronization with the pico cell in the range expansion zone are subject to interference from the macro layer. How well the terminal can handle this will put an upper limit to the amount of range expansion possible. Hence, to fully exploit the benefits of range expansion, interference cancelling receivers are needed in the terminals.

16.4 Time-domain partitioning

An alternative to frequency-domain partitioning is to use time-domain partitioning as illustrated at the bottom of Figure 16.4. In 3GPP, this is known as (Further) Enhanced

Inter-Cell Interference Coordination, (F)eICIC. Work on eICIC started in release 10 and was finalized in release 11 under the name of FeICIC. Also in this case, the transmission points correspond to separate cells, hence the FeICIC name used in 3GPP.

The basic idea with time-domain partitioning is to restrict the transmission power of the overlaid macro cell in some subframes. In these *reduced-power subframes* (or *almost blank subframes*), terminals connected to the pico cell will experience less interference from the overlaid macro cell for both data and control. From a terminal perspective they serve as *protected subframes*. The pico cell can therefore schedule terminals in the range expansion area using the protected subframes and terminals in the inner part of the pico cell using all subframes. The macro cell, on the other hand, primarily schedules terminals outside the protected subframes (see Figure 16.5 for an illustration). The gain from deploying the pico cells must be larger than the loss incurred by the macro cell reducing power in some subframes for the time-domain partitioning scheme to be attractive. Whether this holds or not is highly scenario dependent, although using a reduced but non-zero transmission power in the macro cell for the protected subframes to limit the resource loss in the macro layer is often advantageous.

To support time-domain partitioning in a heterogeneous network, signaling of *protected-subframe patterns*—that is, information about the set of protected subframes—is supported between eNodeBs of different layers using the X2 interface. Note that the set of protected subframes could be different in different cells and more or less dynamic, once again depending on the time scale on which the deployment layers can be coordinated.

It should be noted that the macro cell must not necessarily avoid control-signaling transmission completely in the protected subframes. In particular, it could be beneficial to retain the possibility for a limited amount of control signaling related to uplink transmissions, for example, a limited amount of uplink scheduling grants and/or PHICH transmission, in order to not cause too much impact on the uplink scheduling. As long as the macro-cell control-signaling transmissions are limited and only occupy a small fraction of the overall control region, the interference to terminals in the range expansion region of the pico cell could be kept at an acceptable level. However, the signaling of protected-subframe patterns is also defined so

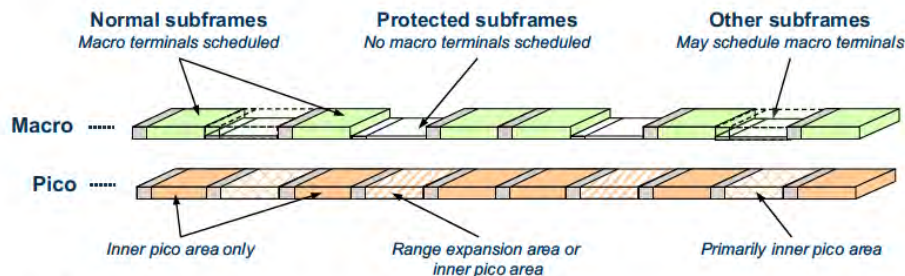


FIGURE 16.5

Protected subframes

that impact on the uplink scheduling is minimized even if no uplink scheduling grants and PHICH can be transmitted in protected subframes. This is achieved by having the protected-subframe patterns matched to the eight-subframe timing of the uplink hybrid-ARQ protocol. It should be noted that this implies that the pattern is not aligned to the 10 ms frame but only to a 40 ms four-frame structure for FDD. For TDD the periodicity also depends on the uplink–downlink configuration.

Up to four different patterns can be exchanged between eNodeBs: two patterns for scheduling and CSI measurement purposes, one pattern for RRM measurements in the serving cell, and one pattern for RRM measurements in neighboring cells. The purpose of having two patterns for CSI purposes is to handle load variations between macro and pico layers without frequent reconfigurations of the terminals, as will be further discussed below (see Figure 16.6 for an illustration). In case of a high load in the range expansion area, which is the case at time t_1 in the figure, it may be beneficial to have a relatively large number of protected subframes to allow these terminals to be served by the pico cell. At a later time instant, t_2 in the figure, the majority of the terminals have moved from the range expansion area into the macro area, calling for a reduction of the number of protected subframes. Thus, by varying the size of the set of currently protected subframes, the configurations can be adjusted to match changes in the scenario. At the same time, there must be a set of subframes that are always protected in order for the pico cell to be able to contact a terminal in the range expansion area as the connection to a terminal otherwise might be lost. Therefore, exchange of two subframe patterns are supported on X2, one of which is intended to be used for the *currently* protected subframes and one for the *always* protected subframes. The intention is to use the former for scheduling coordination across the cells, allowing for relatively frequent updates, while the latter is updated infrequently and used as the basis for configuring protected subframes in the terminals, as described below. Subframes belonging to neither of the two sets above can be thought of as “never protected” subframes.

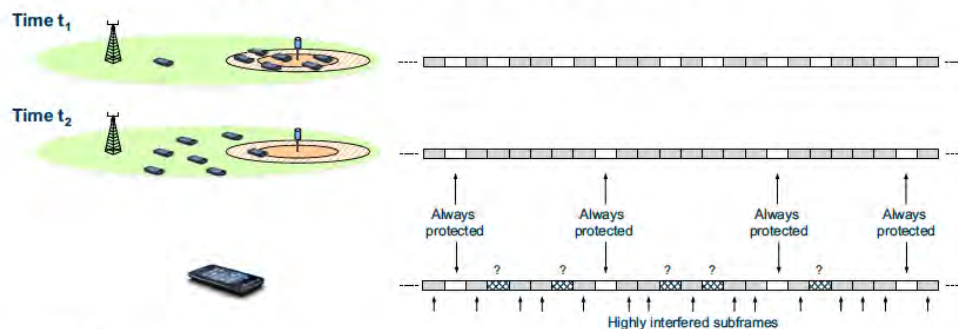


FIGURE 16.6

Exchange of subframe patterns between macro and pico nodes and the corresponding configuration in the terminal

The interference experienced by terminals connected to the pico cell may vary significantly between protected and non-protected subframes. CSI measurements carried out jointly on both the protected and non-protected subframes will thus not accurately reflect the interference of either type of subframe. To address this, the terminal is provided with information about the protected subframes via dedicated RRC signaling using similar bitmaps as described previously.⁴ Two bitmaps can be sent to the terminal, one defining the sets of *protected* subframes and one defining the set of *highly interfered* subframes. Preferably, the protected and highly-interfered subframes correspond to the always protected and never protected subframes derived from the X2 signaling shown earlier. The remaining subframes, if any, not belonging to either of these two subsets have a more unpredictable interference situation as the macro may or may not use reduced power.

CSI reporting is carried out individually for the two subsets. Which subset a certain CSI report reflects depends on in which subframe the CSI is transmitted; the CSI reflects the interference situation in the subset to which the subframe belongs. Hence, the terminal should only average interference measurements during subframes belonging to the same subset. A CSI report transmitted in a subframe not belonging to either of the subsets is undefined from an interference measurement perspective. Through the use of two subsets, the network can predict the radio-channel quality for upcoming transmissions, irrespective of whether they occur in protected subframes or not.

Having two subsets is beneficial for multiple reasons. One example is the situation outlined above, where the set of protected subframes varies over time. Frequent updates of the configurations in all the affected terminals may not be feasible with reasonable overhead. Instead, measuring CSI on a subset of subframes that are always protected is typically preferred as it allows the network to dynamically use reduced power and schedule terminals in the range expansion zone in additional subframes without reconfiguring all the terminals. CSI reports reflecting the situation in the protected subframes are then used for link adaptation in the range expansion zone, while CSI reports from the highly interfered subframes are useful when scheduling terminals in the inner pico area. Another example is when a pico cell is located at the border between, and subject to interference from, two macro-cells. If the macro-cells have differently configured and only partly overlapping sets of protected subframes, the pico-cell scheduling as well as the configuration of the CSI-measurement sets need to take the structure of the protected sets of both macro-cells into account.

The discussion so far has focused on the dynamic part of the interference—that is, the interference part that varies with the traffic load and can be handled by ICIC and time-domain partitioning. However, there is also static interference from the macro cell in the range expansion area. For example, cell-specific reference signals, synchronization signals, and the PBCH still need to be transmitted. To support extensive range expansion, despite the presence of these signals and channels, the interfering signals need to be cancelled. Therefore, cancellation of CRS, PSS/SSS, and PBCH is required to fully exploit the features

⁴Note that transmission mode 10 with multiple CSI processes can be used as an alternative to the two bitmaps.

described previously, functionality that is not mandated in release 10. To assist the terminal in cancelling the interference, RRC signaling provides information about the physical-layer identity of the neighboring cells, the number of antenna ports in those cells, and the MBSFN configuration (MBSFN configuration is needed as there is no CRS in the data region in those subframes).

Time-domain partitioning is supported in RRC_CONNECTED only. Idle terminals will still be able to connect to the pico cell, but cannot exploit range expansion until they enter RRC_CONNECTED.

16.5 Shared cell

In the partitioning schemes described in the previous sections, the transmission points correspond to individual cells, each of which has an individual cell identity that is different from neighboring cells in either of the network layers. Consequently, each pico node transmits unique system information, synchronization signals, and cell-specific reference signals. This is illustrated to the left in Figure 16.7.

Alternatively, CoMP techniques (see Chapter 15) can be used to realize heterogeneous deployments. To understand this approach, remember the distinction between a *cell* and *transmission point*. A cell has a unique physical-layer cell identity from which the position of the cell-specific reference signals is derived. By acquiring the cell identity, a terminal can determine the CRS structure of the cell and receive system information necessary to access the network. A transmission point, on the other hand, is in this context simply one or more co-located antennas from which a terminal can receive data transmissions.

By exploiting the DM-RS introduced in release 10, the PDSCH does not have to be transmitted from the same point as the cell-specific reference signals. Data can instead be transmitted from one of the pico transmission points when beneficial and the time-frequency resources can be reused between spatially separated pico transmission points.

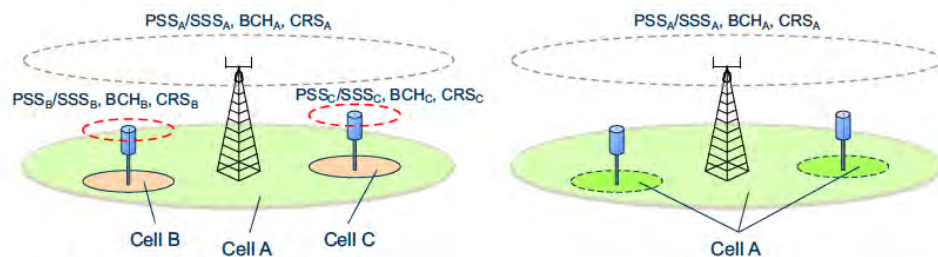


FIGURE 16.7

Independent cells (left) vs. shared cell (right)

Since the pico transmission points neither transmit (unique) cell-specific reference signals, nor system information, they do not define cells but are part of the overlaid macro-cell. This CoMP-approach to heterogeneous deployments is therefore commonly referred to as *shared-cell ID*, illustrated to the right in Figure 16.7.

In Figure 16.8, data is transmitted to terminal number two from the rightmost transmission point. Since the associated DM-RS is transmitted from the same transmission point as the data, the point used for data transmission does not need to be known by the terminal. Spatial reuse gains—that is, reusing the time-frequency resources used for the data transmission across multiple pico nodes within the same macro cell—are hence obtained similarly to the resource-partitioning schemes described in the previous sections.

The control information required in release 10 is based on CRS and the control information therefore needs to be transmitted from (at least) the macro-site as is the case for the first terminal in Figure 16.8. Thus, in many cases, data and the associated control signaling originate from *different* transmission points. This is in theory transparent to the terminal; it only needs to know which reference signal to use for which piece of information, but not from which transmission point the information originates. The quasi-co-location mechanism introduced in release 11 (see Chapter 9) is preferably used to ensure that the terminal only exploits the relevant reference signals.

In Figure 16.8, multiple ways of transmitting the control information are illustrated. The first case, involving terminal 1 with control signaling originating from the macro-site only, has already been described. Alternatively, identical CRS and control channels can be transmitted from the macro- *and* the pico-node as shown for terminal 2. To the terminal this will appear as one composite node as the *same* signal is transmitted from both nodes. The first case is beneficial from a network power-consumption perspective as the pico transmission point is active only when there is data to transmit. The second case, on the other hand, provides an improvement in the signal-to-noise ratio for the control signaling via over-the-air combining of the macro- and pico-transmissions. Furthermore, as an LTE terminal estimates the uplink path loss for power control purposes from the received CRS signal strength (see Chapter 11

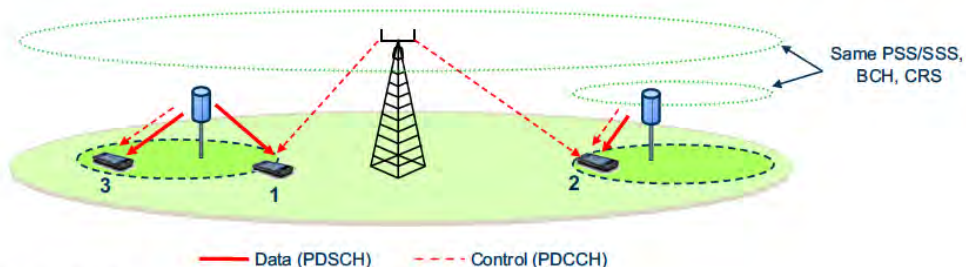


FIGURE 16.8

Heterogeneous deployment using a shared cell

for a discussion on uplink power control), the second case can sometimes result in a more accurate uplink power control.

In both these cases there is no spatial-reuse gain for the control signaling as the macro site is involved in all these transmissions and time-frequency resources cannot be reused across pico nodes. This can be addressed by using the DM-RS-based EPDCCH for control signaling and transmitting it from the same node as used for the PDSCH, as is the case of the third terminal in Figure 16.8.

Terminals not supporting DM-RS-based transmission can still operate in the shared-cell scheme. Data transmissions to these terminals are CRS-based and thus handled in the same manner as the PDCCH control signaling described previously. Although no spatial reuse gain will be possible for these terminals, they will benefit from the pico nodes through an improved signal-to-noise ratio.

Channel-state feedback used for scheduling decisions is preferably based on the CSI-RS. Different pico-nodes, as well as the macro-node, can be configured to use different and non-interfering CSI-RS configurations to allow terminals to estimate the channel conditions to the transmission points corresponding to the different nodes. For terminals not supporting CSI-RS, the channel-state feedback is based on the CRS. In these cases, the eNodeB may need to scale the received reports to account for the difference in the set of transmission points used for CRS and PDSCH.

Deploying a shared-cell scheme can be done by connecting one or multiple remote radio units (RRUs) as well as the macro-site to the same main unit using optical fibers. One reason for this is the tight coupling between the macro- and pico-nodes with control and data originating from different transmission points, calling for low-latency connections.

Centralization of the processing provides benefits also in terms of uplink performance and, in many cases, this alone can motivate the usage of RRUs with centralized processing. Any combination of transmission points, not necessarily those used for downlink transmission to a terminal, can be used for receiving the transmissions from that terminal. By combining the signals from the different antennas in a constructive manner at the central processing node—in essence uplink CoMP (see Chapter 15)—a significant improvement in the uplink data rates can be achieved. In essence, uplink reception and downlink transmission has been *decoupled*, and it is possible to do “uplink range expansion” without causing a downlink interference problem as in separate-cell-ID deployments. The uplink gains can be achieved also for release-8 terminals.

Heterogeneous deployments using shared cells can also provide additional mobility robustness compared to deployments with separate cells. This can be an important aspect, especially when moving from a pico-node to the macro-node. In a separate-cell deployment, a handover procedure is required to change the serving cell. If, during the time it takes to perform the handover procedure, the terminal has moved too far into the macro-area, it may drop the downlink connection from the pico-node before the handover is complete, leading to a radio link failure. In a shared-cell deployment, on the other hand, the transmission point to use for downlink transmission can be rapidly changed *without a handover procedure*. Thus, the probability of dropping connections is reduced.

16.6 Closed subscriber groups

The discussion in the previous sections assumes that the terminals are allowed to connect to the low-power pico-node. This is known as *open access*, and typically the low-power nodes are operator-deployed in such a scenario. Another scenario, giving rise to a similar interference problem, is user-deployed home base stations. The term *closed subscriber groups* (CSG) is commonly used to refer to cases when access to such low-power base stations is limited to a small set of terminals, for example the family living in the house where the home base station is located. CSG results in additional interference scenarios. For example, a terminal located close to but not permitted to connect to the home base station will be subject to strong interference and may not be able to access the macro-cell. In essence, the presence of a home base station and may cause a coverage hole in the operator's macro network, a problem that is particularly worrisome as the home base stations typically are user deployed and their locations are not controlled by the operator. Similarly, reception at the home base station may be severely impacted by uplink transmissions from the terminal connected to the macro-cell. In principle, these problems can to some extent be solved by the same means as above—that is, by relying on interference coordination between the scheduling in the home-eNodeB layer and an overlaid macro. However, note that the interference avoidance must be two-way—that is, one must not only avoid interference from the macro cell to home-eNodeB terminals in the high-interference outer region of the home-eNodeB coverage area, but also home-eNodeB interference to terminals close to the home-eNodeB but not being part of the home-eNodeB CSG. Furthermore, most interference coordination schemes assume the presence of an X2 interface between the macro- and home eNodeB, which may not always hold. Therefore, if closed subscriber groups are supported, it is preferable to use a separate carrier for the CSG cells to maintain the overall performance of the radio-access network. Interference handling between CSG cells, which typically lack backhaul-based coordination schemes, could rely on distributed algorithms for power control and/or resource partitioning between the cells.

This page intentionally left blank

Multimedia Broadcast/Multicast Services

17

In the past, cellular systems have mostly focused on transmission of data intended for a single user and not on multicast/broadcast services. Broadcast networks, exemplified by the radio and TV broadcasting networks, have on the other hand focused on covering very large areas with the same content and have offered no or limited possibilities for transmission of data intended for a single user. *Multimedia Broadcast Multicast Services* (MBMS) support multicast/broadcast services in a cellular system, thereby combining the provision of multicast/broadcast and unicast services within a single network.

With MBMS, introduced in LTE release 9, the same content is transmitted to multiple users located in a specific area, known as the *MBMS service area* and typically comprising multiple cells. In each cell participating in the transmission, a point-to-multipoint radio resource is configured and all users subscribing to the MBMS service simultaneously receive the same transmitted signal. No tracking of users' movement in the radio-access network is performed and users can receive the content without notifying the network.

When providing multicast/broadcast services for mobile devices there are several aspects to take into account, of which two deserve special attention and will be elaborated upon further below: good coverage and low terminal power consumption.

The coverage, or more accurately the data rate possible to provide, is basically determined by the link quality of the worst-case user, as no user-specific adaptation of transmission parameters can be used in a multicast/broadcast system providing the same information to multiple users. As discussed in Chapter 3, OFDM transmission provides specific benefits for provision of multi-cell multicast/broadcast services. If the transmissions from the different cells are time synchronized, the resulting signal will, from a terminal point of view, appear as a transmission from a single point over a time-dispersive channel. As mentioned in Chapter 7, for LTE this kind of transmission is referred to as an *MBMS Single-Frequency Network* (MBSFN). MBSFN transmission provides several benefits:

- Increased received signal strength, especially at the border between cells involved in the MBSFN transmission, as the terminal can utilize the signal energy received from multiple cells.
- Reduced interference level, once again especially at the border between cells involved in the MBSFN transmission, as the signals received from neighboring cells will not appear as interference but as useful signals.

- Additional diversity against fading on the radio channel as the information is received from several, geographically separated locations, typically making the overall aggregated channel appear highly time-dispersive or, equivalently, highly frequency selective.

Altogether, this allows for significant improvements in the multicast/broadcast reception quality, especially at the border between cells involved in the MBSFN transmission, and, as a consequence, significant improvements in the achievable multicast/broadcast data rates.

Providing for power-efficient reception in the terminal in essence implies that the structure of the overall transmission should be such that data for a service-of-interest is provided in short high-data-rate bursts rather than longer low-data-rate bursts. This allows the terminal to occasionally wake up to receive data with long periods of DRX in between. In LTE, this is catered for by time-multiplexing unicast and broadcast transmissions, as well as by the scheduling of different MBMS services, as discussed further below.

17.1 Architecture

An *MBSFN area* is a specific area where one or several cells transmit the same content. For example, in Figure 17.1, cells 8 and 9 both belong to MBSFN area C. Not only can an MBSFN area consist of multiple cells, a single cell can also be part of multiple, up to eight, MBSFN areas, as shown in Figure 17.1 where cells 4 and 5 are part of both MBSFN areas A and B. Note that, from an MBSFN reception point of view, the individual cells are invisible, although the terminal needs to be aware of the different cells for other purposes, such as reading system information and notification indicators, as discussed below. The MBSFN areas are static and do not vary over time.

The usage of MBSFN transmission requires not only time synchronization among the cells participating in an MBSFN area, but also usage of the same set of radio resources in each of the cells for a particular service. This coordination is the responsibility of the *Multi-cell/multicast Coordination Entity* (MCE), which is a logical node in the radio-access

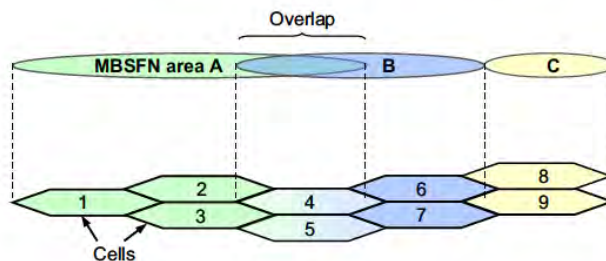
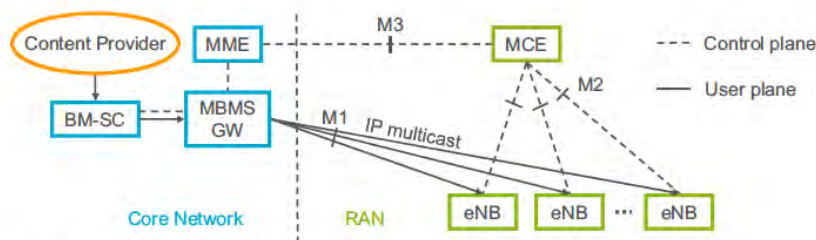


FIGURE 17.1

Example of MBSFN areas

**FIGURE 17.2**

LTE MBMS architecture

network handling allocation of radio resources and transmission parameters (time–frequency resources and transport format) across the cells in the MBSFN area. As shown in Figure 17.2, the MCE¹ can control multiple eNodeBs, each handling one or more cells.

The *Broadcast Multicast Service Center* (BM-SC), located in the core network, is responsible for authorization and authentication of content providers, charging, and the overall configuration of the data flow through the core network. The MBMS gateway (MBMS-GW) is a logical node handling multicast of IP packets from the BM-SC to all eNodeBs involved in transmission in the MBSFN area. It also handles session control signaling via the MME.

From the BM-SC, the MBMS data is forwarded using IP multicast, a method of sending an IP packet to multiple receiving network nodes in a single transmission, via the MBMS gateway to the cells from which the MBMS transmission is to be carried out. Hence, MBMS is not only efficient from a radio-interface perspective, but it also saves resources in the transport network by not having to send the same packet to multiple nodes individually unless necessary. This can lead to significant savings in the transport network.

A terminal receiving MBMS transmission may also receive unicast transmission on the same carrier as MBMS and unicast transmissions are time-multiplexed onto different sub-frames. This assumes the same carrier being used for both MBMS and unicast transmission, which may limit the deployment flexibility in case an operator uses multiple carriers (multiple frequency bands) in an MBSFN area. In release 11, enhancements were introduced to improve operation in such deployments. Briefly, the terminal informs the network about its MBMS interest and capabilities. The network can take this information into account and ensure that the terminal is able to receive the relevant MBMS service, for example by handover of the terminal to the carrier providing the MBMS transmission. A carrier-aggregation capable terminal may receive unicast transmissions on one component carrier and MBMS on another component carrier.

¹There is an alternative architecture supported where MCE functionality is included in every eNodeB. However, as there is no communication between MCEs in different eNodeBs, the MBSFN area would in this case be limited to the set of cells controlled by a single eNodeB.

17.2 Overall channel structure and physical-layer processing

The basis for MBSFN transmission is the *Multicast Channel* (MCH), a transport channel type supporting MBSFN transmission. Two types of logical channels can be multiplexed and mapped to the MCH:

- Multicast Traffic Channel (MTCH)
- Multicast Control Channel (MCCH)

The MTCH is the logical channel type used to carry MBMS data corresponding to a certain MBMS service. If the number of services to be provided in an MBSFN area is large, multiple MTCHs can be configured. As no acknowledgements are transmitted by the terminals, no RLC retransmissions can be used and consequently the RLC unacknowledged mode is used.

The MCCH is the logical channel type used to carry control information necessary for reception of a certain MBMS service, including the subframe allocation and modulation-and-coding scheme for each MCH. There is one MCCH per MBSFN area. Similarly to the MTCH, the RLC uses unacknowledged mode.

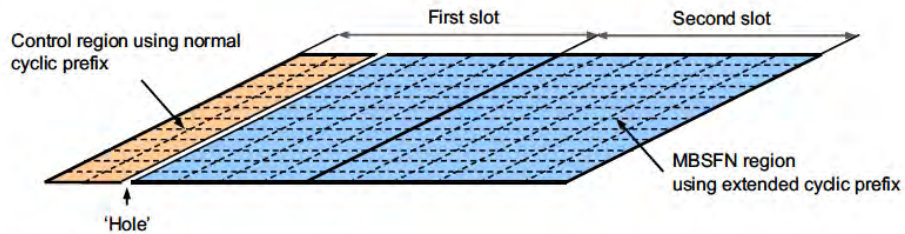
One or several MTCHs and, if applicable,² one MCCH are multiplexed at the MAC layer to form an MCH transport channel. As described in Chapter 8, the MAC header contains information about the logical-channel multiplexing, in this specific case the MTCH/MCCH multiplexing, such that the terminal can demultiplex the information upon reception. The MCH is transmitted using MBSFN in one MBSFN area.

The transport-channel processing for MCH is, in most respects, the same as that for DL-SCH as described in Section 10.1, with some exceptions:

- In the case of MBSFN transmission, the same data is to be transmitted with the same transport format using the same physical resource from multiple cells typically belonging to different eNodeBs. Thus, the MCH transport format and resource allocation cannot be dynamically adjusted by the eNodeB. As described above, the transport format is instead determined by the MCE and signaled to the terminals as part of the information sent on the MCCH.
- As the MCH transmission is simultaneously targeting multiple terminals and therefore no feedback is used, hybrid ARQ is not applicable in the case of MCH transmission.
- As already mentioned, multi-antenna transmission (transmit diversity and spatial multiplexing) does not apply to MCH transmission.

Furthermore, as also mentioned in Chapter 10, the PMCH scrambling should be *MBSFN-area specific*—that is, identical for all cells involved in the MBSFN transmission.

²One MCCH per MBSFN area is needed, but it does not have to occur in every MCH TTI, nor on all MCHs in the MBSFN area.

**FIGURE 17.3**

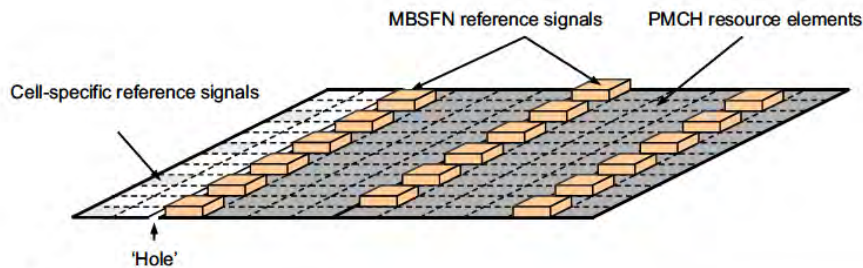
Resource-block structure for MBSFN subframes, assuming normal cyclic prefix for the control region

The MCH is mapped to the PMCH physical channel and transmitted in MBSFN subframes, illustrated in Figure 17.3. As discussed in Chapter 9, an MBSFN subframe consists of two parts: a *control region*, used for transmission of regular unicast L1/L2 control signaling, and an *MBSFN region*, used for transmission of the MCH.³ Unicast control signaling may be needed in an MBSFN subframe, for example to schedule uplink transmissions in a later subframe, but is also used for MBMS-related signaling, as discussed further below.

As discussed in Chapter 3, in the case of MBSFN-based multicast/broadcast transmission, the cyclic prefix should not only cover the main part of the actual channel time dispersion but also the timing difference between the transmissions received from the cells involved in the MBSFN transmission. Therefore, MCH transmissions, which can take place in the MBSFN region only, use an extended cyclic prefix. If a normal cyclic prefix is used for normal subframes, and therefore also in the control region of MBSFN subframes, there will be a small “hole” between the two parts of the MBSFN subframe, as illustrated in Figure 17.3. The reason is to keep the start timing of the MBSFN region fixed, irrespective of the cyclic prefix used for the control region.

As already mentioned, the MCH is transmitted by means of MBSFN from the set of cells that are part of the corresponding MBSFN area. Thus, as seen from the terminal point of view, the radio channel that the MCH has propagated over is the aggregation of the channels of each cell within the MBSFN area. For channel estimation for coherent demodulation of the MCH, the terminal can thus not rely on the normal cell-specific reference signals transmitted from each cell. Rather, in order to enable coherent demodulation for MCH, special MBSFN reference symbols are inserted within the MBSFN part of the MBSFN subframe, as illustrated in Figure 17.4. These reference symbols are transmitted by means of MBSFN over the set of cells that constitute the MBSFN area—that is, they are transmitted at the same time–frequency position and with the same reference-symbol values from each cell. Channel estimation using these reference symbols will thus correctly reflect the overall aggregated channel corresponding to the MCH transmissions of all cells that are part of the MBSFN area.

³ As discussed in Chapter 9, MBSFN subframes can be used for multiple purposes and not all of them have to be used for MCH transmission.

**FIGURE 17.4**

Reference-signal structure for PMCH reception

MBSFN transmission in combination with specific MBSFN reference signals can be seen as transmission using a specific antenna port, referred to as *antenna port 4*.

A terminal can assume that all MBSFN transmissions within a given subframe correspond to the same MBSFN area. Hence, a terminal can interpolate over all MBSFN reference symbols within a given MBSFN subframe when estimating the aggregated MBSFN channel. In contrast, MCH transmissions in different subframes may, as already discussed, correspond to different MBSFN areas. Consequently, a terminal cannot necessarily interpolate the channel estimates across multiple subframes.

As can be seen in Figure 17.4, the frequency-domain density of MBSFN reference symbols is higher than the corresponding density of cell-specific reference symbols. This is needed as the aggregated channel of all cells involved in the MBSFN transmission will be equivalent to a highly time-dispersive or, equivalently, highly frequency-selective channel. Consequently, a higher frequency-domain reference-symbol density is needed.

There is only a single MBSFN reference signal in MBSFN subframes. Thus, multi-antenna transmission such as transmit diversity and spatial multiplexing is not supported for MCH transmission. The main argument for not supporting any standardized transmit diversity scheme for MCH transmission is that the high frequency selectivity of the aggregated MBSFN channel in itself provides substantial (frequency) diversity. Transmit-diversity schemes transparent to the terminal and hence not requiring any specific support in the specifications can still be applied if beneficial.

17.3 Scheduling of MBMS services

Good coverage throughout the MBSFN area is, as already explained, one important aspect of providing broadcast services. Another important aspect, as mentioned in the introduction, is to provide for energy-efficient reception. In essence, for a given service, this translates into transmission of short high-rate bursts in between which the terminal can enter a DRX state to

reduce power consumption. LTE therefore makes extensive use of time-multiplexing of MBMS services and the associated signaling, as well as provides a mechanism to inform the terminal *when* in time a certain MBMS service is transmitted. Fundamental to the description of this mechanism are the *Common Subframe Allocation* (CSA) period and the *MCH Scheduling Period* (MSP).

All MCHs that are part of the same MBSFN area occupy a pattern of MBSFN subframes known as the *Common Subframe Allocation* (CSA). The CSA is periodic, as illustrated in Figure 17.5. The subframes used for transmission of the MCH must be configured as MBSFN subframes, but the opposite does not hold—MBSFN subframes can be configured for other purposes as well, for example to support the backhaul link in the case of relaying, as described in Chapter 18. Furthermore, the allocation of MBSFN subframes for MCH transmission should be identical across the MBSFN area as there otherwise will not be any MBSFN gain. This is the responsibility of the MCE.

Transmission of a specific MCH follows the *MCH Subframe Allocation* (MSA). The MSA is periodic and at the beginning of each *MCH Scheduling Period* (MSP), a MAC control element is used to transmit the *MCH Scheduling Information* (MSI). The MSI indicates which subframes are used for a certain MTCH in the upcoming scheduling period. Not all possible subframes need to be used; if a smaller number than allocated to an MCH is required by the MTCH(s), the MSI indicates the last MCH subframe to be used for this particular MTCH (*MSA end* in Figure 17.5), while the remaining subframes are not used for MBMS

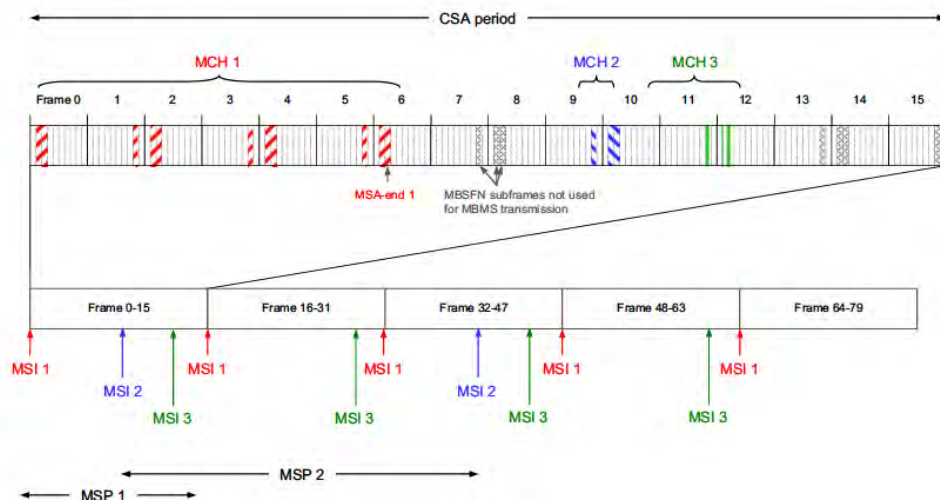


FIGURE 17.5

Example of scheduling of MBMS services

transmission. The different MCHs are transmitted in consecutive order within a CSA period—that is, all subframes used by MCH n in a CSA are transmitted before the subframes used for MCH $n + 1$ in the same CSA period.

The fact that the transport format is signaled as part of the MCCH implies that the MCH transport format may differ between MCHs but must remain constant across subframes used for the same MCH. The only exception is subframes used for the MCCH and MSI, where the MCCH-specific transport format, signaled as part of the system information, is used instead.

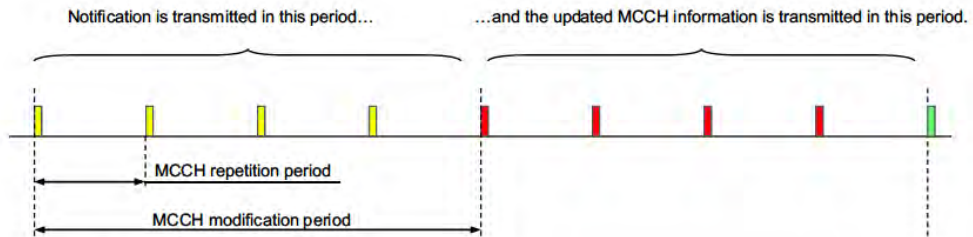
In the example in Figure 17.5, the scheduling period for the first MCH is 16 frames, corresponding to one CSA period, and the scheduling information for this MCH is therefore transmitted once every 16 frames. The scheduling period for the second MCH, on the other hand, is 32 frames, corresponding to two CSA periods, and the scheduling information is transmitted once every 32 frames. The MCH scheduling periods can range from 80 ms to 10.24 s.

To summarize, for each MBSFN area, the MCCH provides information about the CSA pattern, the CSA period, and, for each MCH in the MBSFN area, the transport format and the scheduling period. This information is necessary for the terminal to properly receive the different MCHs. However, the MCCH is a logical channel and is itself mapped to the MCH, which would result in a chicken-and-egg problem—the information necessary for receiving the MCH is transmitted on the MCH. Hence, in TTIs when the MCCH (or MSI) is multiplexed into the MCH, the MCCH-specific transport format is used for the MCH. The MCCH-specific transport format is provided as part of the system information (SIB13; see Chapter 14 for a discussion about system information). The system information also provides information about the scheduling and modifications periods of the MCCH (but not about CSA period, CSA pattern, and MSP, as those quantities are obtained from the MCCH itself). Reception of a specific MBMS service can thus be described by the following steps:

- Receive SIB13 to obtain knowledge on how to receive the MCCH for this particular MBSFN area.
- Receive the MCCH to obtain knowledge about the CSA period, CSA pattern, and MSP for the service of interest.
- Receive the MSI at the beginning of each MSP. This provides the terminal with information on which subframes the service of interest can be found in.

After the second step above, the terminal has acquired the CSA period, CSA pattern, and MSP. These parameters typically remain fixed for a relatively long time. The terminal therefore only needs to receive the MSI and the subframes in which the MTCH carrying the service of interest are located, as described in the third bullet above. This greatly helps to reduce the power consumption in the terminal as it can sleep in most of the subframes.

Occasionally there may be a need to update the information provided on the MCCH, for example when starting a new service. Requiring the terminal to repeatedly receive the MCCH comes at a cost in terms of terminal power consumption. Therefore, a fixed schedule for

**FIGURE 17.6**

MCCH transmission schedule

MCCH transmission is used in combination with a change-notification mechanism, as described below.

The MCCH information is transmitted repeatedly with a fixed repetition period and changes to the MCCH information can only occur at specific time instants. When (parts of) the MCCH information are changed, which can only be done at the beginning of a new modification period, as shown in Figure 17.6, the network notifies the terminals about the upcoming MCCH information change in the preceding MCCH modification period. The notification mechanism uses the PDCCH for this purpose. An eight-bit bitmap, where each bit represents a certain MBSFN area, is transmitted on the PDCCH in an MBSFN subframe using DCI format 1C and a reserved identifier, the M-RNTI. The notification bitmap is only transmitted when there are any changes in the services provided (in release 10, notification is also used to indicate a counting request in an MBSFN area) and follows the modification period described above.

The purpose of the concept of notification indicators and modification periods is to maximize the amount of time the terminal may sleep to save battery power. In the absence of any changes to the MCCH information, a terminal currently not receiving MBMS may enter DRX and only wake up when the notification indicator is transmitted. As a PDCCH in an MBSFN subframe spans at most two OFDM symbols, the duration during which the terminal needs to wake up to check for notifications is very short, translating to a high degree of power saving. Repeatedly transmitting the MCCH is useful to support mobility; a terminal entering a new area or a terminal missing the first transmission does not have to wait until the start of a new modification period to receive the MCCH information.

This page intentionally left blank

Relaying

18

The possibility of a terminal communicating with the network, and the data rate that can be used, depends on several factors, including the path loss between the terminal and the base station. The link performance of LTE is already quite close to the Shannon limit and from a pure link-budget perspective, the highest data rates supported by LTE require a relatively high signal-to-noise ratio. Unless the link budget can be improved, for example with different types of beam-forming solutions, a denser infrastructure is required to reduce the terminal-to-base-station distance and thereby improve the link budget.

A denser infrastructure is mainly a deployment aspect, but in later releases of LTE, various tools enhancing the support for low-power base stations were included. One of these tools is *relaying*, which can be used to reduce the distance between the terminal and the infrastructure, resulting in an improved link budget and an increased possibility for high data rates. In principle this reduction in terminal-to-infrastructure distance could be achieved by deploying traditional base stations with a wired connection to the rest of the network. However, relays with a shorter deployment time can often be an attractive alternative, as there is no need to deploy a specific backhaul.

A wide range of relay types can be envisioned, some of which could already be deployed in release 8.

Amplify-and-forward relays, commonly referred to as *repeaters*, simply amplify and forward the received analog signals and are, on some markets, relatively common as a tool for handling coverage holes. Traditionally, once installed, repeaters continuously forward the received signal regardless of whether there is a terminal in their coverage area or not, although more advanced repeaters can be considered as well. Repeaters are transparent to both the terminal and the base station and can therefore be introduced in existing networks. The fact that the basic principle of a repeater is to amplify whatever it receives, including noise and interference as well as the useful signal, implies that repeaters are mainly useful in high-SNR environments. Expressed differently, the SNR at the output of the repeater can never be higher than at the input.

Decode-and-forward relays decode and re-encode the received signal prior to forwarding it to the served users. The decode-and-re-encode process results in this class of relays not amplifying noise and interference, as is the case with repeaters. They are therefore also useful in low-SNR environments. Furthermore, independent rate adaptation and scheduling for the base station–relay and relay–terminal links is possible. However, the decode-and-re-encode operation implies a larger delay than for an amplify-and-forward repeater, longer than

the LTE subframe duration of 1 ms. As for repeaters, many different options exist depending on supported features (support of more than two hops, support for mesh structures, etc.) and, depending on the details of those features, a decode-and-forward relay may or may not be transparent to the terminal.

18.1 Relays in LTE

LTE release 10 introduced support for a decode-and-forward relaying scheme (repeaters require no additional standardization support other than RF requirements and are already available in release 8). A basic requirement in the development of LTE relaying solutions was that the relay should be transparent to the terminal—that is, the terminal should not be aware of whether it is connected to a relay or to a conventional base station. This ensures that releases-8/9 terminals can also be served by relays, despite relays being introduced in release 10. Therefore, so-called *self-backhauling* was taken as the basis for the LTE relaying solution. In essence, from a logical perspective, a relay is an eNodeB wirelessly connected to the rest of the radio-access network by using the LTE radio interface. It is important to note that, even though the relay from a terminal perspective is identical to an eNodeB, the physical implementation may differ significantly from a traditional base station, for example in terms of output power.

In conjunction with relaying, the terms *backhaul link* and *access link* are often used to refer to the base station–relay connection and the relay–terminal connection respectively. The cell to which the relay is connected using the backhaul link is known as the *donor cell* and the donor cell may, in addition to one or several relays, also serve terminals not connected via a relay. This is illustrated in Figure 18.1.

Since the relay communicates both with the donor cell and terminals served by the relay, interference between the access and backhaul links must be avoided. Otherwise, since the power difference between access-link transmissions and backhaul-link reception at the relay can easily be more than 100 dB, the possibility of receiving the backhaul link may be completely ruined. Similarly, transmissions on the backhaul link may cause significant interference to the reception of the access link. These two cases are illustrated in Figure 18.2.

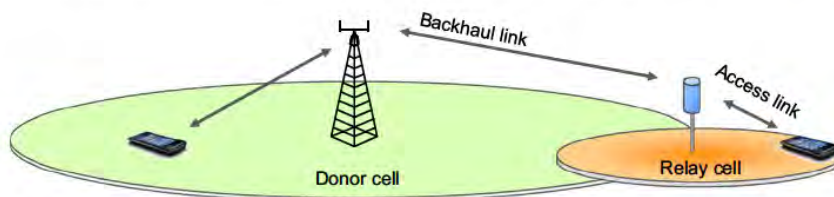
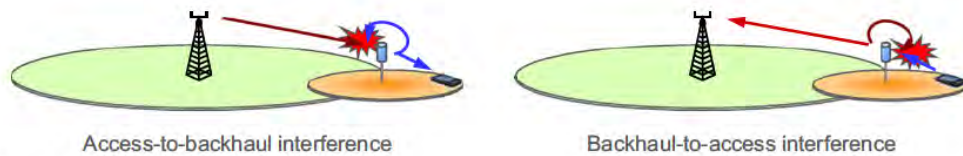


FIGURE 18.1

Access and backhaul links

**FIGURE 18.2**

Interference between access and backhaul links

Therefore, isolation between the access and backhaul links is required—isolation that can be obtained in one or several of the frequency, time, and/or spatial domains.

Depending on the spectrum used for access and backhaul links, relaying can be classified into *outband* and *inband* types.

Outband relaying implies that the backhaul operates in a spectrum separate from that of the access link, using the same radio interface as the access link. Provided that the frequency separation between the backhaul and access links is sufficiently large, interference between the backhaul and access links can be avoided and the necessary isolation is obtained in the frequency domain. Consequently, no enhancements to the release-8 radio interface are needed to operate an outband relay. There are no restrictions on the activity on the access and backhaul links and the relay can in principle operate with full duplex.

Inband relaying implies that the backhaul and access links operate in the same spectrum. Depending on the deployment and operation of the relay, this may, as the access and backhaul link share the same spectrum, require additional mechanisms to avoid interference between the access and backhaul links. Unless this interference can be handled by proper antenna arrangements, for example with the relay deployed in a tunnel with the backhaul antenna placed outside the tunnel, a mechanism to separate activity on the access and backhaul links in the time domain is required. Such a mechanism was introduced as part of release 10 and will be described in more detail. Since the backhaul and access links are separated in the time domain, there is a dependency on the transmission activity and the two links cannot operate simultaneously.

The RF requirements for decode-and-forward relays were introduced in 3GPP release 11. Because of the similarities with operation of the access and backhaul links with base stations and UEs respectively, the requirements are to a large extent very similar to the corresponding ones for base stations and UEs. This is discussed in more detail in section 19.14.

18.2 Overall architecture

From an architectural perspective, a relay can, on a high level, be thought of as having a “base-station side” and a “terminal side.” Towards terminals, it behaves as a conventional eNodeB using the access link, and a terminal is not aware of whether it is communicating

with a relay or a “traditional” base station. Relays are therefore transparent for the terminals and terminals from the first LTE release, release 8, can also benefit from relays. This is important from an operator’s perspective, as it allows a gradual introduction of relays without affecting the existing terminal fleet.

Towards the donor cell, a relay initially operates as a terminal, using the LTE radio interface to connect to the donor cell. Once connection is established and the relay is configured, the relay uses a subset of the “terminal side” functionality for communication on the backhaul link. In this phase, the relay-specific enhancements described in this chapter may be used for the backhaul.

In release 10, the focus is on two-hop relaying and scenarios with a relay connected to the network via another relay are not considered. Furthermore, relays are stationary—that is, handover of a relay from one donor cell to another donor cell is not supported. The case for using mobile relays is not yet clear and therefore it was decided in release 10 not to undertake the relatively large task of adapting existing core-network procedures to handle cells that are moving over time (something that could have been a consequence of a mobile relay).

The overall LTE relaying architecture is illustrated in Figure 18.3. One key aspect of the architecture is that the donor eNodeB acts as a proxy between the core network and the relay. From a relay perspective, it appears as if it is connected directly to the core network, as the donor eNodeB appears as an MME for the S1 interface and an eNodeB for X2 towards the relay. From a core-network perspective, on the other hand, the relay cells appear as if they belong to the donor eNodeB. It is the task of the proxy in the donor eNodeB to connect these two views. The use of a proxy is motivated by the desire to minimize the impact to the core network from the introduction of relays, as well as to allow for features such as tight coordination of radio-resource management between the donor eNodeB and the relay.

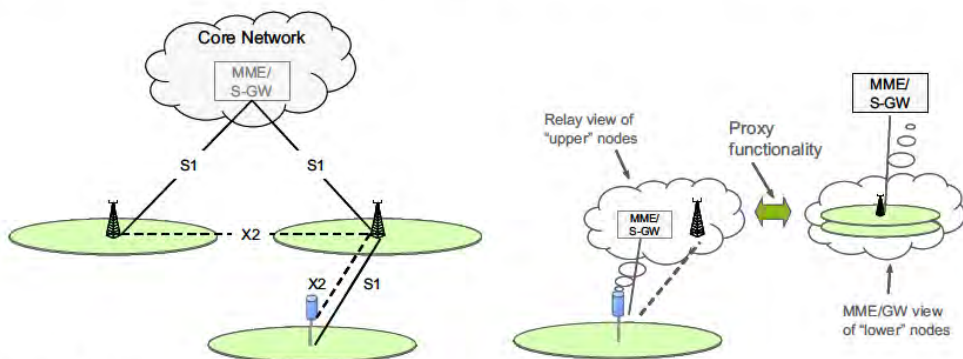


FIGURE 18.3

LTE relaying architecture

18.3 Backhaul design for inband relaying

In the case of inband relaying, the backhaul and access links operate in the same spectrum. As discussed in the previous section, a mechanism to separate activity on the access and backhaul links in the time domain is required unless sufficient isolation between the two links can be achieved in other ways, for example through appropriate antenna arrangements. Such a mechanism should ensure that the relay is not transmitting on the access link at the same time as it is receiving on the backhaul link (and vice versa).

One way to handle this is to “blank” some subframes on the access link to provide the relay with the possibility to communicate with the donor eNodeB on the backhaul link. In the uplink, the scheduler in the relay can in principle schedule such that there is no access-link activity in certain subframes. These subframes can then be used for uplink transmissions on the backhaul link as the relay does not need to receive anything on the access link in these subframes. However, blanking subframes on the access downlink is not possible. Although a release-10 terminal in principle could have been designed to cope with blank subframes, terminals from earlier releases expect at least cell-specific reference signals to be present in all downlink subframes. Hence, to preserve the possibility of also serving releases-8/9 terminals, which was an important requirement during standardization of release 10, the design of the backhaul link must be based on the assumption that the access link can operate with release-8 functionality only.

Fortunately, from the first release LTE included the possibility of configuring MBSFN subframes (see Chapter 9). In an MBSFN subframe, terminals expect cell-specific reference signals and (possibly) L1/L2 control signaling to be transmitted only in the first one or two OFDM symbols, while the remaining part of the subframe can be empty. By configuring some of the access-link subframes as MBSFN subframes, the relay can stop transmitting in the latter part of these subframes and receive transmissions from the donor cell. As seen in Figure 18.4, the gap during which the relay can receive transmissions from the donor cell is shorter than the full subframe duration. In particular, as the first OFDM symbols in the subframe are unavailable for reception of donor-cell transmissions, L1/L2 control signaling from the donor to the relay cannot be transmitted using the regular PDCCH. Instead, a relay-specific control channel, the R-PDCCH, was introduced in release 10.

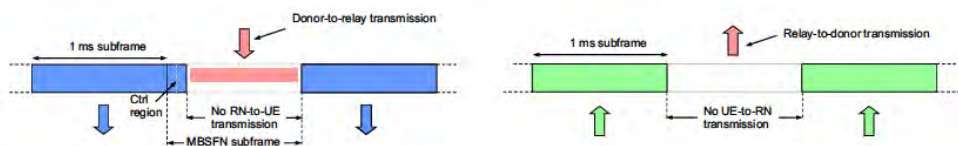


FIGURE 18.4

Multiplexing between access and backhaul links

Not only are transmission gaps in the access downlink required in order to receive transmissions from the donor cell, but also reception gaps in the access link are needed in order to transmit on the backhaul from the relay to the donor cell. As already mentioned, such gaps can be created through proper scheduling of uplink transmissions.

The detailed specifications of the physical-layer enhancements introduced in release 10 to support the backhaul can be found in [75].

18.3.1 Access-link hybrid-ARQ operation

The access-link gaps discussed above, MBSFN subframes in the downlink and scheduling gaps in the uplink, used in order to be able to receive and transmit respectively on the backhaul link, affect the hybrid-ARQ operation. Note that hybrid ARQ is used on both the access and backhaul links. Since compatibility with release 8 was a fundamental requirement in the development of the LTE relaying solution, there are no changes to access-link hybrid-ARQ operation.

For uplink transmissions on PUSCH, hybrid-ARQ acknowledgements are transmitted on PHICH. Since the PHICH can be transmitted by the relay even in MBSFN subframes, the operation is identical to that in earlier releases of the LTE standard. However, although the hybrid-ARQ acknowledgement can be received, the subframe where the retransmission should take place (8 ms after the initial transmission for FDD, configuration dependent for TDD) may be used by the backhaul link and not be available for the access link. In that case the corresponding uplink hybrid-ARQ process needs to be suspended by transmitting a positive acknowledgement on the PHICH, irrespective of the outcome of the decoding. By using PDCCH, a retransmission can instead be requested in a later subframe available for the same hybrid-ARQ process, as described in Chapter 12. The hybrid-ARQ round-trip time will be larger in those cases (for example, 16 ms instead of 8 ms for FDD).

Downlink transmissions on PDSCH trigger hybrid-ARQ acknowledgements to be sent on PUCCH and, for proper operation, the relay should be able to receive those acknowledgements. The possibility to receive PUCCH on the access link depends on the backhaul operation, more specifically on the allocation of subframes for backhaul communication.

In FDD, backhaul subframes are configured such that an uplink subframe occurs 4 ms after a downlink subframe. This is chosen to match the access-link hybrid-ARQ timing relations, where an uplink subframe follows 4 ms after a downlink subframe. As the relay cannot transmit on the access link simultaneously with the backhaul link, there is no access-link transmission in subframe n and, consequently, no hybrid-ARQ transmission in subframe $n + 4$. Hence, the inability to receive access-link hybrid-ARQ acknowledgements in some subframes is of no concern as the corresponding downlink subframes cannot be used for access-link transmission anyway. Downlink retransmissions are not an issue as they are asynchronous and can be scheduled in any suitable downlink subframe on the access link.

In TDD, the relay node may not be able to receive hybrid-ARQ feedback on PUCCH in uplink subframes used for transmission on the backhaul link. One possibility is to restrict the

downlink scheduler such that no terminals transmit PUCCH in uplink subframes the relay cannot receive. However, such a restriction may be too limiting. Alternatively, the relay can schedule without restrictions in the downlink and ignore the hybrid-ARQ acknowledgement. Retransmissions can then either be handled blindly—that is, the relay has to make an educated “guess” on whether a retransmission is required based on, for example, CSI feedback—or RLC retransmissions are used to handle missing packets. Another possibility is to configure repetition of the hybrid-ARQ acknowledgements such that at least some of the repeated acknowledgements are receivable by the relay.

18.3.2 Backhaul-link hybrid-ARQ operation

For the backhaul link, the underlying principle in the design is to maintain the same timing relations as in release 8 for scheduling grants and hybrid-ARQ acknowledgements. As the donor cell may schedule both relays and terminals, such a principle simplifies the scheduling implementation, as scheduling and retransmission decisions for terminals and relays are taken at the same point in time. It also simplifies the overall structure, as release-8 solutions can be reused for the relay backhaul design.

For FDD, the subframes configured for downlink backhaul transmission therefore follow a period of 8 ms in order to match the hybrid-ARQ round-trip time to the extent possible. This also ensures that the PUCCH can be received in the access link, as discussed in the previous section. However, as the possible configurations of MBSFN subframes have an inherent 10 ms structure while the hybrid-ARQ timing follows an 8 ms periodicity, there is an inherent mismatch between the two. Hence, as illustrated in Figure 18.5, some backhaul subframes may be spaced 16 ms apart, as subframes 0, 4, 5, and 9 cannot be configured as MBSFN subframes (see Chapter 9). Uplink backhaul subframes follow 4 ms after a downlink backhaul subframe, following the principle discussed in the previous paragraph.

For TDD, there is an inherent 10 ms component in the hybrid-ARQ timing relations, which matches the 10 ms MBSFN structure and makes it possible to keep a regular spacing of the backhaul transmission attempts. Subframes 0, 1, 5, and 6 cannot be configured as MBSFN subframes. Hence, TDD configuration 0, where subframes 0 and 5 are the only downlink subframes, cannot be used in a relay cell since this configuration does not support any

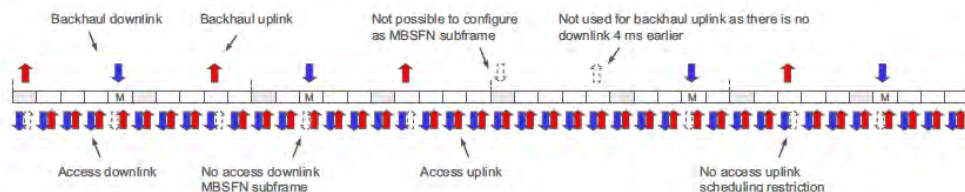


FIGURE 18.5

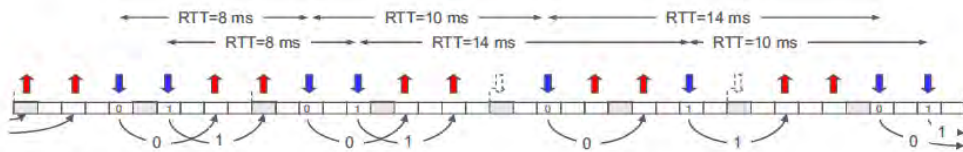
Example of backhaul configuration for FDD

MBSFN subframes. For configuration 5, there is only a single uplink subframe, and in order to support both the backhaul and access links at least two uplink subframes are needed. Therefore, of the seven TDD configurations supported in LTE, only configurations 1, 2, 3, 4, and 6 are supported in relay cells. For each TDD configuration, one or several backhaul configurations are supported, as shown in Table 18.1.

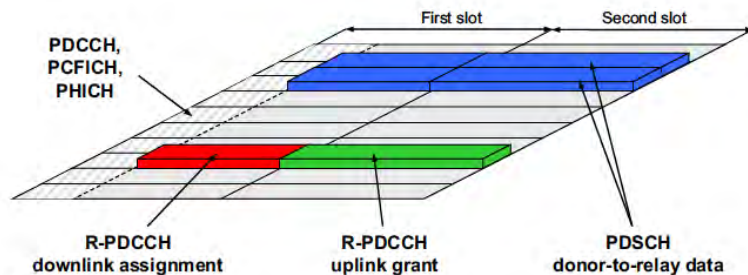
The underlying timing principles for hybrid-ARQ acknowledgements and uplink scheduling grants are, as mentioned above, to keep the same principles as for the access link. However, backhaul transmissions may occur in backhaul subframes only. Therefore, for TDD, the acknowledgement of a transport block on the backhaul link in subframe n is transmitted in subframe $n + k$, where $k \geq 4$ and is selected such that $n + k$ is an uplink *backhaul* subframe if the acknowledgement is to be transmitted from the relay, and a downlink *backhaul* subframe if the acknowledgement is transmitted from the eNodeB.

The numbering of uplink hybrid-ARQ processes on the backhaul link is similar to the TDD numbering on the access link, where uplink hybrid-ARQ process numbers are assigned sequentially to the available backhaul occasions as shown in Figure 18.6, taking into account

Table 18.1 Supported Backhaul Configurations for TDD										
Backhaul Subframe Configuration	Uplink-Downlink Configuration in Relay Cell	Backhaul DL:UL Ratio	Subframe Number							
			0	1	2	3	4	5	6	7
0	1	1:1					D			U
1						U				D
2							D			U
3						U	D			D
4	2	2:2				U	D			U
5					U					D
6						D			U	
7					U		D			D
8	3	2:1				D			U	D
9					U	D	D			D
10						D			U	D
11						U			D	D
12	4	3:1				U			D	D
13						U				D
14						U			D	D
15						U				D
16		4:1				U			D	D
17						U	D		D	D

**FIGURE 18.6**

Example of hybrid-ARQ process numbering for FDD

**FIGURE 18.7**

Example of R-PDCCH transmission

the same processing times as for the access link (see Chapter 12). This is in contrast to the FDD access link, where the uplink hybrid-ARQ process number can be directly derived from the subframe number. The reason for adopting a somewhat different strategy is to minimize the maximum hybrid-ARQ round-trip time. Due to the fact that a pure 8 ms periodicity does not always match the MBSFN allocation, the actual uplink round-trip time, unlike the FDD access link, is not constant but, similar to the TDD access link, is dependent on the subframe number.

18.3.3 Backhaul downlink control signaling

The gap during which the relay can receive transmissions from the donor cell is, as seen in Figure 18.4, shorter than the full subframe duration. In particular, as the first OFDM symbols in the subframe are unavailable for reception of transmissions from the donor cell, L1/L2 control signaling from the donor to the relay cannot be transmitted using the regular PDCCH.¹ Instead, a relay-specific control channel, the R-PDCCH, was introduced in release 10.

The R-PDCCH carries downlink scheduling assignments and uplink scheduling grants, using the same DCI formats as for the PDCCH. However, there is no support for power

¹In principle, the PDCCH could be received if the subframe structures of the access and backhaul links are offset by two to three OFDM symbols, but with the drawback that relay and donor cells would not be time aligned. Time alignment is beneficial, for example, in heterogeneous deployments.

control commands using DCI formats 3/3A. The main function of DCI formats 3/3A is to support semi-persistent scheduling, a feature mainly targeting overhead reduction for low-rate services and not supported for the backhaul link.

In the time domain, the R-PDCCH is, as already mentioned, received in the “MBSFN region” of the subframe, while in the frequency domain, transmission of the R-PDCCH occurs in a set of semi-statically allocated resource blocks. From a latency perspective it is beneficial to locate transmissions of downlink scheduling assignments as early as possible in the subframe. As discussed in Chapter 10, this was the main motivation for dividing normal subframes into a control region and a data region. In principle, a similar approach could be taken for the R-PDCCH, namely dividing the set of resource blocks used for R-PDCCH transmission into a control part and a data part. However, as it is not possible to exploit fractions of a subframe for transmission of PDSCH to terminals connected directly to the donor cell, transmission of a single R-PDCCH could block usage of a relatively large number of resource blocks. From an overhead and scheduling flexibility perspective, a structure where the frequency span of the R-PDCCH is minimized (while still providing sufficient diversity) and resources are allocated mainly in the time dimension is preferable. In the release-10 design of the R-PDCCH, these seemingly contradicting requirements have been addressed through a structure where downlink assignments are located in the first slot and uplink grants, which are less time critical, in the second slot of a subframe (see Figure 18.7). This structure allows the time-critical downlink assignments to be decoded early. To handle the case when there is no uplink grant to transmit to the relay, the R-PDCCH resources in the second slot may be used for PDSCH transmission *to the same relay*.

Coding, scrambling, and modulation for the R-PDCCH follows the same principles as for the PDCCH (see Chapter 10), with the same set of aggregation levels supported (one, two, four, and eight CCEs). However, the mapping of the R-PDCCH to time–frequency resources is, different. Two different mapping methods, illustrated in Figure 18.8, are supported:

- Without cross-interleaving.
- With cross-interleaving.

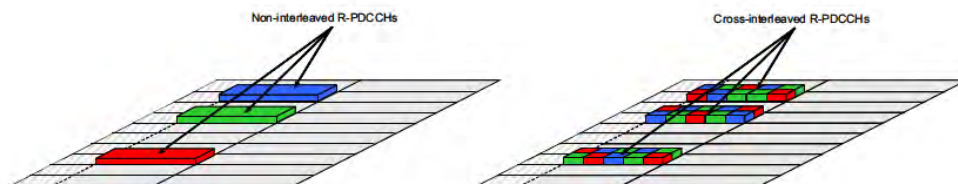


FIGURE 18.8

R-PDCCH mapping types, no cross-interleaving (left) and cross-interleaving (right)

Without cross-interleaving, one R-PDCCH is mapped to one set of virtual resource blocks, where the number of resource blocks (one, two, four, or eight) depends on the aggregation level. No other R-PDCCHs are transmitted using the same set of resource blocks. If the resource blocks are located sufficiently apart in the frequency domain, frequency diversity can be obtained, at least for the higher aggregation levels. Non-interleaved mapping is, for example, useful for beam-forming of the backhaul transmissions or when applying frequency-selective scheduling to the R-PDCCH. Either cell-specific reference signals (CRS) or demodulation reference signals (DM-RS) can be used for demodulation.

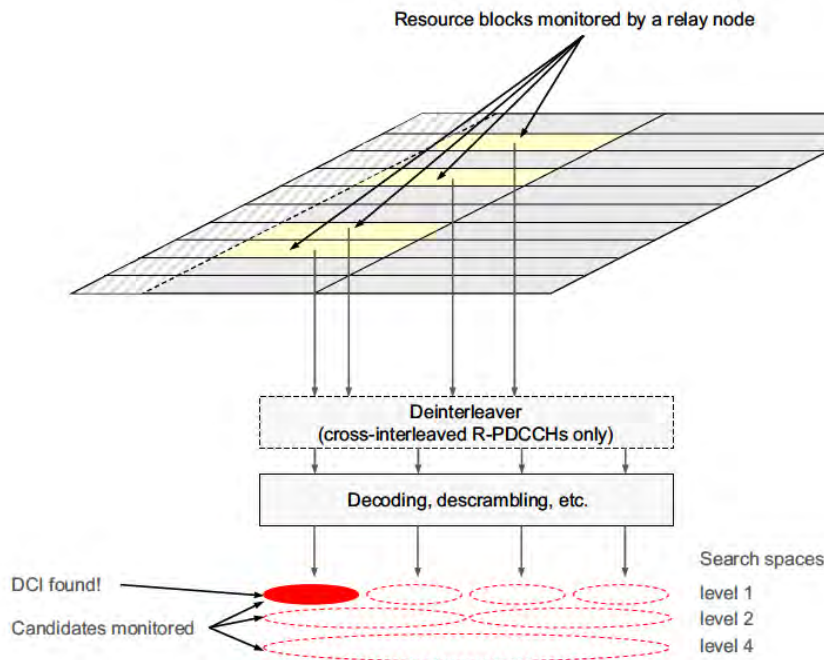
Cross-interleaved mapping is similar to the strategy used for the PDCCH and reuses most of the PDCCH processing structures, except for the mapping to resource elements. A set of R-PDCCHs is multiplexed together, interleaved, and mapped to a set of resource blocks allocated for R-PDCCH transmission. As transmissions to multiple relays may share the same set of resource blocks, cell-specific reference signals are the only possibility for demodulation. The motivation for this mapping method is to obtain frequency diversity for the lowest aggregation level. However, it also comes at the cost of blocking additional resource blocks from PDSCH transmission as, even at low aggregation levels, several resource blocks in the frequency domain are used for the R-PDCCH.

For both mapping cases, cross-interleaved as well as non-cross-interleaved, a set of candidate R-PDCCHs is monitored by the relay node. The set of resource blocks upon which the relay monitors for R-PDCCH transmission is configurable by the donor cell by signaling a set of virtual resource blocks using resource allocation type 0, 1, or 2 (see Chapter 10 for a discussion on resource allocation types). The sets may or may not overlap across multiple relay nodes. In the subframes used for backhaul reception, the relay attempts to receive and decode each of the R-PDCCHs candidates, as illustrated in Figure 18.9 and, if valid downlink control information is found, applies this information to downlink reception or uplink transmission. This approach is in essence similar to the blind decoding procedure used in the terminals, although there are some differences. First, there are no common search spaces for the relays as there is no need to receive broadcast information. Any information necessary for relay operation is transmitted using dedicated signaling. Secondly, the search spaces for the non-interleaved mapping are not time varying as in terminals, but remain static in time.

The number of blind decoding attempts is the same as for a terminal—that is, six, six, two, and two attempts for aggregation levels one, two, four, and eight respectively. However, note that an R-PDCCH can be transmitted in either the first or second slot. Hence, the total number of decoding attempts performed by a relay is 64.²

No PHICH channel is defined for the backhaul. The main reason for the PHICH in release 8 was efficient support of non-adaptive retransmissions for delay-sensitive low-rate applications, such as voice over IP. The backhaul from a relay, on the other hand, typically uses a higher data rate as multiple terminals are served by the relay. Hence, as control signaling

²Two slots and two DCI formats per transmission mode results in $2 \cdot 2 \cdot (6 + 6 + 2 + 2) = 64$.

**FIGURE 18.9**

Principle illustration of R-PDCCH monitoring

overhead is less of an issue, the PHICH was omitted from the backhaul in order to simplify the overall design. Retransmissions are still supported through the use of the R-PDCCH.

18.3.4 Reference signals for the backhaul link

Backhaul reception at the relay can use cell-specific reference signals (CRS) or demodulation reference signals (DM-RS), described in Chapter 10. Different reference-signal types can be used for R-PDCCH and PDSCH, but if the R-PDCCH is received using demodulation reference signals, then demodulation reference signals should be used for PDSCH as well. This is a reasonable restriction as DM-RS for R-PDCCH is motivated by beam-forming. If beam-forming is used for the R-PDCCH there is no incentive not to also use beam-forming for the PDSCH. The opposite scenario, CRS for R-PDCCH and DM-RS for the PDSCH, does make sense though. One example is interleaved mapping of the control signaling, where multiple R-PDCCHs are multiplexed and individual beam-forming cannot be used, together with beam-forming of the PDSCH. The different combinations of reference signals supported for the backhaul link are summarized in Table 18.2.

Table 18.2 Combinations of Reference Signals and R-PDCCH Mapping Schemes

Reference Signal Type Used for Demodulation of		R-PDCCH Mapping Scheme
R-PDCCH	PDSCH	
CRS	CRS	Cross-interleaved or non-cross-interleaved
CRS	DM-RS	Cross-interleaved or non-cross-interleaved
DM-RS	DM-RS	Non-cross-interleaved

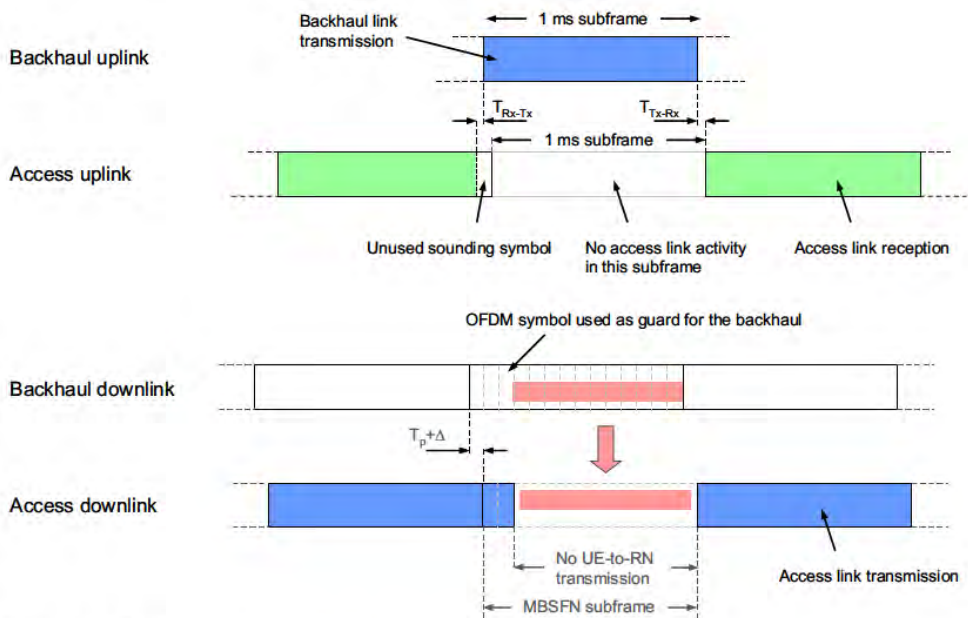
Note also that in the case of (global) time alignment between the donor and relay cell, the last OFDM symbol cannot be received by the relay as it is needed for reception-transmission switching. Hence, the demodulation reference signals on the last OFDM symbols in the subframe cannot be received. For transmission ranks up to 4 this is not a problem, as the necessary reference signals are also available earlier in the subframe. However, for spatial multiplexing with five or more layers, the first set of reference signals in the subframe is used for the lower layers while the second set of reference signals, located at the end of the subframe and that cannot be received, is used for the higher layers. This implies that reference signals for rank 5 and higher cannot be received by the relay, and backhaul transmissions are therefore restricted to at most four-layer spatial multiplexing, irrespective of the timing relation used.

18.3.5 Backhaul–access link timing

To ensure that the relay is able to receive transmissions from the donor cell, some form of timing relation between the downlink transmissions in the donor and relay cells must be defined, including any guard time needed to allow the relay to switch between access-link transmission to backhaul-link reception and vice versa.

A natural choice for the timing of the access link is to synchronize it to the frame timing of the backhaul link as observed by the relay. From this backhaul downlink timing reference, the timing of the access-link transmission is derived as shown at the bottom of Figure 18.10. The backhaul uplink timing is subject to the normal timing advance controlled by the donor cell, ensuring that the backhaul uplink transmissions are time aligned with other uplink transmissions received by the donor base station.

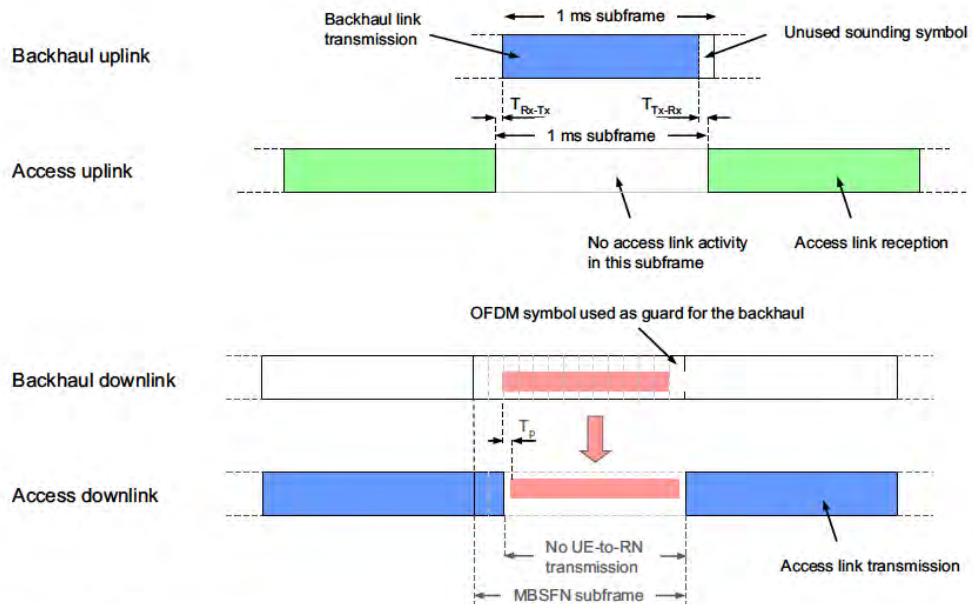
In the backhaul downlink, the first OFDM symbol in the data region is left unused to provide the guard time for relay switching, and a small time offset is used to distribute the guard between Tx–Rx and Rx–Tx switching at the relay. This case is shown at the bottom of Figure 18.11. Locating the guard symbol at the beginning of the data region instead of at the end is beneficial as the guard symbol is needed at the relay side only and can therefore still be used for transmission of PDCCHs to terminals in the donor cell. In principle, the guard time comes “for free” from a donor cell perspective, and the freedom in shifting the relay node frame timing relative to the donor cell timing is used to move the “free” guard period to where it is needed.

**FIGURE 18.10**

Backhaul timing relations in case the relay cell timing is derived from the backhaul timing

The backhaul uplink is subject to the normal timing advance controlled by the donor cell, ensuring that the backhaul uplink transmissions are time aligned with other uplink transmissions received by the donor base station. Similarly to the guard time needed to switch from access-link transmission to backhaul-link reception, which influenced the downlink timing relation between the access and backhaul links, there may also be the need for a guard time in the uplink direction to switch from access-link reception to backhaul-link transmission. However, unlike the downlink case, how to handle this is not standardized but left for implementation, noting that functionality already present in release 8 is sufficient for providing the necessary guard time.

In principle, if the relay could switch from access-link reception to backhaul-link transmission within the cyclic prefix, no provisions for additional switching time would be necessary. However, the switching time is implementation dependent and typically larger than the cyclic prefix. For larger switching times, one possibility is to use the shortened transmission format on the access link, originally intended for sounding, as shown in the top part of Figure 18.10. By configuring all the terminals in the relay cell to reserve the last OFDM symbol of the preceding subframe for sounding-reference signals but not to transmit

**FIGURE 18.11**

Backhaul timing relations in access link transmission in the relay and donor cells are time synchronized

any sounding-reference signals, a guard period of one OFDM symbol is created. This guard time can then be divided into Rx–Tx and Tx–Rx switching times through a time offset between the frame timing of the backhaul and access links.

For some deployments, it is desirable to align the access-link transmission timing of the relay with the transmission timing of the donor cell—that is, to use a global timing reference for all the cells. One example hereof is TDD. In such deployments, the necessary guard times are obtained in a slightly different manner compared to the case of using reception timing of the backhaul downlink. In this case it is not possible to obtain the necessary guard time by shifting the subframe timing at the relay. Hence, the guard time for switching from backhaul-link reception to access-link transmission will also be visible at the donor cell, as the last OFDM symbol in the resource blocks used for the backhaul transmission cannot be used for other transmissions in the relay cell. If the time for Tx–Rx switching is longer than the donor-cell-to-relay-node propagation delay, then the first OFDM symbol has to be left unused as well. This case is shown in the bottom part of Figure 18.11.

In the backhaul uplink, the guard time necessary, similar to the previous timing case, is obtained through configuration of (unused) sounding instances. However, unlike the previous

case, sounding is configured in the *backhaul* link, as shown at the top of Figure 18.11. Note that this implies that sounding cannot be used for the backhaul link, as the OFDM symbol intended as a sounding-reference symbol is used as guard time.

In the case of TDD operation, guard time for the access-backhaul switch can, in addition to the methods discussed above, be obtained from the guard period required for TDD operation itself. This is shown in Figure 18.12 and is a matter of using the appropriate settings of timing advance and timing offsets.

Backhaul downlink transmissions consist of data transmitted on the PDSCH and L1/L2 control signaling transmitted on the R-PDCCH, as already discussed. Both these types of transmission must follow one of the timing scenarios discussed above. In order to allow for different implementations and deployments, the LTE specifications provide not only the possibility to configure which of the two access-backhaul downlink timing relations to use, but also flexibility in terms of the time span of the channels transmitted on the backhaul link.

PDSCH transmissions intended for a relay can be semi-statically configured to start on the second, third, or fourth OFDM symbol to cater for different control region sizes in the donor cell and relay cells. The PDSCH transmission ends at the last or second last OFDM symbol, depending on which of the two timing cases above is used.

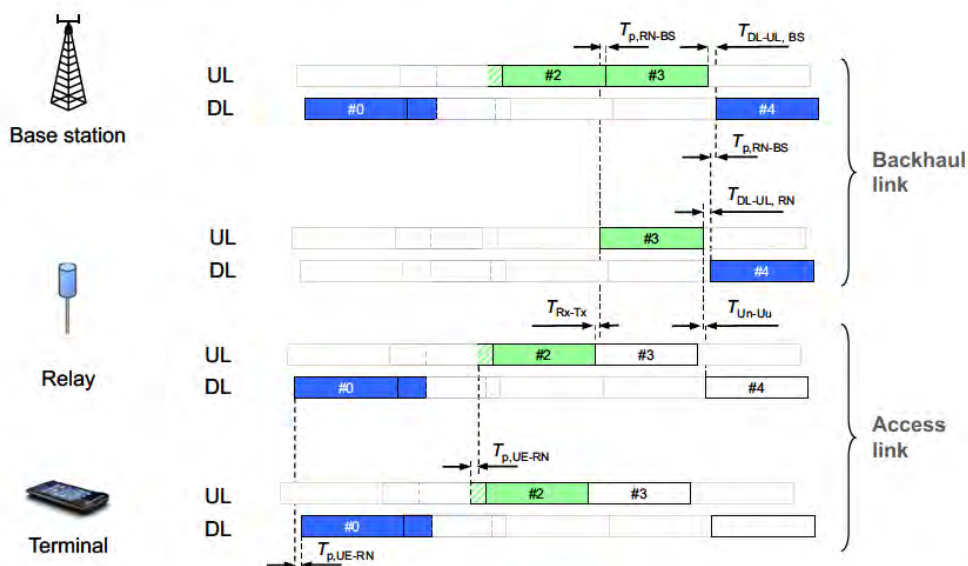


FIGURE 18.12

Example of uplink timing relation for TDD

R-PDCCH transmissions intended for a relay always start at the fourth OFDM symbol. A fixed starting position was chosen to simplify the overall structure. Since the amount of resource blocks occupied by an R-PDCCH is relatively small compared to the PDSCH, the overhead reduction possible with a configurable starting position is small and does not justify the additional specification and testing complexity.

This page intentionally left blank

Spectrum and RF Characteristics

19

Spectrum flexibility is, as mentioned in Chapter 7, a key feature of LTE radio access and is set out in the LTE design targets [10]. It consists of several components, including deployment in different-sized spectrum allocations and deployment in diverse frequency ranges, both in paired and unpaired frequency bands.

There are a number of frequency bands identified for mobile use and specifically for IMT today. Many of these bands were first defined for operation with WCDMA/HSPA, but are now shared also with LTE deployments. However, new bands are today often defined for LTE only in 3GPP. Both paired and unpaired bands are included in the LTE specifications. The additional challenge with LTE operation in some bands is the possibility of using channel bandwidths up to 20 MHz with a single carrier and even beyond that with aggregated carriers.

The use of OFDM in LTE gives flexibility both in terms of the size of the spectrum allocation needed and in the instantaneous transmission bandwidth used. The OFDM physical layer also enables frequency-domain scheduling, as briefly discussed in Chapter 7. Beyond the physical layer implications described in Chapters 10 and 11, these properties also impact the RF implementation in terms of filters, amplifiers, and all other RF components that are used to transmit and receive the signal. This means that the RF requirements for the receiver and transmitter will have to be expressed with flexibility in mind.

19.1 Spectrum for LTE

LTE can be deployed both in existing IMT bands and in future bands that may be identified. The possibility of operating radio-access technology in different frequency bands is, in itself, nothing new. For example, 2G and 3G terminals are multi-band capable, covering bands used in the different regions of the world to provide global roaming. From a radio-access functionality perspective, this has no or limited impact and the LTE physical-layer specifications [61–64] do not assume any specific frequency band. What may differ, in terms of specification, between different bands is mainly the more specific RF requirements, such as the allowed maximum transmit power, requirements/limits on out-of-band (OOB) emission, and so on. One reason for this is that external constraints, imposed by regulatory bodies, may differ between different frequency bands.

4G: LTE/LTE-Advanced for Mobile Broadband. <http://dx.doi.org/10.1016/B978-0-12-419985-9.00019-2>

Copyright © 2014 Erik Dahlman, Stefan Parkvall and Johan Sköld. Published by Elsevier Ltd. All rights reserved.

431

19.1.1 Spectrum defined for IMT systems by the ITU-R

The global designations of spectrum for different services and applications are done within the ITU-R. The *World Administrative Radio Congress* WARC-92 identified the bands 1885–2025 and 2110–2200 MHz as intended for implementation of IMT-2000. Of these 230 MHz of 3G spectrum, 2×30 MHz were intended for the satellite component of IMT-2000 and the rest for the terrestrial component. Parts of the bands were used during the 1990s for deployment of 2G cellular systems, especially in the Americas. The first deployment of 3G in 2001–2002 by Japan and Europe was done in this band allocation, and for that reason it is often referred to as the IMT-2000 “core band.”

Additional spectrum for IMT-2000 was identified at the World Radio-communication Conference WRC-2000, where it was considered that an additional need for 160 MHz of spectrum for IMT-2000 was forecasted by the ITU-R. The identification includes the bands used for 2G mobile systems at 806–960 and 1710–1885 MHz, and “new” 3G spectrum in the bands at 2500–2690 MHz. The identification of bands previously assigned for 2G was also recognition of the evolution of existing 2G mobile systems into 3G. Additional spectrum was identified at WRC’07 for IMT, encompassing both IMT-2000 and IMT-Advanced. The bands added were 450–470, 698–806, 2300–2400, and 3400–3600 MHz, but the applicability of the bands varies on a regional and national basis. At WRC’12 there were no additional spectrum allocations identified for IMT, but the issue was put on the agenda for WRC’15. It was also determined to study the use of the band 694–790 MHz for mobile services in Region 1 (Europe, Middle East, and Africa).

The somewhat diverging arrangement between regions of the frequency bands assigned to IMT means that there is not one single band that can be used for 3G and 4G roaming worldwide. Large efforts have, however, been put into defining a minimum set of bands that can be used to provide roaming. In this way, multi-band devices can provide efficient worldwide roaming for 3G and 4G devices.

19.1.2 Frequency bands for LTE

The frequency bands where LTE will operate are in both paired and unpaired spectrum, requiring flexibility in the duplex arrangement. For this reason, LTE supports both FDD and TDD operation, as discussed in the previous chapters.

Release 11 of the 3GPP specifications for LTE includes 27 frequency bands for FDD and twelve for TDD. The paired bands for FDD operation are numbered from 1 to 29 [76], as shown in [Table 19.1](#), while the unpaired bands for TDD operation are numbered from 33 to 44, as shown in [Table 19.2](#). Note that the frequency bands for UTRA FDD use the same numbers as the paired LTE bands, but are labeled with Roman numerals. All bands for LTE are summarized in [Figures 19.1 and 19.2](#), which also show the corresponding frequency allocation defined by the ITU.

Some of the frequency bands are partly or fully overlapping. In most cases this is explained by regional differences in how the bands defined by the ITU are implemented.

Table 19.1 Paired Frequency Bands Defined by 3GPP for LTE

Band	Uplink Range (MHz)	Downlink Range (MHz)	Main Region(s)
1	1920–1980	2110–2170	Europe, Asia
2	1850–1910	1930–1990	Americas (Asia)
3	1710–1785	1805–1880	Europe, Asia (Americas)
4	1710–1755	2110–2155	Americas
5	824–849	869–894	Americas, Asia
6	830–840	875–885	Japan (only for UTRA)
7	2500–2570	2620–2690	Europe, Asia
8	880–915	925–960	Europe, Asia
9	1749.9–1784.9	1844.9–1879.9	Japan
10	1710–1770	2110–2170	Americas
11	1427.9–1447.9	1475.9–1495.9	Japan
12	698–716	728–746	US
13	777–787	746–756	US
14	788–798	758–768	US
17	704–716	734–746	US
18	815–830	860–875	Japan
19	830–845	875–890	Japan
20	832–862	791–821	Europe
21	1447.9–1462.9	1495.9–1510.9	Japan
22	3410–3490	3510–3590	Europe
23	2000–2020	2180–2200	Americas
24	1626.5–1660.5	1525–1559	Americas
25	1850–1915	1930–1995	Americas
26	814–849	859–894	Americas
27	807–824	852–869	Americas
28	703–748	758–803	Asia/Pacific
29	N/A	717–728	Americas

At the same time, a high degree of commonality between the bands is desired to enable global roaming. A set of bands was first specified as bands for UTRA, with each band originating in global, regional, and local spectrum developments. The complete set of UTRA bands was then transferred to the LTE specifications in release 8, and additional ones have been added since then in later releases.

Bands 1, 33, and 34 are the same paired and unpaired bands that were defined first for UTRA in release 99 of the 3GPP specifications, also called the 2 GHz “core band.” *Band 2* was added later for operation in the US PCS1900 band and *Band 3* for 3G operation in the GSM1800 band. The unpaired *Bands 35, 36, and 37* are also defined for the PCS1900

Table 19.2 Unpaired Frequency Bands Defined by 3GPP for LTE

Band	Frequency Range (MHz)	Main Region(s)
33	1900–1920	Europe, Asia (not Japan)
34	2010–2025	Europe, Asia
35	1850–1910	(Americas)
36	1930–1990	(Americas)
37	1910–1930	—
38	2570–2620	Europe
39	1880–1920	China
40	2300–2400	Europe, Asia
41	2496–2690	US
42	3400–3600	Europe
43	3600–3800	Europe
44	703–803	Asia/Pacific

frequency ranges, but are not deployed anywhere today. *Band 39* is an extension of the unpaired *Band 33* from 20 to 40 MHz for use in China.

Band 4 was introduced as a new band for the Americas following the addition of the 3G bands at WRC-2000. Its downlink overlaps completely with the downlink of *Band 1*, which facilitates roaming and eases the design of dual *Band 1 + 4* terminals. *Band 10* is an extension of *Band 4* from 2×45 to 2×60 MHz.

Band 9 overlaps with *Band 3*, but is intended only for Japan. The specifications are drafted in such a way that implementation of roaming dual *Band 3 + 9* terminals is possible. The 1500 MHz frequency band is also identified in 3GPP for Japan as *Bands 11 and 21*. It is allocated globally to mobile service on a co-primary basis and was previously used for 2G in Japan.

With WRC-2000, the band 2500–2690 MHz was identified for IMT-2000 and it is identified as *Band 7* in 3GPP for FDD and *Band 38* for TDD operation in the “center gap” of the FDD allocation. The band has a slightly different arrangement in North America, where a US-specific *Band 41* is defined. *Band 40* is an unpaired band specified for the new frequency range 2300–2400 MHz identified for IMT and has a widespread allocation globally.

WRC-2000 also identified the frequency range 806–960 MHz for IMT-2000, complemented by the frequency range 698–806 MHz in WRC’07. As shown in Figure 19.2, several bands are defined for FDD operation in this range. *Band 8* uses the same band plan as GSM900. *Bands 5, 18, 19, 26, and 27* overlap, but are intended for different regions. *Band 5* is based on the US cellular band, while *Bands 18 and 19* are restricted to Japan in the

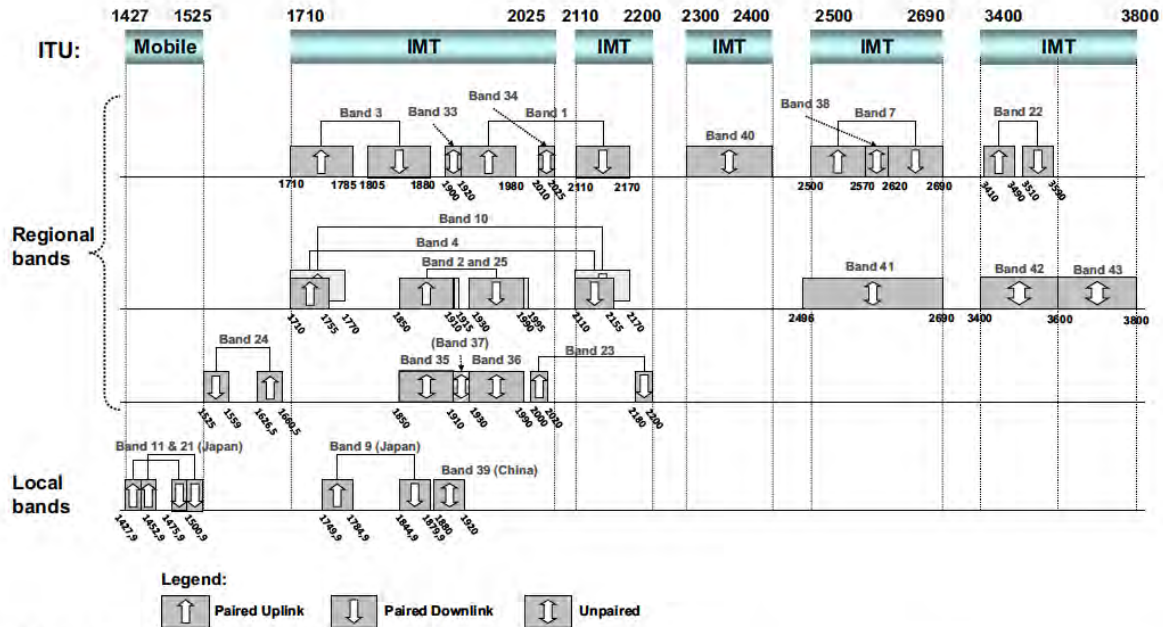
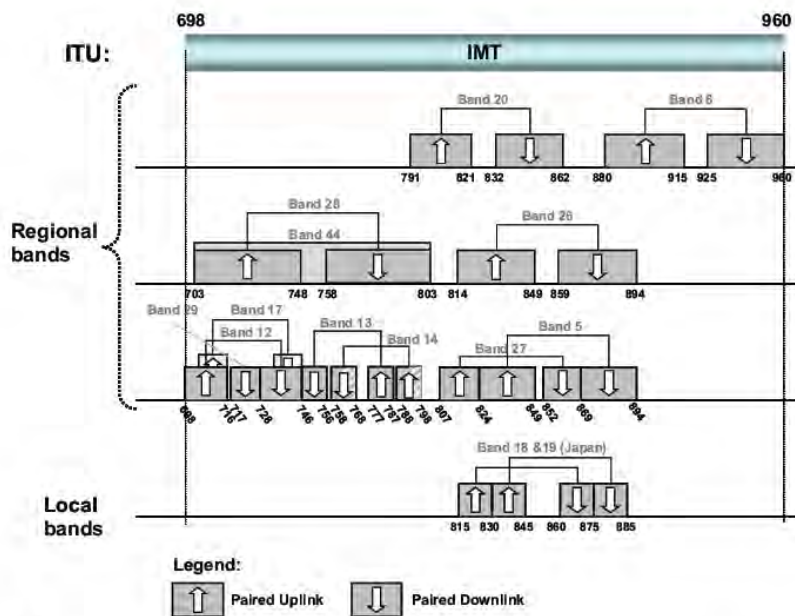


FIGURE 19.1

Operating bands specified for LTE in 3GPP above 1 GHz and the corresponding ITU allocation

**FIGURE 19.2**

Operating bands specified for LTE in 3GPP below 1 GHz and the corresponding ITU allocation

specifications. 2G systems in Japan had a very specific band plan and Bands 18 and 19 are a way of partly aligning the Japanese spectrum plan in the 810–960 MHz range to that in other parts of the world. Note that Band 6 was originally defined in this frequency range for Japan, but it is not used for LTE.

An extensive study was performed in 3GPP to create an extension of Band 5 (850 MHz), which is one of the bands with the most widespread deployment globally. The extension adds additional frequency ranges below the present Band 5 and is done with two new operating bands. Band 26 is the “Upper Extending 850 MHz” band, which encompasses the band 5 range, adding 2x10 MHz to create an extended 2x35 MHz band. Band 27 is the “Lower Extending 850 MHz” band, which consists of the 2x17 MHz frequency range right below and adjacent to Band 5.

Bands 12, 13, 14, and 17 make up the first set of bands defined for what is called the *digital dividend*—that is, for spectrum previously used for broadcasting. This spectrum is partly migrated to be used by other wireless technologies, since TV broadcasting is migrating from analog to more spectrum-efficient digital technologies. Other regional bands for the digital dividend are Band 20 that is defined in Europe and Band 28 for the Asia/Pacific region. An alternative unpaired arrangement in the Asia/Pacific region is the unpaired Band 44.

Band 29 is the only “paired” band that consists of a downlink without an identified uplink. It is intended for carrier aggregation, primarily with other downlink carriers in Band 2, 4, and 5.

The paired *Band 22* and unpaired *Band 42 and 43* are specified for the frequency range 3.4–3.8 GHz [78]. In Europe, a majority of countries already license the band at 3.4–3.6 GHz for both Fixed Wireless Access and mobile use, and there is a European spectrum decision for 3.4–3.8 GHz with “flexible usage modes” for deployment of fixed, nomadic, and mobile networks. In Japan, not only 3.4–3.6 GHz but also 3.6–4.2 GHz will be available to terrestrial mobile services in the future. The band 3.4–3.6 GHz has also been licensed for wireless access in Latin America.

Several Mobile Satellite Service (MSS) operators in the USA are planning to deploy an Ancillary Terrestrial Component (ATC) using LTE. For this purpose two new frequency bands are defined, *Band 23* with 2×20 MHz band for the S-band MSS operators at 2 GHz, and *Band 24* with 2×34 MHz band for the L-band MSS operators at 1.5 GHz.

19.1.3 New frequency bands

Additional frequency bands are continuously specified for UTRA and LTE. WRC’07 identified additional frequency bands for IMT, which encompasses both IMT-2000 and IMT-Advanced. Several of the bands defined by WRC’07 are already available for LTE as described previously, or will become available partly or fully for deployment on a global basis:

- 450–470 MHz was identified for IMT globally. It is already allocated to mobile service globally, but it is only 20 MHz wide and has a number of different arrangements.
- 698–806 MHz was allocated to mobile service and identified for IMT to some extent in all regions. Together with the band at 806–960 MHz identified at WRC-2000, it forms a wide frequency range from 698 to 960 MHz that is partly identified to IMT in all regions, with some variations. A number of LTE bands are defined in this frequency.
- 2300–2400 MHz was identified for IMT on a worldwide basis in all three regions. It is defined as LTE Band 40.
- 3400–3600 MHz was allocated to the mobile service on a primary basis in Europe and Asia and partly in some countries in the Americas. There is also satellite use in the bands today. It is defined as LTE Bands 22, 42, and 43.

Additional bands for IMT were not on the agenda for WRC’12, but are on the agenda for WRC’15. For the frequency ranges below 1 GHz identified at WRC-07, 3GPP has already specified several operating bands, as shown in [Figure 19.2](#). The bands with the widest use are Bands 5 and 8, while most of the other bands have regional or more limited use. With the identification of bands down to 698 MHz for IMT use and the switchover from analog to digital TV broadcasting, Bands 12, 13, 14, and 17 are defined in the United States, Band 20 in Europe, and Bands 29 and 44 in Asia/Pacific for the digital dividend.

For release 12, 3GPP is defining one frequency arrangement for the 450 MHz band for Brazil. There is also work ongoing to specify a paired 2x10 MHz arrangement in the 2300-2400MHz band for the United States, called the *Wireless Communications Service* (WCS) Band.

19.2 Flexible spectrum use

Most of the frequency bands identified previously for deployment of LTE are existing IMT-2000 bands and some bands also have legacy systems deployed, including WCDMA/HSPA and GSM. Bands are also in some regions defined in a “technology neutral” manner, which means that coexistence between different technologies is a necessity.

The fundamental LTE requirement to operate in different frequency bands [79] does not, in itself, impose any specific requirements on the radio-interface design. There are, however, implications for the RF requirements and how those are defined, in order to support the following:

- *Coexistence between operators in the same geographical area in the band.* These other operators may deploy LTE or other IMT-2000 technologies, such as UMTS/HSPA or GSM/EDGE. There may also be non-IMT-2000 technologies. Such coexistence requirements are to a large extent developed within 3GPP, but there may also be regional requirements defined by regulatory bodies in some frequency bands.
- *Co-location of base station equipment between operators.* There are in many cases limitations to where base-station equipment can be deployed. Often, sites must be shared between operators or an operator will deploy multiple technologies in one site. This puts additional requirements on both base-station receivers and transmitters.
- *Coexistence with services in adjacent frequency bands and across country borders.* The use of the RF spectrum is regulated through complex international agreements, involving many interests. There will therefore be requirements for coordination between operators in different countries and for coexistence with services in adjacent frequency bands. Most of these are defined in different regulatory bodies. Sometimes the regulators request that 3GPP includes such coexistence limits in the 3GPP specifications.
- *Coexistence between operators of TDD systems* in the same band is provided by inter-operator synchronization, in order to avoid interference between downlink and uplink transmissions of different operators. This means that all operators need to have the same downlink/uplink configurations and frame synchronization, not in itself an RF requirement, but it is implicitly assumed in the 3GPP specifications. RF requirements for unsynchronized systems become much stricter.
- *Release-independent frequency-band principles.* Frequency bands are defined regionally and new bands are added continuously. This means that every new release of 3GPP specifications will have new bands added. Through the “release independence” principle,

it is possible to design terminals based on an early release of 3GPP specifications that support a frequency band added in a later release.

- *Aggregation of spectrum allocations.* Operators of LTE systems have quite diverse spectrum allocations, which in many cases do not consist of a block that easily fits exactly one LTE carrier. The allocation may even be non-contiguous, consisting of multiple blocks spread out in a band. Many operators also have allocations in multiple bands to use for LTE deployment. For these scenarios, the LTE specifications support *Carrier Aggregation*, where multiple carriers in contiguous or non-contiguous blocks within a band, or in multiple bands, can be combined to create larger transmission bandwidths.

19.3 Flexible channel bandwidth operation

The frequency allocations in Figures 19.1 and 19.2 are up to 2×75 MHz, but the spectrum available for a single operator may be from 2×20 MHz down to 2×5 MHz for FDD and down to 1×5 MHz for TDD. Furthermore, the migration to LTE in frequency bands currently used for other radio-access technologies must often take place gradually to ensure that a sufficient amount of spectrum remains to support the existing users. Thus, the amount of spectrum that can initially be migrated to LTE can be relatively small, but may then gradually increase, as shown in Figure 19.3. The variation of possible spectrum scenarios implies a requirement for spectrum flexibility for LTE in terms of the transmission bandwidths supported.

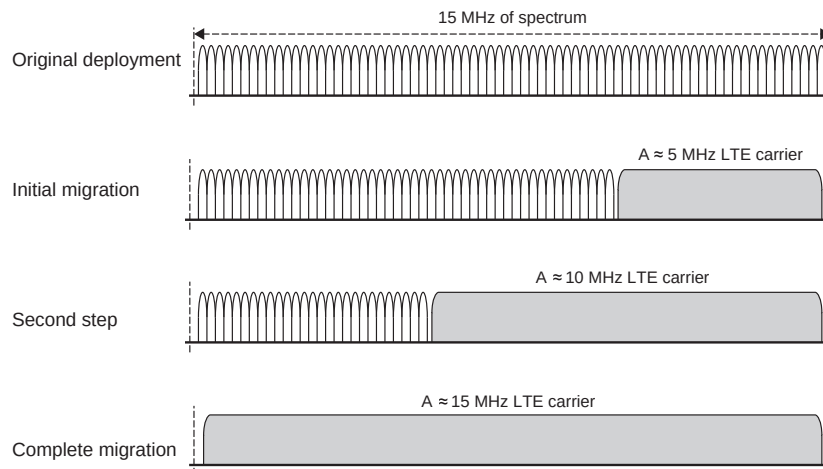


FIGURE 19.3

Example of how LTE can be migrated step-by-step into a spectrum allocation with an original GSM deployment

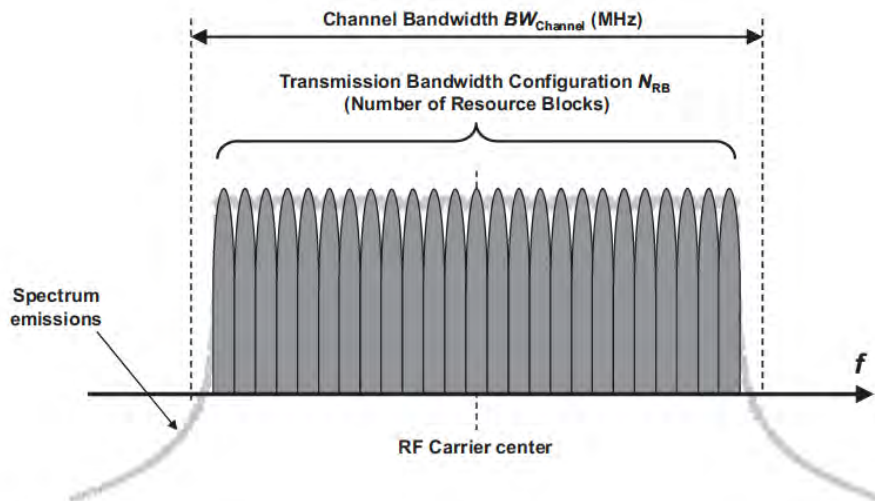
The spectrum flexibility requirement points out the need for LTE to be scalable in the frequency domain. This flexibility requirement is stated in [10] as a list of LTE spectrum allocations from 1.25 to 20 MHz. Note that the final channel bandwidths selected differ slightly from this initial assumption.

As shown in Chapter 9, the frequency-domain structure of LTE is based on resource blocks consisting of 12 subcarriers with a total bandwidth of $12 \times 15 \text{ kHz} = 180 \text{ kHz}$. The basic radio-access specification including the physical-layer and protocol specifications enables *transmission bandwidth configurations* from six up to 110 resource blocks on one LTE RF carrier. This allows for channel bandwidths ranging from 1.4 MHz up to beyond 20 MHz in steps of 180 kHz and is fundamental to providing the required spectrum flexibility.

In order to limit implementation complexity, only a limited set of bandwidths are defined in the RF specifications. Based on the frequency bands available for LTE deployment today and in the future, as described above, and considering the known migration and deployment scenarios in those bands, a limited set of six channel bandwidths is specified. The RF requirements for the base station and terminal are defined only for those six channel bandwidths. The channel bandwidths range from 1.4 to 20 MHz, as shown in Table 19.3. The lower bandwidths, 1.4 and 3 MHz, are chosen specifically to ease migration to LTE in spectrum where CDMA2000 is operated, and also to facilitate migration of GSM and TD-SCDMA to LTE. The specified bandwidths target relevant scenarios in different frequency bands. For this reason, the set of bandwidths available for a specific band is not necessarily the same as in other bands. At a later stage, if new frequency bands are made available that have other spectrum scenarios requiring additional channel bandwidths, the corresponding RF parameters and requirements can be added in the RF specifications, without actually having to update the physical-layer specifications. The process of adding new channel bandwidths would in this way be similar to adding new frequency bands.

Figure 19.4 illustrates in principle the relationship between the channel bandwidth and the number of resource blocks N_{RB} for one RF carrier. Note that for all channel bandwidths except 1.4 MHz, the resource blocks in the transmission bandwidth configuration fill up 90% of the channel bandwidth. The spectrum emissions shown in Figure 19.4 are for a pure OFDM

Table 19.3 Channel Bandwidths Specified in LTE	
Channel Bandwidth (BW_{Channel})	Number of Resource Blocks (N_{RB})
1.4 MHz	6
3 MHz	15
5 MHz	25
10 MHz	50
15 MHz	75
20 MHz	100

**FIGURE 19.4**

The channel bandwidth for one RF carrier and the corresponding transmission bandwidth configuration

signal, while the actual transmitted emissions will also depend on the transmitter RF chain and other components. The emissions outside the channel bandwidth are called *unwanted emissions* and the requirements for those are discussed further below.

19.4 Carrier aggregation for LTE

The possibility from 3GPP release 10 to aggregate two or more component carriers in order to support wider transmission bandwidths has several implications for the RF characteristics. The impact for the base station and terminal RF characteristics are also quite different. Release 10 has some restrictions on carrier aggregation in the RF specification, compared to what has been specified for physical layer and signaling, while in release 11 there is support for carrier-aggregation within and between a much larger number of bands.

There is, from an RF point of view, a substantial difference between the two types of Carrier Aggregation (CA) defined for LTE (see also Section 7.3.1 for more details):

- **Intra-Band Carrier Aggregation** implies that two or more carriers within the same operating band are aggregated (see also the first two examples in Figure 7.4). RF requirements are restricted in release 10 to contiguous intra-band aggregation and a maximum of two carriers, while non-contiguous carrier aggregation is possible in release

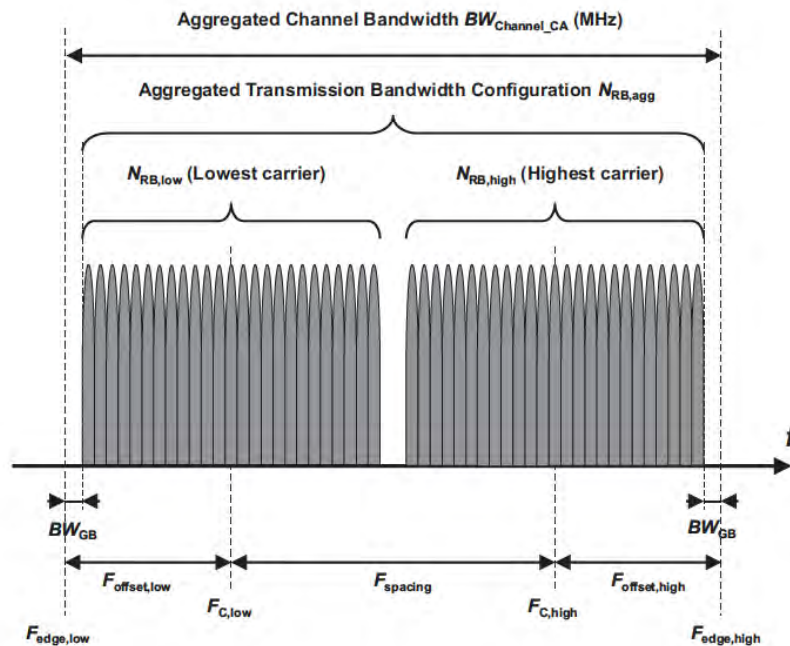
11. Since aggregated carriers from an RF perspective have similar RF properties as a corresponding wider carrier being transmitted and received, there are many implications for the RF requirements. This is especially true for the terminal. For the base station, it corresponds in practice to a multicarrier configuration (non-aggregated) already supported in earlier releases, which also means that the impact is less than for the terminal. For *non-contiguous* carrier aggregation, there is additional impact in terms of defining the requirements in the “gap” between aggregated sub-blocks (see [Section 19.4](#)).

- **Inter-Band Carrier Aggregation** implies that carriers in different operating bands are aggregated (see also the last example in [Figure 7.4](#)). Many RF properties within a band can, to a large extent, remain the same as for a single carrier case. There is, however, impact for the terminal, due to the possibility for intermodulation and cross-modulation within the terminal when multiple transmitter and/or receiver chains are operated simultaneously. For the base station it has very little impact, since in practice it corresponds to a base station supporting multiple bands. There is however additional BS impact, if the inter-band carrier aggregation is deployed with a multi-band base station (see [Section 19.13](#)).

Intra-band carrier aggregation is limited to two component carriers in two paired bands (Band 1 and 7) and three unpaired bands (Bands 38, 40, and 41) in release 11. Inter-band carrier aggregation is specified for 21 different band combinations in release 11. There is also one band defined (Band 25) in which the terminal can operate with non-contiguous carrier aggregation. The band or set of bands over which carriers are aggregated is defined as a UE capability (the term UE, User Equipment, is used in 3GPP specifications instead of terminal).

For intra-band carrier aggregation, the definitions of BW_{channel} and N_{RB} shown in [Figure 19.4](#) still apply for each component carrier, while new definitions are needed for the Aggregated Channel Bandwidth ($BW_{\text{Channel_CA}}$) and the Aggregated Transmission bandwidth Configuration ($N_{\text{RB,agg}}$) shown in [Figure 19.5](#). In connection with this, a new capability is defined for the terminal called *Carrier Aggregation Bandwidth Class*. There are six classes, where each class corresponds to a range for $N_{\text{RB,agg}}$ and a maximum number of component carriers, as shown in [Table 19.4](#). The classes corresponding to aggregation of more than two component carriers or consisting of more than 200 RBs are under study for later releases.

The terminal capability *E-UTRA CA Configuration* [76] is defined as a combination of the operating band (or bands) in which the terminal can operate with carrier aggregation and a bandwidth class. For example, the terminal capability to operate with inter-band carrier aggregation in Bands 1 and 5 in Bandwidth Class A is called CA_1A_5A. For each E-UTRA CA configuration, one or more *Bandwidth Combination Sets* is defined, setting the channel bandwidths that can be used in each band, and what the maximum aggregated bandwidth is. A terminal can declare capability to support multiple bandwidth combination sets.

**FIGURE 19.5**

Definitions for Intra-band Carrier Aggregation RF parameters, example with two aggregated carriers

Table 19.4 UE Carrier Aggregation Bandwidth Classes

Channel Aggregation Bandwidth Classes	Aggregated Transmission BW Configuration	Number of Component Carriers
A	≤ 100	1
B	≤ 100	2
C	101 to 200	2
D, E, F	Under study (201 to 500)	Under study

A fundamental parameter for intra-band CA is the channel spacing. A tighter channel spacing than the nominal spacing for any two single carriers could potentially lead to an increase in spectral efficiency, since there would be a smaller unused “gap” between carriers. On the other hand, there is also a requirement for the possibility to support legacy single-carrier terminals of earlier releases. An additional complication is that

the component carriers should be on the same 15 kHz subcarrier raster in order to allow reception of multiple adjacent component carriers using a single FFT instead of an FFT per subcarrier.¹ As discussed in Section 9.5, this property, together with the fact that the frequency numbering scheme is on a 100 kHz raster, results in the spacing between two component carriers having to be a multiple of 300 kHz, which is the least common denominator of 15 and 100 kHz.

For the specification, RF requirements are based on a nominal channel spacing that is derived from the channel bandwidth of the two adjacent carriers $BW_{\text{Channel}(1)}$ and $BW_{\text{Channel}(2)}$ as follows:²

$$F_{\text{Spacing,Nominal}} = \left\lfloor \frac{BW_{\text{Channel}(1)} + BW_{\text{Channel}(2)} - 0.1 |BW_{\text{Channel}(1)} - BW_{\text{Channel}(2)}|}{2 \cdot 0.3} \right\rfloor 0.3 \quad (19.1)$$

In order to allow for a tighter packing of component carriers, the value of F_{Spacing} can be adjusted to any multiple of 300 kHz that is smaller than the nominal spacing, as long as the subcarriers do not overlap.

RF requirements for LTE are normally defined relative to the channel bandwidth edges. For intra-band CA, this is generalized so that requirements are defined relative to the edges of the Aggregated Channel Bandwidth, identified in Figure 19.5 as $F_{\text{edge,low}}$ and $F_{\text{edge,high}}$. In this way many RF requirements can be reused, but with new reference points in the frequency domain. The aggregated channel bandwidth for both terminal and base station is defined as:

$$BW_{\text{Channel_CA}} = F_{\text{edge,high}} - F_{\text{edge,low}} \quad (19.2)$$

The location of the edges is defined relative to the carriers at the edges through a new parameter F_{offset} (see Figure 19.5) using the following relation to the carrier center positions F_C of the lowest and highest carriers:

$$F_{\text{edge,low}} = F_{C,\text{low}} - F_{\text{offset,low}} \quad (19.3)$$

$$F_{\text{edge,high}} = F_{C,\text{high}} + F_{\text{offset,high}} \quad (19.4)$$

The value of F_{offset} for the edge carriers and the corresponding location of the edges are, however, not defined in the same way for terminal and base station.

For the base station, there are legacy scenarios where the base station receives and transmits adjacent independent carriers, supporting legacy terminals of earlier releases using single carriers. This scenario will also have to be supported for a configuration of aggregated carriers. In addition, for backward s-compatibility reasons, a fundamental parameter such as channel bandwidth and the corresponding reference points (the channel edge) for all RF

¹In case of independent frequency errors between component carriers, multiple FFTs and frequency-tracking functionality may be needed anyway.

²[...] denotes the “floor” operator, which rounds the number down.

requirements will have to remain the same. The implication is that the channel edges shown in Figure 19.4 for each component carrier will also remain as reference points when the carriers are aggregated. This results in the following base station definition of F_{offset} , for carrier aggregation, which is “inherited” from the single carrier scenario:

$$F_{\text{offset}} = \frac{BW_{\text{channel}}}{2} \quad (\text{for base station}) \quad (19.5)$$

Unlike the base station, the terminal is not restricted by legacy operation, but rather from the nonlinear properties of the PA and the resulting unwanted emissions mask. At both edges of the aggregated channel bandwidth, a guard band BW_{GB} will be needed, in order for the emissions to reach a level where the out-of-band emissions limits in terms of an emission mask are applied. Whether a single wide carrier or multiple aggregated carriers of the same or different sizes are transmitted, the guard band needed will have to be the same at both edges, since the emission mask roll-off is the same. A problem with the backwards-compatible base station definition is that the resulting guard BW_{GB} is proportional to the channel BW and would therefore be *different* if carriers of different channel BW are aggregated.

For this reason, a different definition is used for the terminal, based on a “symmetrical” guard band. For the edge carriers (low and high), F_{offset} is half of the transmission bandwidth configuration, plus a symmetrical guard band BW_{GB} :

$$F_{\text{offset}} = \frac{0.18 \text{ MHz} \cdot N_{\text{RB}}}{2} + BW_{\text{GB}} \quad (\text{for terminal}) \quad (19.6)$$

where 0.18 MHz is the bandwidth of one resource block and BW_{GB} is proportional to the channel BW of the largest component carrier. For the CA bandwidth classes defined in release 11 and where the edge carriers have the same channel bandwidth, F_{offset} will be the same for terminals and base stations and $BW_{\text{Channel_CA}}$ will be the same.

It may look like an anomaly that the definitions may potentially lead to slightly different aggregated channel BW for the terminal and the base station, but this is in fact not a problem. Terminal and base-station requirements are defined separately and do not have to cover the same frequency ranges. The aggregated channel BW for both terminal and base station do, however, have to be within an operator’s license block in the operating band.

Once the frequency reference point is set, the actual RF requirements are to a large extent the same as for a single carrier configuration. Which requirements are affected is explained for each requirement in the discussion later in this chapter.

19.5 Operation in non-contiguous spectrum

Some spectrum allocations used for LTE deployments consist of fragmented parts of spectrum for different reasons. The spectrum may be recycled 2G spectrum, where the original

licensed spectrum was “interleaved” between operators. This was quite common for original GSM deployments for implementation reasons (the original combiner filters used were not easily tuned when spectrum allocations were expanded). In some regions, operators have also purchased spectrum licenses on auctions and have for different reasons ended up with multiple allocations in the same band that are not adjacent.

For deployment of non-contiguous spectrum allocations there are a couple of implications:

- If the full spectrum allocation in a band is to be operated with a single base station, the base station has to be capable of operation in non-contiguous spectrum.
- If a larger transmission bandwidth is to be used than what is available in each of the spectrum fragments, both the terminal and the base station have to be capable of *Intra-band non-contiguous carrier aggregation* in that band.

Note that the capability for the base station to operate in non-contiguous spectrum is not directly coupled to carrier aggregation as such. From an RF point-of-view, what will be required by the base stations is to receive and transmit carriers over an RF bandwidth that is split in two (or more) separate sub-blocks, with a sub-block gap in between as shown in Figure 19.6. The spectrum in the sub-block gap can be deployed by any other operator, which means that the RF requirements for the base station in the sub-block gap will be based on co-existence for un-coordinated operation. This has a few implications for some of the base station RF requirements within an operating band.

If the non-contiguous spectrum is operated with carrier aggregation, the RF requirements for the base station will be fundamentally the same as in general for the non-contiguous

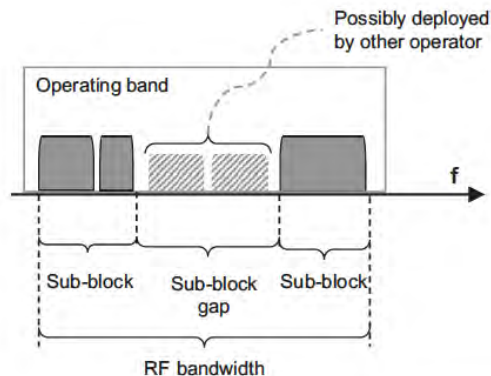
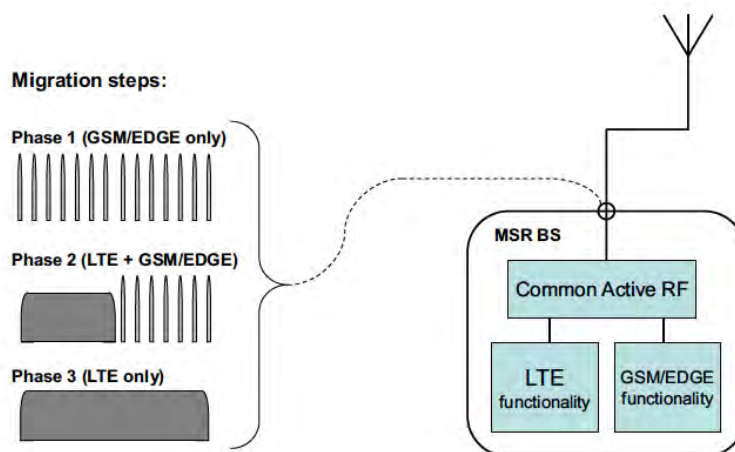


FIGURE 19.6

Example of non-contiguous spectrum operation, illustrating the definitions of *RF bandwidth*, *sub-block* and *sub-block gap*

**FIGURE 19.7**

Example of migration from GSM to LTE using an MSR base station for all migration phases

spectrum. For the terminal however, there are additional implications and limitations to handle the simultaneously received carriers.

19.6 Multi-standard radio base stations

Traditionally the RF specifications have been developed separately for the different 3GPP radio-access technologies GSM/EDGE, UTRA, and E-UTRA (LTE). The rapid evolution of mobile radio and the need to deploy new technologies alongside the legacy deployments has, however, lead to implementation of different Radio-Access Technologies (RAT) at the same sites, often sharing antennas and other parts of the installation. A natural further step is to also share the base-station equipment between multiple RATs. This requires multi-RAT base stations.

The evolution to multi-RAT base stations is also fostered by the evolution of technology. While multiple RATs have traditionally shared parts of the site installation, such as antennas, feeders, backhaul, or power, the advance of both digital baseband and RF technologies enables a much tighter integration. A base station consisting of two separate implementations of both baseband and RF, together with a passive combiner/splitter before the antenna, could in theory be considered a multi-RAT base station. 3GPP has, however, made a narrower but more forward-looking definition.

In a *Multi-Standard Radio* (MSR) base station, both the receiver and the transmitter are capable of simultaneously processing multiple carriers of different RATs in common *active*

RF components. The reason for this stricter definition is that the true potential of multi-RAT base stations, and the challenge in terms of implementation complexity, comes from having a common RF. This principle is illustrated in Figure 19.7 with an example base station capable of both GSM/EDGE and LTE. Much of the GSM/EDGE and LTE baseband functionality may be separate in the base station, but is possibly implemented in the same hardware. The RF must, however, be implemented in the same active components as shown in the figure.

The main advantages of an MSR base station implementation are twofold:

- Migration between RATs in a deployment, for example from GSM/EDGE to LTE, is possible using the same base station hardware. In the example in Figure 19.7, a migration is performed in three phases using the same MSR base station. In the first phase, the base station is deployed in a network for GSM/EDGE-only operation. In the second phase, the operator migrates part of the spectrum to LTE. The same MSR base station will now operate one LTE carrier, but still supports the legacy GSM/EDGE users in half of the band available. In the third phase, when the GSM/EDGE users have migrated from the band, the operator can configure the MSR base station to LTE-only operation with double the channel bandwidth.
- A single base station designed as an MSR base station can be deployed in various environments for single-RAT operation for each RAT supported, as well as for multi-RAT operation where that is required by the deployment scenario. This is also in line with the recent technology trends seen in the market, with fewer and more generic base-station designs. Having fewer varieties of base station is an advantage both for the base-station vendor and for the operator, since a single solution can be developed and implemented for a variety of scenarios.

The single-RAT 3GPP radio-access standards, with requirements defined independently per RAT, do not support such migration scenarios with an implementation where common base-station RF hardware is shared between multiple access technologies, and hence a separate set of requirements for multi-standard radio equipment is needed.

An implication of a common RF for multiple RATs is that carriers are no longer received and transmitted independently of each other. For this reason, a common RF specification must be used to specify the MSR base station. From 3GPP release 9 there is a set of MSR base station specifications for the core RF requirements [80] and for test requirements [81]. Those specifications support GSM/EDGE,³ UTRA and E-UTRA, and all combinations thereof. To support all possible RAT combinations, the MSR specifications have many generic requirements applicable regardless of RAT combination, together with single-access-technology-specific requirements to secure the integrity of the systems in single-RAT operation.

The MSR concept has a substantial impact for many requirements, while others remain completely unchanged. A fundamental concept introduced for MSR base stations is *RF*

³The MSR specifications are not applicable to single-RAT operation of GSM/EDGE.

bandwidth, which is defined as the total bandwidth over the set of carriers transmitted and received. Many receiver and transmitter requirements for GSM/EDGE and UTRA are specified relative to the carrier center and for LTE in relation to the channel edges. For an MSR base station, they are instead specified relative to the *RF bandwidth edges*, in a way similar to carrier aggregation in release 10. In the same way as for carrier aggregation, a parameter F_{offset} is also introduced to define the location of the RF bandwidth edges relative to the edge carriers. For GSM/EDGE carriers, F_{offset} is set to 200 kHz, while it is in general half the channel bandwidth for UTRA and E-UTRA. By introducing the RF bandwidth concept and introducing generic limits, the requirements for MSR shift from being carrier centric towards being frequency block centric, thereby embracing technology neutrality by being independent of the access technology or operational mode.

While E-UTRA and UTRA carriers have quite similar RF properties in terms of bandwidth and power spectral density, GSM/EDGE carriers are quite different. The operating bands for which MSR base stations are defined are therefore divided into three *Band Categories* (BC):

- BC1—All paired bands where UTRA FDD and E-UTRA FDD can be deployed.
- BC2—All paired bands where in addition to UTRA FDD and E-UTRA FDD, GSM/EDGE can also be deployed.
- BC3—All unpaired bands where UTRA TDD and E-UTRA TDD can be deployed.

Since the carriers of different RATs are not transmitted and received independently, it is necessary to perform parts of the testing with carriers of multiple RATs being activated. This is done through a set of multi-RAT *Test Configurations* defined in [81], specifically tailored to stress transmitter and receiver properties. These test configurations are of particular importance for the unwanted emission requirements for the transmitter and for testing of the receiver susceptibility to interfering signals (blocking, etc.). An advantage of the multi-RAT test configurations is that the RF performance of multiple RATs can be tested simultaneously, thereby avoiding repetition of test cases for each RAT. This is of particular importance for the very time-consuming tests of requirements over the complete frequency range outside the operating band.

The requirement with the largest impact from MSR is the spectrum mask, or the *operating band unwanted emissions* requirement. The spectrum mask requirement for MSR base stations is applicable for multi-RAT operation where the carriers at the RF bandwidth edges are either GSM/EDGE, UTRA, or E-UTRA carriers of different channel bandwidths. The mask is generic and applicable to all cases and covers the complete operating band of the base station. There is an exception for the 150 kHz closest to the RF bandwidth edge, where the mask is aligned with the GSM/EDGE modulation spectrum for the case when a GSM/EDGE carrier or a 1.4/3 MHz E-UTRA carrier is transmitted adjacent to the edge.

An important aspect of MSR is the declaration by the base station vendor of the supported RF bandwidth, power levels, multicarrier capability, etc. All testing is based on the capability of the base station through a declaration of the supported *Capability Set* (CS), which defines

Table 19.5 Capability Sets (CSx) Defined for MSR Base Stations and the Corresponding RAT Configurations		
Capability Set CSx Supported by a Base Station	Applicable Band Categories	Supported RAT Configurations
CS1	BC1, BC2 or BC3	Single-RAT: UTRA
CS2	BC1, BC2 or BC3	Single-RAT: UTRA
CS3	BC1, BC2 or BC3	Single-RAT: UTRA or E-UTRA Multi-RAT: UTRA + E-UTRA
CS4	BC2	Single-RAT: GSM or UTRA Multi-RAT: GSM + UTRA
CS5	BC2	Single-RAT: GSM or E-UTRA Multi-RAT: GSM + E-UTRA
CS6	BC2	Single-RAT: GSM, UTRA or E-UTRA Multi-RAT: GSM + UTRA, GSM + E-UTRA, UTRA + E-UTRA or GSM + UTRA + E-UTRA

all supported single RATs and multi-RAT combinations. There are currently six capability sets CS1 to CS6 defined in the MSR test specification [81], allowing full flexibility for implementing and deploying base stations compliant to the MSR specification. These capability sets are listed in Table 19.5 together with the band categories where the capability set is applicable and the RAT configurations that are supported by the BS. Note the difference between the capability of a base station (as declared by the manufacturer) and the configuration in which a BS is operating. CS1 and CS2 define capabilities for base stations that are only single-RAT capable and make it possible to apply the MSR base station specification for such base stations, instead of the corresponding single-RAT UTRA or E-UTRA specifications. There is no capability set defined for base stations that are *only* single-RAT GSM capable, since that is a type of BS that is solely covered by the single-RAT GSM/EDGE specifications.

For a large part of the base station RF requirements, multi-RAT testing is not necessary and the actual test limits are unchanged for the MSR base station. In these cases, both the requirements and the test cases are simply incorporated through direct references to the corresponding single-RAT specifications.

Carrier aggregation as described in Section 19.4 is also applicable to MSR base stations. Since the MSR specification has most of the concepts and definitions in place for defining multi-carrier RF requirements, whether aggregated or not, the differences for the MSR requirements compared to non-aggregated carriers are very minor.

19.7 Overview of RF requirements for LTE

The RF requirements define the receiver and transmitter RF characteristics of a base station or terminal. The base station is the physical node that transmits and receives RF signals on one or more antenna connectors. Note that a base station is not the same thing as an eNodeB, which is the corresponding logical node in the LTE Radio-Access Network. The terminal is denoted UE in all RF specifications.

The set of RF requirements defined for LTE is fundamentally the same as those defined for UTRA or any other radio system. Some requirements are also based on regulatory requirements and are more concerned with the frequency band of operation and/or the place where the system is deployed, than with the type of system.

What is particular to LTE is the flexible bandwidth and the related multiple channel bandwidths of the system, which makes some requirements more complex to define. These properties have special implications for the transmitter requirements on unwanted emissions, where the definition of the limits in international regulation depends on the channel bandwidth. Such limits are harder to define for a system where the base station may operate with multiple channel bandwidths and where the terminal may vary its channel bandwidth of operation. The properties of the flexible OFDM-based physical layer also have implications for specifying the transmitter modulation quality and how to define the receiver selectivity and blocking requirements.

The type of transmitter requirements defined for the terminal is very similar to what is defined for the base station, and the definitions of the requirements are often similar. The output power levels are, however, considerably lower for a terminal, while the restrictions on the terminal implementation are much higher. There is tight pressure on cost and complexity for all telecommunications equipment, but this is much more pronounced for terminals, due to the scale of the total market being close to *two billion* devices per year. In cases where there are differences in how requirements are defined between terminal and base station, they are treated separately in this chapter.

The detailed background of the RF requirements for LTE is described in [82,83], with further details of the additional requirements in release 10 (for LTE-Advanced) in [84,85]. The RF requirements for the base station are specified in [86] and for the terminal in [76]. The RF requirements are divided into transmitter and receiver characteristics. There are also *performance characteristics* for base station and terminal that define the receiver baseband performance for all physical channels under different propagation conditions. These are not strictly RF requirements, though the performance will also depend on the RF to some extent.

Each RF requirement has a corresponding test defined in the LTE test specifications for the base station [87] and the terminal [74]. These specifications define the test setup, test procedure, test signals, test tolerances, etc. needed to show compliance with the RF and performance requirements.

19.7.1 Transmitter characteristics

The transmitter characteristics define RF requirements for the wanted signal transmitted from the terminal and base station, but also for the unavoidable unwanted emissions outside the transmitted carrier(s). The requirements are fundamentally specified in three parts:

- **Output power level** requirements set limits for the maximum allowed transmitted power, for the dynamic variation of the power level, and in some cases for the transmitter OFF state.
- **Transmitted signal quality** requirements define the “purity” of the transmitted signal and also the relation between multiple transmitter branches.
- **Unwanted emissions** requirements set limits to all emissions outside the transmitted carrier(s) and are tightly coupled to regulatory requirements and coexistence with other systems.

A list of the terminal and base-station transmitter characteristics arranged according to the three parts defined above is shown in Table 19.6. A more detailed description of the requirements can be found later in this chapter.

Table 19.6 Overview of LTE Transmitter Characteristics		
	Base Station Requirement	Terminal Requirement
Output power level	Maximum output power Output power dynamics ON/OFF power (TDD only)	Transmit power Output power dynamics Power control
Transmitted signal quality	Frequency error Error Vector Magnitude (EVM) Time alignment between transmitter branches	Frequency error Transmit modulation quality In-band emissions
Unwanted emissions	Operating band unwanted emissions Adjacent Channel Leakage Ratio (ACLR and CACLR) Spurious emissions Occupied bandwidth Transmitter intermodulation	Spectrum emission mask Adjacent Channel Leakage Ratio (ACLR and CACLR) Spurious emissions Occupied bandwidth Transmit intermodulation

19.7.2 Receiver characteristics

The set of receiver requirements for LTE is quite similar to what is defined for other systems such as UTRA, but many of them are defined differently, due to the flexible bandwidth properties. The receiver characteristics are fundamentally specified in three parts:

- **Sensitivity and dynamic range** requirements for receiving the wanted signal.
- **Receiver susceptibility to interfering signals** defines receivers' susceptibility to different types of interfering signals at different frequency offsets.
- **Unwanted emissions** limits are also defined for the receiver.

A list of the terminal and base-station receiver characteristics arranged according to the three parts defined above is shown in Table 19.7. A more detailed description of each requirement can be found later in this chapter.

19.7.3 Regional requirements

There are a number of regional variations to the RF requirements and their application. The variations originate in different regional and local regulations of spectrum and its use. The most obvious regional variation is the different frequency bands and their use, as discussed above. Many of the regional RF requirements are also tied to specific frequency bands.

When there is a regional requirement on, for example, spurious emissions, this requirement should be reflected in the 3GPP specifications. For the base station it is entered as an

Table 19.7 Overview of LTE Receiver Characteristics

	Base Station Requirement	Terminal Requirement
Sensitivity and dynamic range	Reference sensitivity	Reference sensitivity power level
	Dynamic range	Maximum input level
	In-channel selectivity	
Receiver susceptibility to interfering signals	Out-of-band blocking	Out-of-band blocking
		Spurious response
	In-band blocking	In-band blocking
	Narrowband blocking	Narrowband blocking
	Adjacent channel selectivity	Adjacent Channel Selectivity
	Receiver intermodulation	Intermodulation characteristics
Unwanted emissions from the receiver	Receiver spurious emissions	Receiver spurious emissions

optional requirement and is marked as “regional.” For the terminal, the same procedure is not possible, since a terminal may roam between different regions and will therefore have to fulfill all regional requirements that are tied to an operating band in the regions where the band is used. For LTE, this becomes more complex than for UTRA, since there is an additional variation in the transmitter (and receiver) bandwidth used, making some regional requirements difficult to meet as a mandatory requirement. The concept of *network signaling* of RF requirements is therefore introduced for LTE, where a terminal can be informed at call setup of whether some specific RF requirements apply when the terminal is connected to a network.

19.7.4 Band-specific terminal requirements through network signaling

For the terminal, the channel bandwidths supported are a function of the LTE operating band, and also have a relation to the transmitter and receiver RF requirements. The reason is that some RF requirements may be difficult to meet under conditions with a combination of maximum power and high number of transmitted and/or received resource blocks.

Some additional RF requirements apply for the terminal when a specific Network Signaling Value (NS_x) is signaled to the terminal as part of the cell handover or broadcast message. For implementation reasons, these requirements are associated with restrictions and variations to RF parameters such as terminal output power, maximum channel bandwidth, and number of transmitted resource blocks. The variations of the requirements are defined together with the Network Signaling Value (NS_x) in the terminal RF specification [76], where each value corresponds to a specific condition. The default value for all bands is NS_01. All NS_x values are connected to an allowed power reduction called *Additional Maximum Power Reduction* (A-MPR) and apply for transmission using a certain minimum number of resource blocks, depending also on the channel bandwidth. The following are examples of terminal requirements that have a related Network Signaling Value for some bands:

- **NS_03, NS_04, or NS_06** is signaled when specific FCC requirements [88] on terminal unwanted emissions apply for operation in a number of US bands.
- **NS_05** is signaled for protection of the PHS band in Japan when a terminal operates in the 2 GHz band (Band 1).

In some bands the NS_x signaling is also applied for testing of receiver sensitivity, since the active transmitted signal can affect the receiver performance.

There are also additional RF requirements and restrictions that may apply in case of LTE carrier aggregation in the uplink. These can be signaled to a terminal configured for carrier aggregation using specific CA network signaling values CA_NS_x and will in this case replace the usual network signaling values NS_x and their related requirements.

19.7.5 Base-station classes

In the base-station specifications, there is one set of RF requirements that is generic, applicable to what is called “general purpose” base stations. This is the original set of LTE requirements developed in 3GPP release 8. It has no restrictions on base-station output power and can be used for any deployment scenario. When the RF requirements were derived, however, the scenarios used were macro scenarios [89]. For this reason, in release 9 additional base-station classes were introduced that were intended for pico-cell and femto-cell scenarios. An additional class for micro-cell scenarios was added in release 11, together with base station classes applicable for multi-standard base stations. It is also clarified that the original set of “general purpose” RF parameters are applicable for macro-cell scenarios. The terms macro, micro, pico, and femto are not used in 3GPP to identify the base-station classes; instead the following terminology is used:

- **Wide Area base stations.** This type of base station is intended for macro-cell scenarios, defined with a minimum coupling loss between base station and terminal of 70 dB.
- **Medium Range base stations.** This type of base station is intended for micro-cell scenarios, defined with a minimum coupling loss between base station and terminal of 53 dB. Typical deployments are outdoor below-rooftop installations, giving both outdoor hotspot coverage and outdoor-to-indoor coverage through walls.
- **Local Area base stations.** This type of base station is intended for pico-cell scenarios, defined with a minimum coupling loss between base station and terminal of 45 dB. Typical deployments are indoor offices and indoor/outdoor hotspots, with the BS mounted on walls or ceilings.
- **Home base stations.** This type of base station is intended for femto-cell scenarios, which are not explicitly defined. Minimum coupling loss between base station and terminal of 45 dB is also assumed here. Home base stations can be used both for open access and in closed subscriber groups.

The Local Area, Medium Range and Home base station classes have modifications to a number of requirements compared to Wide Area base stations, mainly due to the assumption of a lower minimum coupling loss:

- Maximum base station power is limited to 38 dBm output power for Medium Range base stations, 24 dBm output power for Local Area base stations, and to 20 dBm for Home base stations. This power is defined per antenna and carrier, except for Home base stations, where the power over all antennas (up to four) is counted. There is no maximum base station power defined for Wide Area base stations.
- Home base stations have an additional requirement for protecting systems operating on adjacent channels. The reason is that a terminal connected to a base station belonging to another operator on the adjacent channel may be in close proximity to the Home base station. To avoid an interference situation where the adjacent terminal is blocked, the Home base station must make measurements on the adjacent channel to detect adjacent

base-station operations. If an adjacent base-station transmission (UTRA or LTE) is detected under certain conditions, the maximum allowed Home base-station output power is reduced in proportion to how weak the adjacent base-station signal is, in order to avoid interference to the adjacent base station.

- The spectrum mask (operating band unwanted emissions) has lower limits for Medium Range, Local Area, and Home base stations, in line with the lower maximum power levels.
- Unwanted emission limits for protecting Home base-station operation (from other Home base stations) are lower, since a stricter through-the-wall indoor interference scenario is assumed. Limits for co-location for Medium Range and Local Area are, however, less strict, corresponding to the relaxed reference sensitivity for the base station.
- Receiver reference sensitivity limits are higher (more relaxed) for Medium Range, Local Area and Home base stations. Receiver dynamic range and in-channel selectivity are also adjusted accordingly.
- All Medium Range, Local Area, and Home base-station limits for receiver susceptibility to interfering signals are adjusted to take the higher receiver sensitivity limit and the lower assumed minimum coupling loss (base station-to-terminal) into account.

19.8 Output power level requirements

19.8.1 Base-station output power and dynamic range

There is no general maximum output power requirement for base stations. As mentioned in the discussion of base-station classes above, there is, however, a maximum output power limit of 38 dBm for Medium Range base stations, 24 dBm for Local Area base stations, and 20 dBm for Home base stations. In addition to this there is a tolerance specified, defining how much the actual maximum power may deviate from the power level declared by the manufacturer.

The base station also has a specification of the total power control dynamic range for a resource element, defining the power range over which it should be possible to configure. There is also a dynamic range requirement for the total base-station power.

For TDD operation, a power mask is defined for the base-station output power, defining the Off power level during the uplink subframes and the maximum time for the *transmitter transient period* between the transmitter On and Off states.

19.8.2 Terminal output power and dynamic range

The terminal output power level is defined in three steps:

- *UE power class* defines a *nominal* maximum output power for QPSK modulation. It may be different in different operating bands, but the main terminal power class is today set at 23 dBm for all bands.

- *Maximum Power Reduction (MPR)* defines an allowed reduction of maximum power level for certain combinations of modulation used and the number of resource blocks that are assigned.
- *Additional Maximum Power Reduction (A-MPR)* may be applied in some regions and is usually connected to specific transmitter requirements such as regional emission limits and to certain carrier configurations. For each such set of requirements there is an associated network signaling value NS_x that identifies the allowed A-MPR and the associated conditions, as explained in [section 19.7.4](#).

The terminal has a definition of the transmitter Off power level, applicable to conditions when the terminal is not allowed to transmit. There is also a general On/Off time mask specified, plus specific time masks for PRACH, SRS, subframe boundary, and PUCCH/PUSCH/SRS.

The terminal transmit power control is specified through requirements for the *absolute power tolerance* for the initial power setting, the *relative power tolerance* between two subframes, and the *aggregated power tolerance* for a sequence of power-control commands.

19.9 Transmitted signal quality

The requirements for transmitted signal quality specify how much the transmitted base station or terminal signal deviates from an “ideal” modulated signal in the signal and the frequency domains. Impairments on the transmitted signal are introduced by the transmitter RF parts, with the nonlinear properties of the power amplifier being a major contributor. The signal quality is assessed for base station and terminal through requirements on *EVM* and *frequency error*. An additional terminal requirement is UE in-band emissions.

19.9.1 EVM and frequency error

While the theoretical definitions of the signal quality measures are quite straightforward, the actual assessment is a very elaborate procedure, described in great detail in the 3GPP specification. The reason is that it becomes a multidimensional optimization problem, where the best match for the timing, the frequency, and the signal constellation is found.

The *Error Vector Magnitude (EVM)* is a measure of the error in the modulated signal constellation, taken as the root mean square of the error vectors over the active subcarriers, considering all symbols of the modulation scheme. It is expressed as a percentage value in relation to the power of the ideal signal. The EVM fundamentally defines the maximum SINR that can be achieved at the receiver, if there are no additional impairments to the signal between transmitter and receiver.

Since a receiver can remove some impairments of the transmitted signal, such as time dispersion, the EVM is assessed after cyclic prefix removal and equalization. In this way, the

EVM evaluation includes a standardized model of the receiver. The frequency offset resulting from the EVM evaluation is averaged and used as a measure of the *frequency error* of the transmitted signal.

19.9.2 Terminal in-band emissions

In-band emissions are emissions within the channel bandwidth. The requirement limits how much a terminal can transmit into non-allocated resource blocks within the channel bandwidth. Unlike the out-of-band emissions, the in-band emissions are measured after cyclic prefix removal and FFT, since this is how a terminal transmitter affects a real base-station receiver.

19.9.3 Base-station time alignment

Several LTE features require the base station to transmit from two or more antennas, such as transmitter diversity and MIMO. For carrier aggregation, the carriers may also be transmitted from different antennas. In order for the terminal to properly receive the signals from multiple antennas, the timing relation between any two transmitter branches is specified in terms of a maximum time alignment error between transmitter branches. The maximum allowed error depends on the feature or combination of features in the transmitter branches.

19.10 Unwanted emissions requirements

Unwanted emissions from the transmitter are divided into *out-of-band (OOB) emissions* and *spurious emissions* in ITU-R recommendations [90]. OOB emissions are defined as emissions on a frequency close to the RF carrier, which results from the modulation process. Spurious emissions are emissions outside the RF carrier that may be reduced without affecting the corresponding transmission of information. Examples of spurious emissions are harmonic emissions, intermodulation products, and frequency conversion products. The frequency range where OOB emissions are normally defined is called the *OOB domain*, whereas spurious emissions limits are normally defined in the *spurious domain*.

ITU-R also defines the boundary between the OOB and spurious domains at a frequency separation from the carrier center of 2.5 times the necessary bandwidth, which corresponds to 2.5 times the channel bandwidth for LTE. This division of the requirements is easily applied for systems that have a fixed channel bandwidth. It does, however, become more difficult for LTE, which is a flexible bandwidth system, implying that the frequency range where requirements apply would then vary with the channel bandwidth. The approach taken for defining the boundary in 3GPP is slightly different for base-station and terminal requirements.

With the recommended boundary between OOB emissions and spurious emissions set at 2.5 times the channel bandwidth, third- and fifth-order intermodulation products from the carrier will fall inside the OOB domain, which will cover a frequency range of twice the

channel bandwidth on each side of the carrier. For the OOB domain, two overlapping requirements are defined for both base station and terminal: *Spectrum Emission Mask* (SEM) and *Adjacent Channel Leakage Ratio* (ACLR). The details of these are further explained ahead.

19.10.1 Implementation aspects

As shown in Chapter 3, the spectrum of an OFDM signal decays rather slowly outside of the transmission bandwidth configuration. Since the transmitted signal for LTE occupies 90% of the channel bandwidth, it is not possible to directly meet the unwanted emission limits directly outside the channel bandwidth with a “pure” OFDM signal. The techniques used for achieving the transmitter requirements are, however, not specified or mandated in LTE specifications. Time-domain windowing is one method commonly used in OFDM-based transmission systems to control spectrum emissions. Filtering is always used—both time-domain digital filtering of the baseband signal and analog filtering of the RF signal.

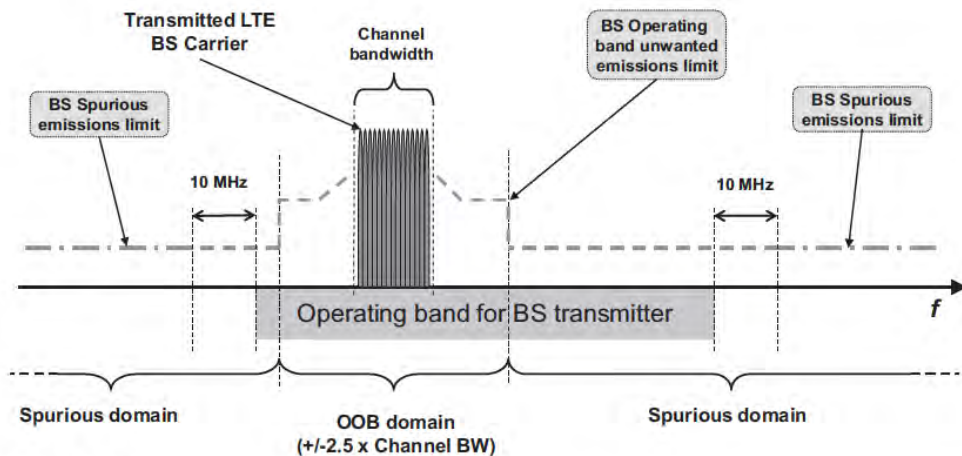
The nonlinear characteristics of the *Power Amplifier* (PA) used to amplify the RF signal must also be taken into account, since it is the source of intermodulation products outside the channel bandwidth. Power back-off to give a more linear operation of the PA can be used, but at the cost of a lower power efficiency. The power back-off should therefore be kept to a minimum. For this reason, additional linearization schemes can be employed. These are especially important for the base station, where there are fewer restrictions on implementation complexity and use of advanced linearization schemes is an essential part of controlling spectrum emissions. Examples of such techniques are feed-forward, feedback, predistortion, and postdistortion.

19.10.2 Spectrum emission mask

The spectrum emission mask defines the permissible out-of-band spectrum emissions outside the necessary bandwidth. As explained above, how to take the flexible channel bandwidth into account when defining the frequency boundary between OOB emissions and spurious emissions is done differently for the LTE base station and terminal. Consequently, the spectrum emission masks are also based on different principles.

19.10.2.1 Base-station operating band unwanted emission limits

For the LTE base station, the problem of the implicit variation of the boundary between OOB and spurious domain with the varying channel bandwidths is handled by not defining an explicit boundary. The solution is a unified concept of *operating band unwanted emissions* (UEM) for the LTE base station instead of the spectrum mask usually defined for OOB emissions. The operating band unwanted emissions requirement applies over the whole base station transmitter operating band, plus an additional 10 MHz on each side, as shown in [Figure 19.8](#). All requirements outside of that range are set by the regulatory spurious emission limits, based on the ITU-R recommendations [90]. As seen in the figure, a large part of the operating band unwanted emissions are defined over a frequency range that for smaller

**FIGURE 19.8**

Frequency ranges for Operating band unwanted emissions and Spurious emissions applicable to LTE base station

channel bandwidths can be both in spurious and OOB domains. This means that the limits for the frequency ranges that may be in the spurious domain also have to align with the regulatory limits from the ITU-R. The shape of the mask is generic for all channel bandwidths from 5 to 20 MHz, with a mask that consequently has to align with the ITU-R limits starting 10 MHz from the channel edges. Special masks are defined for the smaller 1.4 and 3 MHz channel bandwidths. The operating band unwanted emissions are defined with a 100 kHz measurement bandwidth.

In case of carrier aggregation for a base station, the UEM requirements (as other RF requirements) apply as for any multicarrier transmission, where the UEM will be defined relative to the carriers on the edges of the RF bandwidth. In case of non-contiguous carrier aggregation, the UEM within a sub-block gap is partly calculated as the cumulative sum of contributions from each sub-block.

There are also special limits defined to meet a specific regulation set by the FCC [88] for the operating bands used in the USA and by the ECC for some European bands. These are specified as separate limits in addition to the operating band unwanted emission limits.

19.10.2.2 Terminal spectrum emission mask

For implementation reasons, it is not possible to define a generic terminal spectrum mask that does not vary with the channel bandwidth, so the frequency ranges for OOB limits and spurious emissions limits do not follow the same principle as for the base station. The SEM extends out to a separation Δf_{OOB} from the channel edges, as illustrated in Figure 19.9. For 5

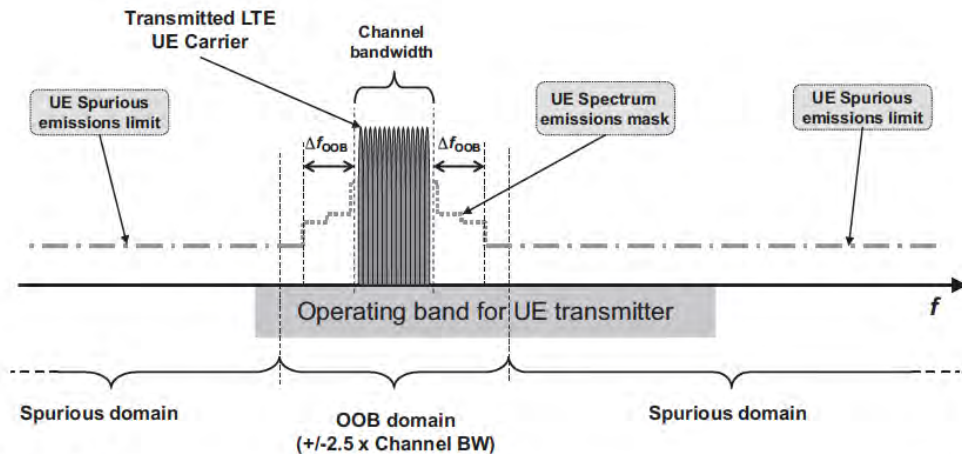


FIGURE 19.9

Frequency ranges for Spectrum emissions mask and Spurious emissions applicable to LTE terminal

MHz channel bandwidth, this point corresponds to 250% of the necessary bandwidth as recommended by the ITU-R, but for higher channel bandwidths it is set closer than 250%.

The SEM is defined as a general mask and a set of additional masks that can be applied to reflect different regional requirements. Each additional regional mask is associated with a specific network signaling value NS_x.

19.10.3 Adjacent channel leakage ratio

In addition to a spectrum emissions mask, the OOB emissions are defined by an *Adjacent Channel Leakage Ratio* (ACLR) requirement. The ACLR concept is very useful for analysis of coexistence between two systems that operate on adjacent frequencies. The ACLR defines the ratio of the power transmitted within the assigned channel bandwidth to the power of the unwanted emissions transmitted on an adjacent channel. There is a corresponding receiver requirement called *Adjacent Channel Selectivity* (ACS), which defines a receiver's ability to suppress a signal on an adjacent channel.

The definitions of ACLR and ACS are illustrated in Figure 19.10 for a wanted and an interfering signal received in adjacent channels. The interfering signal's leakage of unwanted emissions at the wanted signal receiver is given by the ACLR and the ability of the receiver of the wanted signal to suppress the interfering signal in the adjacent channel is defined by the ACS. The two parameters when combined define the total leakage between two transmissions on adjacent channels. That ratio is called the *Adjacent Channel Interference Ratio* (ACIR) and is defined as the ratio of the power transmitted on one channel to the total interference

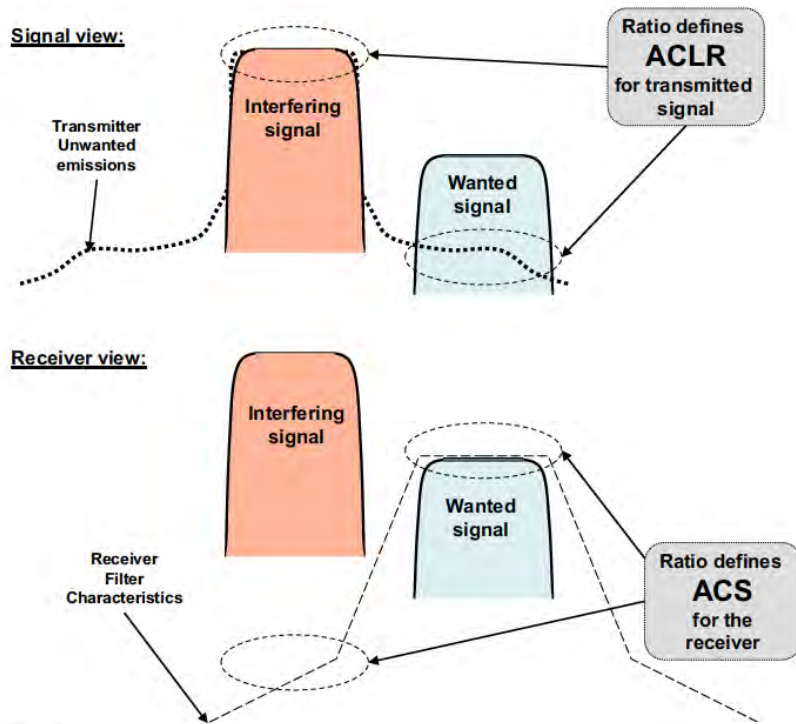
**FIGURE 19.10**

Illustration of ACLR and ACS, with example characteristics for an “aggressor” interferer and a receiver for a “victim” wanted signal

received by a receiver on the adjacent channel, due to both transmitter (ACLR) and receiver (ACS) imperfections.

This relation between the adjacent channel parameters is [91]:

$$\text{ACIR} = \frac{1}{\frac{1}{\text{ACLR}} + \frac{1}{\text{ACS}}} \quad (19.7)$$

ACLR and ACS can be defined with different channel bandwidths for the two adjacent channels, which is the case for some requirements set for LTE due to the bandwidth flexibility. The equation above will also apply for different channel bandwidths, but only if the same two channel bandwidths are used for defining all three parameters ACIR, ACLR, and ACS used in the equation.

The ACLR limits for LTE terminal and base station are derived based on extensive analysis [89] of LTE coexistence with LTE or other systems on adjacent carriers.

The LTE base-station requirements on ACLR and operating band unwanted emissions both cover the OOB domain, but the operating band unwanted emission limits are in general set slightly more relaxed compared to the ACLR, since they are defined in a much narrower measurement bandwidth of 100 kHz. This allows for some variations in the unwanted emissions due to intermodulation products from varying power allocation between resource blocks within the channel. For an LTE base station, there are ACLR requirements both for an adjacent channel with a UTRA receiver and with an LTE receiver of the same channel bandwidth.

In case of carrier aggregation for a base station, the ACLR (as other RF requirements) apply as for any multicarrier transmission, where the ACLR requirement will be defined for the carriers on the edges of the RF bandwidth. In case of non-contiguous carrier aggregation where the sub-block gap is so small that the ACLR requirements at the edges of the gap will “overlap,” a special *Cumulative ACLR* requirement (CACLR) is defined for the gap. For CACLR, contributions from carriers on both sides of the sub-block gap are accounted for in the CACLR limit.

ACLR limits for the terminal are set both with assumed UTRA and LTE receivers on the adjacent channel. As for the base station, the limits are also set stricter than the corresponding SEM, thereby accounting for variations in the spectrum emissions resulting from variations in resource-block allocations. In case of carrier aggregation, the terminal ACLR requirement applies to the aggregated channel bandwidth instead of per carrier.

19.10.4 Spurious emissions

The limits for base station spurious emissions are taken from international recommendations [90], but are only defined in the region outside the frequency range of operating band unwanted emissions limits as illustrated in Figure 19.8—that is, at frequencies that are separated from the base-station transmitter operating band by at least 10 MHz. There are also additional regional or optional limits for protection of other systems that LTE may coexist with or even be co-located with. Examples of other systems considered in those additional spurious emissions requirements are GSM, UTRA FDD/TDD, CDMA2000, and PHS.

Terminal spurious emission limits are defined for all frequency ranges outside the frequency range covered by the SEM. The limits are generally based on international regulations [90], but there are also additional requirements for coexistence with other bands when the mobile is roaming. The additional spurious emission limits can have an associated network signaling value.

In addition, there are base-station and terminal emission limits defined for the receiver. Since receiver emissions are dominated by the transmitted signal, the receiver spurious emission limits are only applicable when the transmitter is Off, and also when the transmitter is On for an LTE FDD base station that has a separate receiver antenna connector.

19.10.5 Occupied bandwidth

Occupied bandwidth is a regulatory requirement that is specified for equipment in some regions, such as Japan and the USA. It is originally defined by the ITU-R as a maximum bandwidth, outside of which emissions do not exceed a certain percentage of the total emissions. The occupied bandwidth is for LTE equal to the channel bandwidth, outside of which a maximum of 1% of the emissions are allowed (0.5% on each side).

In the case of carrier aggregation, the occupied bandwidth is equal to the aggregated channel bandwidth. For non-contiguous carrier aggregation, the occupied bandwidth applies per sub-block.

19.10.6 Transmitter intermodulation

An additional implementation aspect of an RF transmitter is the possibility of intermodulation between the transmitted signal and another strong signal transmitted in the proximity of the base station or terminal. For this reason there is a requirement for *transmitter intermodulation*.

For the base station, the requirement is based on a stationary scenario with a co-located other base-station transmitter, with its transmitted signal appearing at the antenna connector of the base station being specified, but attenuated by 30 dB. Since it is a stationary scenario, there are no additional unwanted emissions allowed, implying that all unwanted emission limits also have to be met with the interferer present.

For the terminal, there is a similar requirement based on a scenario with another terminal transmitted signal appearing at the antenna connector of the terminal being specified, but attenuated by 40 dB. The requirement specifies the minimum attenuation of the resulting intermodulation product below the transmitted signal.

19.11 Sensitivity and dynamic range

The primary purpose of the *reference sensitivity requirement* is to verify the receiver *Noise Figure*, which is a measure of how much the receiver's RF signal chain degrades the SNR of the received signal. For this reason, a low-SNR transmission scheme using QPSK is chosen as reference channel for the reference sensitivity test. The reference sensitivity is defined at a receiver input level where the throughput is 95% of the maximum throughput for the reference channel.

For the base station, reference sensitivity could potentially be defined for a single resource block up to a group covering all resource blocks. For reasons of complexity, a maximum granularity of 25 resource blocks has been chosen. This means that for channel bandwidths larger than 5 MHz, sensitivity is verified over multiple adjacent 5 MHz blocks, while it is only defined over the full channel for smaller channel bandwidths.

For the terminal, reference sensitivity is defined for the full channel bandwidth signals and with all resource blocks allocated for the wanted signal. For the higher channel bandwidths (>5

MHz) in some operating bands, the nominal reference sensitivity needs to be met with a minimum number of allocated resource blocks. For larger allocation, a certain relaxation is allowed.

The intention of the *dynamic range requirement* is to ensure that the receiver can also operate at received signal levels considerably higher than the reference sensitivity. The scenario assumed for base-station dynamic range is the presence of increased interference and corresponding higher wanted signal levels, thereby testing the effects of different receiver impairments. In order to stress the receiver, a higher SNR transmission scheme using 16QAM is applied for the test. In order to further stress the receiver to higher signal levels, an interfering AWGN signal at a level 20 dB above the assumed noise floor is added to the received signal. The dynamic range requirement for the terminal is specified as a *maximum signal level* at which the throughput requirement is met.

19.12 Receiver susceptibility to interfering signals

There is a set of requirements for base station and terminal, defining the receiver's ability to receive a wanted signal in the presence of a stronger interfering signal. The reason for the multiple requirements is that, depending on the frequency offset of the interferer from the wanted signal, the interference scenario may look very different and different types of receiver impairments will affect the performance. The intention of the different combinations of interfering signals is to model as far as possible the range of scenarios with interfering signals of different bandwidths that may be encountered inside and outside the base-station and terminal receiver operating band.

While the types of requirements are very similar between base station and terminal, the signal levels are different, since the interference scenarios for the base station and terminal are very different. There is also no terminal requirement corresponding to the base-station *In-Channel Selectivity* (ICS) requirement.

The following requirements are defined for LTE base station and terminal, starting from interferers with large frequency separation and going close in (see Figure 19.11). In all cases where the interfering signal is an LTE signal it has the same bandwidth as the wanted signal, but at most 5 MHz.

- *Blocking*. This corresponds to the scenario with strong interfering signals received outside the operating band (out-of-band blocking) or inside the operating band (in-band blocking), but not adjacent to the wanted signal. In-band blocking includes interferers in the first 20 MHz outside the operating band for the base station and the first 15 MHz for the terminal. The scenarios are modeled with a *Continuous Wave* (CW) signal for the out-of-band case and an LTE signal for the in-band case. There are additional (optional) base-station blocking requirements for the scenario when the base station is co-located with another base station in a different operating band. For the terminal, a fixed number of *exceptions* are allowed from the out-of-band blocking requirement for each assigned

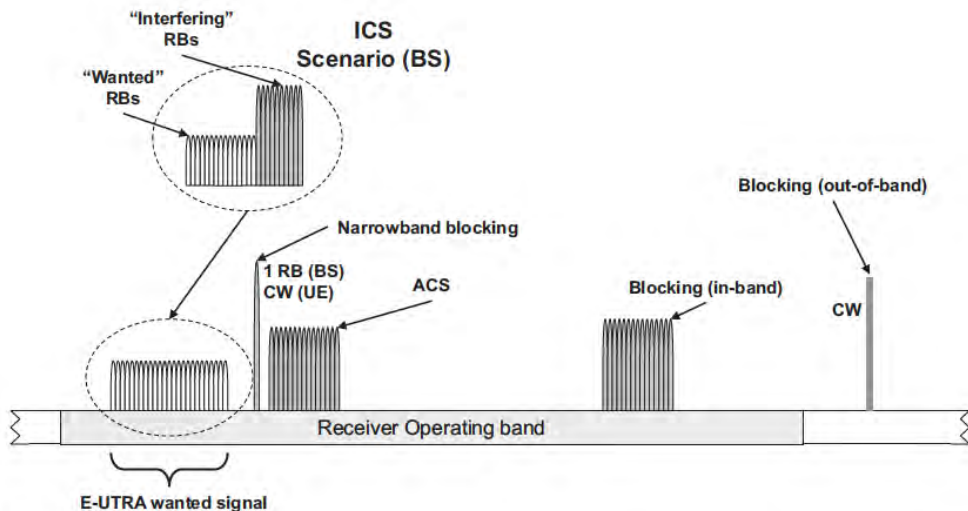


FIGURE 19.11

Base-station and terminal requirements for receiver susceptibility to interfering signals in terms of blocking, ACS, narrowband blocking, and in-channel selectivity (BS only)

frequency channel and at the respective *spurious response frequencies*. At those frequencies, the terminal must comply with the more relaxed spurious response requirement.

- **Adjacent channel selectivity.** The ACS scenario is a strong signal in the channel adjacent to the wanted signal and is closely related to the corresponding ACLR requirement (see the discussion in [Section 19.10.3](#)). The adjacent interferer is an LTE signal. For the terminal, the ACS is specified for two cases with a lower and a higher signal level. For MSR base stations, there is no specific ACS requirement defined. It is instead replaced by the narrowband blocking requirement, which covers the adjacent channel properties fully.
- **Narrowband blocking.** The scenario is an adjacent strong narrowband interferer, which in the requirement is modeled as a single resource block LTE signal for the base station and a CW signal for the terminal.
- **In-channel selectivity (ICS).** The scenario is multiple received signals of different received power levels inside the channel bandwidth, where the performance of the weaker “wanted” signal is verified in the presence of the stronger “interfering” signal. ICS is only specified for the base station.
- **Receiver intermodulation.** The scenario is *two* interfering signals near to the wanted signal, where the interferers are one CW and one LTE signal (not shown in [Figure 19.11](#)).

The purpose of the requirement is to test receiver linearity. The interferers are placed in frequency in such a way that the main intermodulation product falls inside the wanted signal's channel bandwidth. There is also a *narrowband intermodulation* requirement for the base station where the CW signal is very close to the wanted signal and the LTE interferer is a single RB signal.

For all requirements except in-channel selectivity, the wanted signal uses the same reference channel as in the corresponding reference sensitivity requirement. With the interference added, the same 95% relative throughput is met as for the reference sensitivity but at a “desensitized” higher wanted signal level.

19.13 Multi-band-capable base stations

The 3GPP specifications have been continuously developed to support larger RF bandwidths for transmission and reception through multicarrier and multi-RAT operation and carrier aggregation over contiguous and non-contiguous spectrum allocations. This has been made possible with the evolution of RF technology supporting larger bandwidths for both transmitters and receivers. The next step in RF technology for base stations is to support simultaneous transmission and/or reception in multiple bands through a common radio. A multi-band base station could cover multiple bands over a frequency range of a few hundred MHz.

The RF requirements for such multi-band-capable base stations are defined in 3GPP release 11. The specification supports multi-band operation both with a single RAT and with multiple RATs, also called Multi-Band Multi-Standard Radio (MB-MSR) base stations. The specifications cover all combinations of RATs, except pure single-RAT GSM operation across the supported bands.

There are several scenarios envisioned for multi-band base station implementation and deployment. The possibilities for the multi-band capability are:

- Multi-band transmitter + multi-band receiver
- Multi-band transmitter + single-band receiver
- Single-band transmitter + multi-band receiver

The first case is demonstrated in [Figure 19.12](#), which shows an example base station with a common RF implementation of both transmitter and receiver for two operating bands X and Y. Through a duplex filter, the transmitter and receiver are connected to a common antenna connector and a common antenna. The example is also a multi-RAT capable MB-MSR base station, with LTE+GSM configured in Band X and LTE configured in Band Y. Note that the figure has only one diagram showing the frequency range for the two bands, which could either be the receiver or transmitter frequencies.

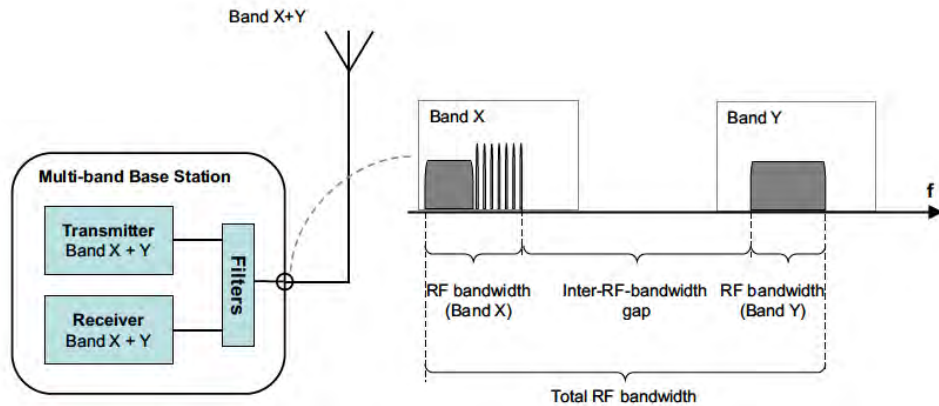


FIGURE 19.12

Example of multi-band base station with multi-band transmitter and receiver for two bands with one common antenna connector

Figure 19.12 also illustrates some new parameters that are defined for multi-band base stations.

- **RF bandwidth** has the same definition as for a multi-standard base station, but is defined individually for each band.
- **Inter-RF-bandwidth gap** is the gap between the RF bandwidths in the two bands. Note that the inter-RF-bandwidth gap may span a frequency range where other mobile operators can be deployed in bands X and Y, as well as a frequency range between the two bands that may be used for other services.
- **Total RF bandwidth** is the full bandwidth supported by the base station to cover the multiple carriers in both bands.

In principle, a multi-band base station can be capable of operating in more than two bands. However, the requirements and testing developed for the new type of base stations in 3GPP release 11 will in general only cover a two-band capability.

While having only a single antenna connector and a common feeder that connects to a common antenna is desirable to reduce the amount of equipment needed in a site, it is not always possible. It may also be desirable to have separate antenna connectors, feeders, and antennas for each band. An example of a multi-band base station with separate connectors for two operating bands X and Y is shown in Figure 19.13. Note that while the antenna connectors are separate for the two bands, the RF implementation for transmitter and receiver are in this case common for the bands. The RF for the two bands is separated into individual paths for Band X and Band Y before the antenna connectors through a filter. As for multi-band base

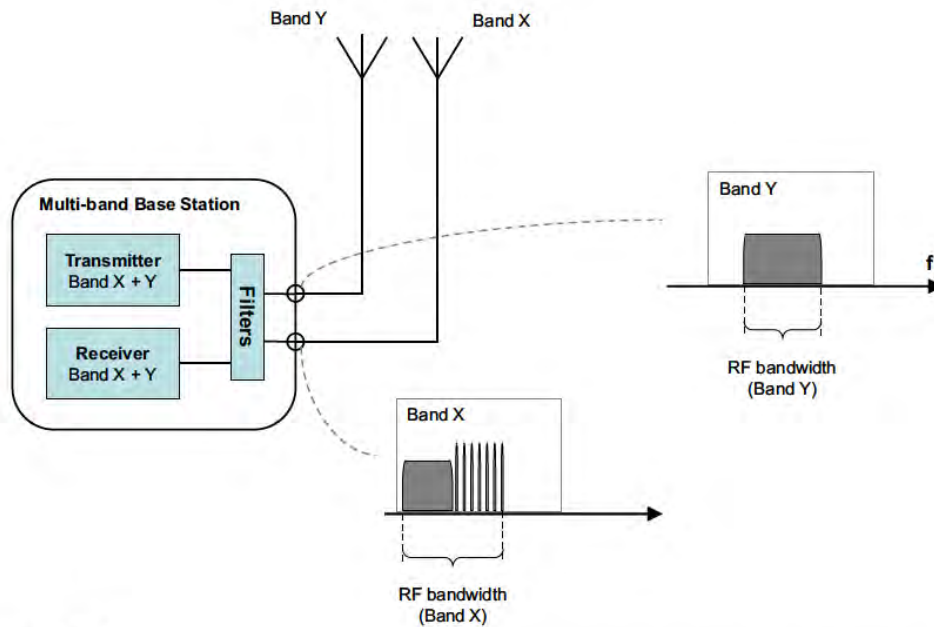


FIGURE 19.13

Multi-band base station with multi-band transmitter and receiver for two bands with separate antenna connectors for each band

stations with a common antenna connector for the bands, it is also possible to have either the transmitter or receiver be a single-band implementation, while the other is multi-band.

Further possibilities are base station implementations with separate antenna connectors for receiver and transmitter, in order to give better isolation between the receiver and transmitter paths. This may be desirable for a multi-band base station, considering the large total RF bandwidths which will in fact also overlap between receiver and transmitter.

For a multi-band base station, with a possible capability to operate with multiple RATs and several alternative implementations with common or separate antenna connectors for the bands and/or for the transmitter and receiver, the declaration of the base station capability becomes quite complex. What requirements will apply to such a base station and how they are tested will also depend on these declared capabilities.

Most RF requirements for a multi-band base station remain the same as for a single-band implementation. There are however some notable exceptions:

- **Transmitter spurious emissions:** For LTE base stations, the requirements exclude frequencies in the operating band plus an additional 10 MHz on each side of the

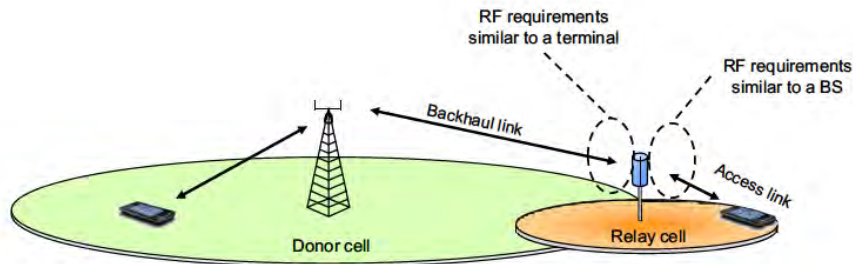
operating band, since this frequency range is covered by the UEM limits. For a multi-band base station, the exclusion applies to both operating bands (plus 10MHz on each side), and only the UEM limits apply in those frequency ranges. This is called “joint exclusion band.”

- **Operating band unwanted emissions mask (UEM):** For multi-band operation, when the inter-RF-bandwidth gap is less than 20 MHz, the UEM limit applies as a cumulative limit with contributions counted from both bands, in a way similar to operation in non-contiguous spectrum.
- **ACLR:** For multi-band operation, when the inter-RF-bandwidth gap is less than 20 MHz, the Cumulative ACLR (CACLR) will apply with contributions counted from both bands, in a way similar to operation in non-contiguous spectrum.
- **Transmitter intermodulation:** For a multi-band base station, when the inter-RF-bandwidth gap is less than 15 MHz, the requirement only applies for the case when the interfering signals fit within the gap.
- **Blocking requirement:** For multi-band base stations, the in-band blocking limits apply for the in-band frequency ranges of *both* operating bands. This can be seen as a “joint exclusion,” similar to the one for spurious emissions. The blocking and receiver intermodulation requirements also apply inside the inter-RF-bandwidth gap.
- **Receiver spurious emissions:** For a multi-band base station, a “joint exclusion band” similar to the one for transmitter spurious emissions will apply, covering both operating bands plus 10MHz on each side.

In the case where the two operating bands are mapped on separate antenna connectors, as shown in Figure 19.13, the preceding exceptions for transmitter/receiver spurious emissions, UEM, ACLR, and transmitter intermodulation do not apply. Those limits will instead be the same as for single-band operation for each antenna connector. In addition, if such a multi-band base station with separate antenna connectors per band is operated in only one band with the other band (and other antenna connector) inactive, the base station will from a requirement point-of-view be seen as a single-band base station. In this case all requirements will apply as single-band requirements.

19.14 RF requirements for relays

As described in Chapter 18, the LTE specifications support decode-and-forward relays in release 10. The RF requirements for such relays were introduced in release 11 and are described in this chapter. The baseline for setting the relay RF requirements has been the existing RF requirements for the base station and the terminal, with the observation that seen from the access link side, a relay would have many similarities with a base station, while on the backhaul side, it would have similarities with a terminal. This is illustrated in Figure 19.14.

**FIGURE 19.14**

RF requirements for a relay on the backhaul and access sides

The RF requirements for LTE relays are defined in a separate specification [RELAYRF], but many of the requirements are set by direct reference to the terminal specification [76] for backhaul link requirements and base station specification [86] for access link requirements. It is in particular the Local Area base station requirements that are referenced, since the deployment scenario in that case is similar to a relay. Many RF requirements are, however, specific for relays, the following in particular:

Output power: Two power classes are defined for relays. Class 1 has a maximum of 24 dBm and Class 2 has a maximum 30 dBm output power for the access link. The maximum for the backhaul link is 24 dBm for both classes. All power levels are counted as the sum over all antennas (up to eight for the access link and four for the backhaul link).

ACLR: In order for the backhaul link to provide proper co-existence properties and not deteriorate the base station uplink, the relay ACLR limits are set equivalent to the ones for a Local Area base station on both the access and backhaul links.

Operating band unwanted emissions (UEM): New UEM limits for the access and backhaul links are set in relation to the output power levels defined by the Relay Power Class.

Adjacent Channel Selectivity (ACS): The relay backhaul link relative ACS limit is set at a level similar to what is used for a base station, but the requirement is defined at a higher input signal level corresponding to the maximum downlink signal level at which ACS is defined for a terminal. The ACS requirement for the access link is taken from the Local Area base station.

Blocking requirements: The in-band blocking levels for the backhaul link are set higher than for a terminal, but are also defined for a correspondingly higher wanted signal level. The blocking requirement for the access link is taken from the Local Area base station.

This page intentionally left blank

20.1 Performance assessment

Computer simulation of mobile systems is a very powerful tool to assess the system performance. The “real-life” performance can, of course, be measured and evaluated in the field for an already deployed system, and such values represent a valid example of performance for a certain system configuration. But there are several advantages with computer simulations:

- Evaluation can be made of system concepts that are not deployed or still under development.
- There is full control of the environment, including propagation parameters, traffic, system layout, etc., and full traceability of all parameters affecting the result.
- Well-controlled “experiments” comparing similar system concepts or parts of concepts can be done under repeatable conditions.

In spite of the advantages, simulation results do not give a full picture of the performance of a system. It is impossible to model all aspects of the mobile environment and to properly model the behavior of all components in a system. Still, a very good picture of system performance can be obtained and it can often be used to find the potential limits for performance. Because of the difficulty in modeling all relevant aspects, relative capacity measures for introducing features will be more accurate than absolute capacity numbers, if a good model is introduced of the feature in question.

The capacity of a system is difficult to assess without comparisons, since a system performance number in itself does not provide very much information. It is when set in relation to how other systems can perform that the number becomes interesting. Since making comparisons is an important component in assessing performance, it also makes performance numbers a highly contentious issue. One always has to watch out for “apples and oranges” comparisons, since the system performance depends on so many parameters. If parameters are not properly selected to give comparable conditions for two systems, the performance numbers will not be comparable either.

In this context, it is also essential to take into account that system performance and capacity will be feature dependent. Many features such as MIMO and advanced antenna techniques that have been introduced in 3G (HSPA) and 4G (LTE) systems are very similar

between systems. If a certain feature is a viable option for several systems that are evaluated in parallel, the feature should be included in the evaluation for all systems.

Any simulated performance number should be viewed in the context that real radio network performance will depend on many parameters that are difficult to control or model, including:

- The mobile environment, including channel conditions, angular spreads, clutter type, terminal speeds, indoor/outdoor usage, and coverage holes.
- User-related behavior, such as voice activity, traffic distribution, and service distribution.
- System tuning of service quality and network quality.
- Deployment aspects such as site types, antenna heights and types, and frequency planning.
- A number of additional parameters that are usually not modeled, such as signaling capacity and performance, and measurement quality.

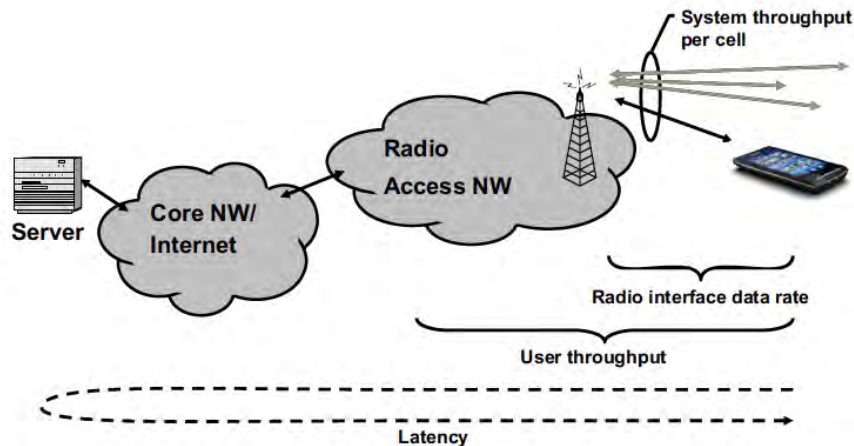
There is no single universal measure of performance for a telecommunications system. Indeed, end-users (subscribers) and system operators define good performance quite differently. On the one hand, end-users want to experience the highest possible level of quality. On the other hand, operators want to derive maximum revenue, for example by squeezing as many users as possible into the system. Performance-enhancing features can improve perceived quality of service (end-user viewpoint) or system performance (operator viewpoint). The good news, however, is that LTE and its evolution has the potential to do both. Compared to 3G systems, LTE yields better data rates and shorter delays. That is, LTE and its evolution can greatly improve both the service experience (end-user viewpoint) and the system capacity (operator viewpoint).

20.1.1 End-user perspective of performance

Users of circuit-switched services are assured of a fixed data rate. The quality of service in the context of speech or video telephony services is defined by perceived speech or video quality. Superior-quality services have fewer bit errors in the received signal.

By contrast, users who download a web page or movie clip via packet data describe quality of service in terms of the delay they experience from the time they start the download until the web page or movie clip is displayed. Best-effort services do not guarantee a fixed data rate. Instead, users are allocated whatever data rate is available under present conditions. This is a general property of packet-switched networks—that is, network resources are not reserved for each user. Given that the delay increases with the size of the object to be downloaded, absolute delay is not a fair measure of quality of service.

Performance of a packet data service in cellular systems can be characterized by several different measures depending on the perspective taken (see [Figure 20.1](#)). A single user in a radio network experiencing good radio conditions may enjoy the *peak data rate* of the radio interface. A user will, however, normally share radio resources with other users. If radio

**FIGURE 20.1**

Definitions of data rates for performance

conditions are less than optimal or there is interference from other users, the radio-interface data rate will be less than the peak data rate. In addition, some data packets might be lost, in which case the missing data must be retransmitted, further reducing the effective data rate as seen from higher protocol layers. Additionally, the effective data rate diminishes even further as the distance from the cell increases (due to poorer radio conditions at cell edges). The data rate experienced above the MAC layer, after sharing the channel with other users, is denoted *user throughput*.

The *Transmission Control Protocol* (TCP)—the protocol at the transport layer—is commonly used together with IP traffic. However, due to the so-called slow-start algorithm, which is sensitive to latency in the network, it is especially prone to cause delay for small files. The slow-start algorithm is meant to ensure that the packet transmission rate from the source does not exceed the capability of network nodes and interfaces.

Network *latency*, which in principle is a measure of the time it takes for a packet to travel from a client to a server and back again, has a direct impact on performance with TCP. Therefore, an important design objective for LTE has been to reduce network latency. One other quality-related criterion (end-user viewpoint) relates to the setup time for initiating, for example, a web-browsing session.

20.1.2 Operator perspective

Radio resources need to be shared when multiple users are in the network. As a result, all data must be queued before it can be transmitted, which restricts the effective data rate to each

user. Notwithstanding this fact, by scheduling radio resources, operators can improve system throughput or the total number of bits per second transmitted over the radio interface. A common measure of system performance is “spectral efficiency,” which is the system throughput per MHz of spectrum in each cell of the system.

LTE employs intelligent scheduling methods to optimize performance, from both end-user and operator viewpoints.

An important performance measure for operators is the number of active users who can be connected simultaneously. Given that system resources are limited, there will thus be a trade-off between number of active users and perceived quality of service in terms of user throughput.

20.2 Performance in terms of peak data rates and latency

As described in Chapter 1, LTE has been developed in a process where design targets for performance parameters play an important role. One target is for the peak data rate over the radio interface. The original design targets for the first release of LTE are documented in 3GPP TR 25.913 [10]. The target capability when operating in a 20 MHz spectrum allocation is a peak data rate of 100 Mbit/s in the downlink and 50 Mbit/s in the uplink. The numbers assume two receive antennas in the terminal for the downlink capability and one transmit antenna for the uplink capability. These target numbers are exceeded by a good margin by the peak data rate capability of the specified LTE standard. LTE release 8 supports peak data rates of 300 Mbit/s in the downlink and 75 Mbit/s in the uplink by using spatial multiplexing of four layers (4×4 MIMO) in the downlink and 64QAM in both downlink and uplink. With the assumptions in the design targets—that is, spatial multiplexing of two layers—the downlink peak data rate is 150 Mbit/s, which is still significantly higher than the target.

The design targets for LTE release 10 (“LTE-Advanced”) are documented in 3GPP TR 36.913 [10], based on the targets set by ITU-R [65]. There is no absolute peak data rate target expressed for LTE release 10; it is instead expressed relative to the channel bandwidth as a peak spectral efficiency, with targets of 15 bit/s/Hz for downlink and 6.75 bit/s/MHz for uplink [65]. LTE release 10 exceeds those numbers by a good margin, as shown in Table 20.1.

Table 20.1 LTE Peak Spectral Efficiency [TR 36912]					
Peak Spectral Efficiency	3ITU Requirement [bit/s/Hz]	LTE Fulfillment			
		Release 8		Release 10	
		FDD	TDD	FDD	TDD
Downlink	15	15.3	15.0	30.6	30.0
Uplink	6.75	4.2	4.0	16.8	16.0

The assumptions for deriving the peak spectral efficiency numbers are a deployment with 20 MHz channel bandwidth, 8×8 MIMO in the downlink, and 4×4 MIMO in the uplink.

The ITU-R requirement for downlink peak spectral efficiency is in fact fulfilled by LTE release 8, assuming 4×4 MIMO in the downlink [92].

The ITU-R target for *latency* in [65] is set with a different definition than in Figure 20.1. Instead of a two-way *latency*, there is a 10 ms target for the *one-way latency* in both downlink and uplink. The one-way latency is defined as the one-way transit time between a packet being available at the IP layer in the user base station and the availability of this packet at the IP layer in the user terminal, and vice versa. This is achieved with a good margin for LTE, where the latency achieved is 4 ms for LTE FDD and 4.9 ms for LTE TDD.

20.3 Performance evaluation of LTE-advanced

An essential part of the submission of LTE release 10 to ITU-R as a candidate for IMT-Advanced was the performance evaluation of the spectral efficiency. The technical requirements are set by ITU-R Report M.2134 [65] and the detailed evaluation methodology is described in ITU-R Report M.2135 [93].

For the work on LTE release 10, 3GPP performed a large simulation campaign with input from several 3GPP members. The resulting performance numbers formed an essential part of the submission of LTE-Advanced as a candidate for IMT-Advanced and is reported in detail in 3GPP TR 36.912 [92]. In addition to the ITU-R evaluation criteria, an additional test environment with higher performance targets than the ones set by ITU-R was defined in 3GPP. Performance numbers for this additional environment are also reported in [92].

An evaluation based on the ITU-R methodology and assumptions is presented in [94] and is also outlined below. The evaluation is based on time-dynamic system simulations, where users are dropped independently with uniform distribution over a simulated LTE system with a large number of cells. The simulations include overhead for control channels and time-dynamic models of feedback channels. With the level of details applied to model the protocols and the simulation assumptions used for the evaluation described below, the technology potential of LTE release 10 is demonstrated, in addition to showing that the IMT-Advanced performance requirements are exceeded.

Detailed evaluations of LTE release 8 in comparison to earlier 3GPP releases of HSPA can be found in [95,96].

20.3.1 Models and assumptions

This section presents the models and assumptions used for the evaluation in [94]. A summary of the test environments is given in Table 20.2, and the assumptions for the LTE-Advanced system characteristics are given in Table 20.3.

Table 20.2 Test Environment and Deployment Parameters for the Evaluation				
	Test Environment			
	Indoor	Microcellular	Base Coverage Urban	High Speed
Deployment scenario	Indoor hotspot	Urban Micro	Urban Macro	Rural Macro
Channel model	InH	UMi	UMa	RMa
Carrier frequency	3.4 GHz	2.5 GHz	2.0 GHz	0.8 GHz
Inter-site distance	60 m	200 m	500 m	1732 m
Terminal speed	3 km/h	3 km/h	30 km/h	120 km/h
User distribution	100% indoor	50% indoor 50% outdoor	100% outdoor	100% outdoor
BS antenna height	6 m	10 m	25 m	35 m
BS antenna gain	0 dBi	17 dBi	17 dBi	17 dBi
BS output power	21 dBm/ 20 MHz	44 dBm/ 20 MHz	49 dBm/ 20 MHz	49 dBm/ 20 MHz
UE output power	21 dBm	24 dBm	24 dBm	24 dBm

The evaluation is performed in four deployment scenarios, each corresponding to a different test environment defined by ITU-R [93]:

- **Indoor Hotspot:** Deployment scenario for the *indoor* environment, having isolated office cells or hotspots for stationary or pedestrian users, with high user density and high user throughput.
- **Urban Micro:** Deployment scenario for the *microcellular* environment, having small cells with outdoor and outdoor-to-indoor coverage for pedestrian and slow vehicular users, provided by below rooftop outdoor base stations. It has high traffic per unit area and will be interference limited.
- **Urban Macro:** Deployment scenario for the *base coverage urban* environment, having large cells and continuous coverage for pedestrians up to fast vehicular users, provided by above rooftop outdoor base stations. It will be interference limited and has mostly non-line-of-sight propagation conditions.
- **Rural Macro:** Deployment scenario for the *high speed* environment, having large cells and continuous coverage for high-speed vehicular and train users. It will be noise and/or interference limited.

Each deployment scenario described in [93] corresponds to a set of deployment parameters, including detailed propagation models for line-of-sight, non-line-of-sight, and/or outdoor-to-indoor propagation. The channel model is based on the so-called WINNER II channel model [97], which is a Spatial Channel Model (SCM) applicable for MIMO simulations [98]. Further details of the test environments and deployment parameters are given in Table 20.2.

Table 20.3 LTE-Advanced System Characteristics

General Characteristics	
Duplex method assumptions	FDD TDD: Configuration 1, DwPTS/GP/UpPTS length set to 12/1/1 OFDM symbols (see Chapter 9)
Spectrum allocation	10 MHz DL + 10 MHz UL for FDD, 10 MHz for TDD Double bandwidth for InH case
Antenna configuration at base station (BS)	Vertically co-polarized antennas separated 4 (InH) or 0.5 (UMi, UMa, RMa) wavelengths separation
Antenna configuration at terminal (UE)	Vertically co-polarized 0.5 wavelengths separation
Network synchronization	Synchronized, not explicitly utilized other than for avoiding UE-UE and BS-BS interference for TDD
Detailed Radio Interface Characteristics and Models	
Scheduler	DL: Proportional Fair in Time and Frequency UL: Quality-based Frequency Domain Multiplexing
Downlink transmission scheme	InH: Transmission mode 4; closed-loop codebook-based pre-coded adaptive rank spatial multiplexing UMi, UMa, RMa: Transmission mode 5; coordinated beam-forming (within site) with MU-MIMO
Uplink transmission scheme	1 Tx, 4Rx antennas, no MU-MIMO
Receiver type	Minimum Mean Squared Error (MMSE) in DL and UL
Uplink power control	Open loop with fractional path-loss compensation, parameters chosen according to the deployment scenario. Effective noise rise below 10 dB.
Hybrid-ARQ scheme	Incremental redundancy, synchronous, adaptive
Link adaptation	Non-ideal, based on delayed feedback
Channel estimation	Non-ideal channel estimation Non-ideal channel-state reports in downlink; CQI error per resource block is $N(0,1)$ dB, error-free feedback of the reports, 6 ms reporting delay, 5 ms reporting periodicity Uplink quality estimated from PUSCH, 6 ms delay, 20 ms sounding period
Feedback channel errors	Error-free, but quantized and delayed
Control channel overhead	DL: 3 OFDM symbols per subframe UL: 4 resource blocks Overhead for common control channels (synchronization, broadcast and random access; $\sim 1\%$ for 10 MHz) has not been deducted

The simulation methodology is “time dynamic,” where a system of base stations and terminals is simulated over a limited time frame (20–100 seconds). This is repeated to create a number of samples to reach sufficient statistical confidence. For each simulation, terminals are randomly positioned over a model of a radio network and the radio channel between each base station and terminal antenna pair is simulated according to the propagation and fading models. A full buffer traffic model with an average of 10 users per cell is assumed. This results in the system operating at the maximum 100% system load.

Based on the channel realizations and the active interferers, a *signal-to-interference-and-noise ratio* (SINR) is calculated for each terminal (or base-station) receive antenna. The SINR values are then mapped to *block error probability* for the modulation and coding scheme employed for each user. MIMO and the modulation-and-coding scheme according to the LTE standard are selected based on delayed feedback. Retransmissions are explicitly modeled and the user throughput for each active user i will be the number of correctly received bits χ_i (above the MAC layer) divided by the simulated time T . The distribution of user throughput between users serves as a basis for assessing end-user quality. The served traffic per cell is calculated from the total number of received bits χ_i for all users, averaged over all cells and divided by the simulated time T . Statistics are collected from each simulation run and then new terminals are randomly positioned for the next sample.

20.3.2 Evaluation criteria

ITU-R defines two requirements related to the efficiency of the radio interface for evaluating the performance of the IMT-Advanced candidate radio-interface technologies (RITs) [65]. The first is *cell spectral efficiency*, defining the operator perspective, and the second is *cell-edge spectral efficiency*, defining the end-user perspective.

The *cell spectral efficiency* is the aggregated throughput over all users, averaged over all cells and divided by the channel bandwidth. The measure relates to the system throughput in Figure 20.1 and is a measure of the maximum total “capacity” available in the system to be shared between users; it is measured in bits/s/Hz/cell.

The cell spectral efficiency η is defined as:

$$\eta = \frac{\sum_{i=1}^N \chi_i}{T \cdot \omega \cdot M} \quad (20.1)$$

where χ_i denotes the number of correctly received bits for user i in a system with N users and M cells, ω is the channel bandwidth, and T is the time over which the data bits are received.

The *cell-edge user spectral efficiency* is based on the distribution between users of the *normalized user throughput* (see also Figure 20.1), which is defined as the average user throughput over a certain period of time divided by the channel bandwidth, and is measured in bit/s/Hz. The cell-edge user spectral efficiency is defined as the 5% point of the cumulative distribution function (CDF) of the normalized user throughput. It is thus a measure of the end-user perceived “quality of service” for the 5% of the users with the lowest user throughput.

Table 20.4 ITU-R Requirements for IMT-Advanced Spectral Efficiency

Test Environment and Corresponding Deployment Scenario	Cell Spectral Efficiency [bit/s/Hz/cell]		Cell-Edge User Spectral Efficiency [bit/s/Hz]	
	Downlink	Uplink	Downlink	Uplink
Indoor (InH)	3	2.25	0.1	0.07
Microcellular (UMi)	2.6	1.8	0.075	0.05
Base coverage urban (UMa)	2.2	1.4	0.06	0.03
High speed (RMa)	1.1	0.7	0.4	0.015

The normalized user throughput for user i is defined as:

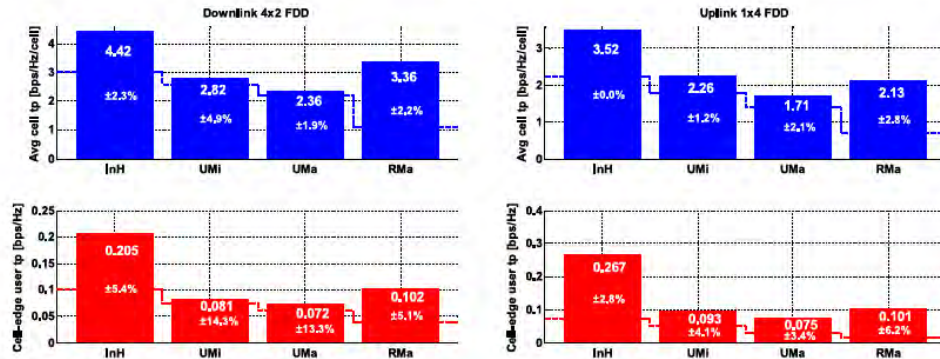
$$\gamma_i = \frac{\chi_i}{T_i \cdot \omega} \quad (20.2)$$

where T_i is the active session time for user i .

The requirements defined by ITU-R [65] for cell spectral efficiency and cell-edge user spectral efficiency are listed in Table 20.4. The requirement levels are shown as dashed lines together with the simulated performance values in Figures 20.2 and 20.5.

20.3.3 Performance numbers

Simulations for FDD and TDD and for all test environments were submitted for the ITU-R evaluation and are also presented here. The numbers here, however, are not identical to the

**FIGURE 20.2**

FDD cell spectral efficiency and cell-edge user spectral efficiency, compared with ITU-R requirements (downlink and uplink)

numbers put forward to ITU-R in [92], but will be seen as one sample of results from the evaluation in [94].

For FDD, the cell spectral efficiency and cell-edge spectral efficiency are shown in Figure 20.2. All results exceed the ITU-R performance requirements in Table 20.4, in some cases by a very good margin. For the indoor scenario and the downlink, this is achieved using an uncorrelated antenna configuration together with single-user MIMO. For the other scenarios, a correlated antenna setup and intra-site coordinated beam-forming (CBF) with MU-MIMO are used in the downlink. A quite basic uplink configuration not utilizing MIMO

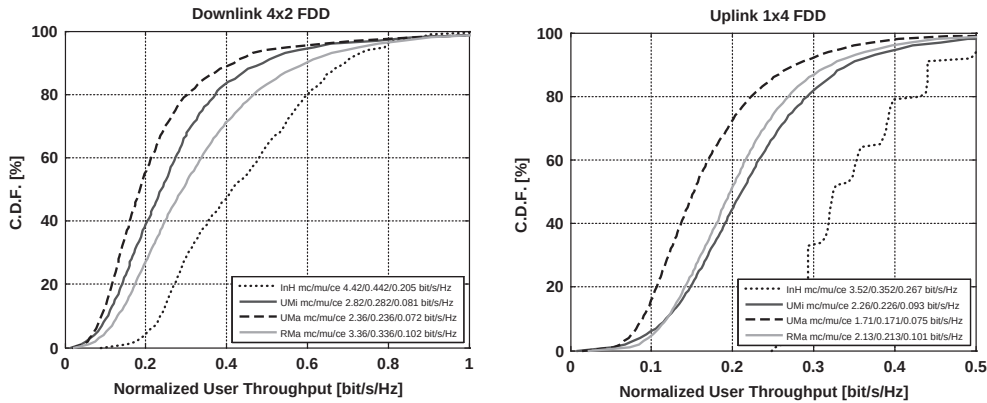


FIGURE 20.3

FDD normalized user throughput distributions (downlink and uplink)

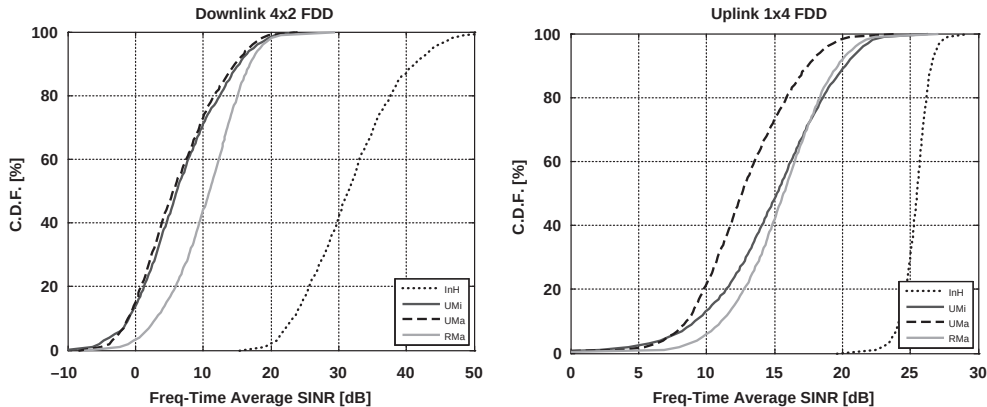


FIGURE 20.4

FDD SINR distributions (downlink and uplink)

is used for all environments. It should be noted that the cell-edge spectral efficiency is highly dependent on the system load and that the ITU-R simulations are performed with maximum load and a full buffer model. An example of LTE performance numbers including lower and varying loads can be found in [96].

The normalized user throughput distributions are given in Figure 20.3 and Signal-to-Noise-and-Interference Ratio (SINR) distributions are shown in Figure 20.4. The figures demonstrate the impact of the very advantageous SINR distributions achieved in the indoor

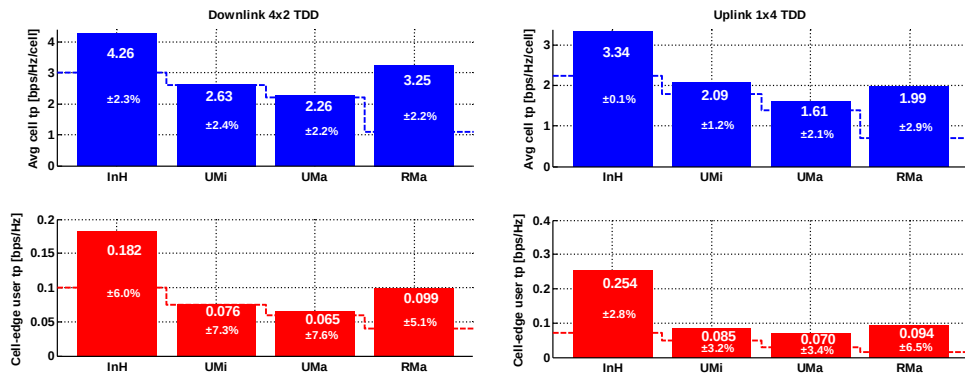


FIGURE 20.5

TDD cell spectral efficiency and cell-edge user spectral efficiency, compared with ITU-R requirements (downlink and uplink)

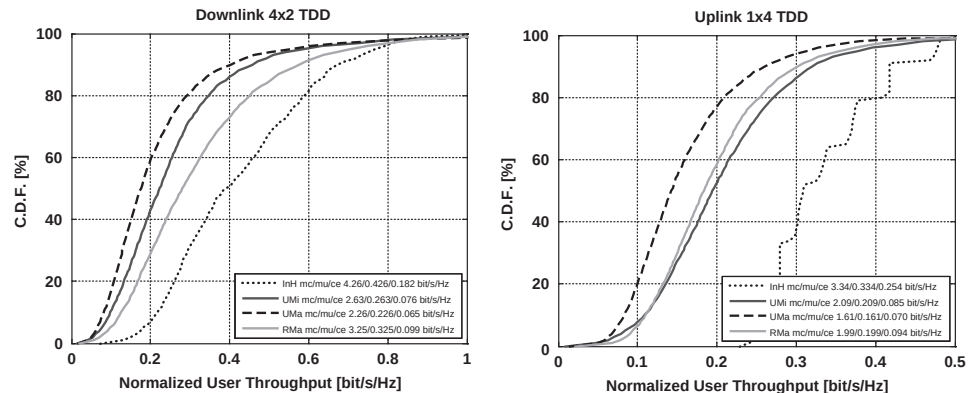


FIGURE 20.6

TDD normalized user throughput distributions (downlink and uplink)

hotspot scenario and to some extent also for rural macro, resulting in higher spectral efficiency numbers in those environments.

For TDD, the cell spectral efficiency and cell-edge spectral efficiency are shown in Figure 20.5, while normalized user throughput distributions are illustrated in Figure 20.6. There is a small drop in the spectral efficiency numbers compared to FDD, but all results exceed the ITU-R performance requirements. The difference compared to FDD is due to a higher overhead from the guard period between uplink and downlink. The average delay between making measurements in the terminal and receiving the measurement result at the base station is also increased compared to FDD, due to the TDD time-domain structure. This has some impact on the performance of scheduling and link adaptation. The channel reciprocity is not utilized in the simulated TDD scheme.

20.4 Conclusion

The simulation results presented in this chapter demonstrate the high potential of LTE/LTE-Advanced in terms of both overall spectral efficiency to the benefit of operators and high cell-edge performance to the benefit of the end-user. The performance is assessed in four different test environments. The results for FDD and TDD in downlink and uplink all exceed the requirements set up by ITU-R for the evaluation of IMT-Advanced candidates. In addition, the peak radio-interface data rates and the latency achieved by LTE meet the ITU-R requirements.

Future Radio Access—Final Thoughts

21

This book has provided an extensive description of the LTE radio-access technology, from its emergence as part of 3GPP release 8, via releases 9 and 10 to current LTE release 11. However, the radio-access evolution does not end with release 11. New releases of LTE will emerge in the future, introducing new features and technology components to further enhance system performance and address new usage scenarios. Also, at some point in time new “beyond-LTE” radio-access technology will undoubtedly appear. The aim of this last chapter is to provide an outline of future radio access, including the future evolution of LTE as well as potential new radio-access technology.

The overall traffic in mobile-communication networks has grown tremendously in recent years, fueled primarily by an enormous uptake in the use of mobile broadband. This trend is expected to continue in the future, with traffic demands predicted to increase around 12 times from 2012 to 2018 [112]. In a longer-term perspective, mobile-communication networks will have to be able to handle more than 1000 times higher traffic than today [123]. This increase in traffic will to a large extent be driven by further increase in mobile-broadband usage. However, especially in a longer-term perspective, the traffic growth will also be fueled by emerging applications and new use cases associated with “communicating machines” as elaborated on below.

There are in the order of five billion wirelessly connected mobile devices in service today, most of which are hand-held terminals, tablets, or mobile-broadband devices in portable computers. In the future, these devices may be surpassed tenfold or more by communicating machines including, for example, surveillance cameras, smart-home and smart-grid devices, and connected sensors. The transition from roughly five billion to more than 50 billion, and in a longer-term perspective several 100 billion, connected devices will pose a formidable challenge for future mobile-communication systems [124].

Future mobile-communication systems will also have to cope with a much wider range of applications and use cases, with a corresponding much wider range of requirements and characteristics, compared to today. Mobile broadband services, including web browsing, video streaming, and different forms of data sharing will remain a key application and will continue to drive a demand for higher data rates. Future demands for data rates exceeding 10 Gbit/s in specific scenarios and general availability of data rates at least in the order of 100 Mbit/s can be anticipated.

In addition, the large-scale introduction of communicating machines will go hand in hand with the emergence of many new applications and use cases with characteristics and requirements that will not only differ significantly from the currently dominating mobile-broadband applications, but which will also vary significantly between applications.

- Some applications, such as safety or control mechanisms in the process industry or in the electrical grid, may have very strict requirements on low latency while for other applications, such as remote-meter reading for billing purposes, the latency requirements may be very relaxed.
- Applications related to the control of critical infrastructure, for example the electrical grid or industrial control, may require extremely high reliability whereas in other cases, for example for certain types of sensors such as temperature or moisture sensors in the home, the reliability requirements may be much lower.
- Some applications, such as remote surveillance cameras, may be associated with major volumes of information to be conveyed while other applications, such as cargo tracking in the shipping industry, may be associated with very small data payloads.
- In some cases, for example for battery-powered sensor networks, low device cost and/or energy consumption are extremely important, while for other applications, this may be less of an issue.

In order to address the above challenges of massive traffic increase, massive number of connected devices, and a very wide range of requirements and characteristics, in an affordable and sustainable way, new technology components are required, both as part of the continued evolution of existing radio-access technologies or as building blocks for new beyond-LTE radio access.

21.1 Continued LTE evolution beyond release 11

As already mentioned, the evolution of LTE does not end with LTE release 11. Rather LTE will continue to evolve into release 12, release 13, and so on, with each new release bringing additional capabilities and further enhanced system performance into the LTE radio-access technology. Not only will this provide better performance for existing use cases, it will also allow for new use cases.

The starting point for the 3GPP activities on LTE release 12 was the workshop “Release 12 and beyond,” taking place in June 2012. At that workshop it was agreed that important aspects to be addressed by the evolution include not only increased capacity and enhanced end-user experience but also improved energy efficiency, reduced cost, better support for diverse data applications, and backhaul enhancements. The major technology areas identified at the workshop comprised [114]:

- Further enhanced local-area access
- Extended multi-antenna/multi-point transmission

- Further improved support for machine-type communication (MTC)
- Support for direct device-to-device communication

In the following text, these areas will be discussed, primarily focusing on release 12.

21.1.1 Further enhanced local-area access

Network densification, typically by deploying low-power nodes under the coverage of an existing macro-node layer as described in Chapter 16, is a key tool to improving traffic capacity and extending the achievable data rates. The low-power nodes provide *local-area access*, complementing the wide-area-covering macro layer in selected areas. As most users are stationary or slowly moving, there is an increasing focus on high data rates in (quasi-) stationary situations, and improved support for local-area access was identified as a key area at the release-12 workshop. To some extent it can be seen as the, in a wide sense, continuation of the work on improved support for heterogeneous deployments in releases 10/11 as described in Chapter 16.

In the following, examples of some aspects related to local-area access considered by 3GPP as part of release 12 are given. Additionally, 3GPP is also studying various forms of interference coordination to handle dense local-area deployments, a new carrier type with less overhead as a mean to reduce interference and increase energy efficiency, and physical-layer modifications such as the introduction of 256QAM.

21.1.1.1 Frequency-separated local area access

The 3GPP activities on local-area access and heterogeneous deployments have up to and including release 11 primarily focused on *same-frequency* operation—that is, when the wide-area and local-area layers operate on the same carrier frequency. The main reason for this has been the assumption that, especially for operators with a limited spectrum situation, it is not justifiable to split the available spectrum between the layers, reducing the bandwidth and thus also the achievable data rates available in each layer. Same-frequency operation also maximizes the spectral efficiency but, as discussed in Chapter 16, results in an inter-layer interference problem that needs to be handled. Hence, the features in focus for releases 10/11 have primarily targeted handling of inter-layer interference between the different layers in a same-frequency deployment.

However, in the future additional spectrum will be primarily available at higher frequencies, 3.5 GHz and above, as lower frequencies are already heavily used by cellular as well as non-cellular services. In general, higher frequency bands are, due to propagation characteristics, more challenging to use for wide-area deployments. Furthermore, there are regulatory limitations on the output power and outdoor usage of some frequency bands in certain parts of the world.

With the availability of higher frequency bands more challenging to use for wide-area coverage, it is more relevant to consider band-separated local-area access operating on higher frequency bands with the overlaid macro layer operating on lower cellular bands. Not

only does such a *frequency-separated* deployment avoid the inter-layer interference issues present in a same-frequency deployment, but it also provides some additional benefits compared to same-frequency operation, for example the possibility to use different duplex schemes in the different network layers.

21.1.1.2 *Dynamic TDD*

With an increased interest in local-area deployments, TDD is expected to become more important compared to the situation for wide-area deployments today. For example, an existing wide-area FDD network could be complemented by a local-area layer using TDD. To better handle the high traffic dynamics in a local-area scenario, where the number of terminals associated with a certain access node can be very small, *dynamic TDD* is beneficial. In dynamic TDD, the network can *dynamically* use resources for either uplink or downlink transmissions to match the *instantaneous* traffic situation, which leads to an improvement of the end-user performance compared to the conventional static split of resources between uplink and downlink.

21.1.1.3 *Dual connectivity*

Time-domain partitioning and shared cell, described in Chapter 16, are two examples of inter-layer coordination schemes. In the former, the local-area nodes transmit all the signals associated with a cell, including cell-specific reference signals, synchronization signals, and the full set of system information. A terminal communicates with the *same* node, macro or local-area node, in both uplink and downlink directions and the transmission activity in the different layers is coordinated on a relatively slow basis to handle the inter-layer interference. Shared cell, on the other hand, is based on a much tighter interaction between the cell layers and in principle a terminal can communicate with different nodes in uplink and downlink. However, the benefits are limited to nodes being part of the same cell and low-latency backhaul is required.

On a high level, *dual connectivity* can be seen as taking the principle of operating the wide-area and local-area layers in a more integrated manner one step further. In this scenario, the terminal is simultaneously connected to *two* network nodes. The nodes may or may not form separate cells and there is no need for low-latency backhaul. The integrated operation of the two layers can lead to benefits in terms of:

- *Uplink-downlink separation*—that is, uplink reception may occur in a network node different from that used for downlink transmissions. Uplink-downlink separation is particularly advantageous in a heterogeneous deployment where the node with the strongest received signal at the terminal (and thus typically used for downlink transmissions) may not be the node with the lowest path loss to the terminal (and thus not the best node for uplink reception) as the difference in transmission power or load between the wide-area and local-area nodes can be quite large, as discussed in Chapter 16. Uplink CoMP, discussed in Chapter 15, is one way of implementing uplink-downlink separation assuming a low-latency backhaul is available. With dual connectivity, the benefits can be extended to the case of non-ideal backhaul.

- *Mobility robustness*—that is, vital control-plane information such as handover commands can be transmitted from the overlaid wide-area layer even if the user data is provided by the local-area layer. A terminal being connected to a single node only may lose the connection if it has moved outside of the coverage area of the local-area node before the handover procedure is completed, a problem that is avoided if the overlaid wide-area layer is responsible for transmitting handover commands. Alternatively, the handover commands can be transmitted from both the source and the destination nodes.
- *Interference reduction* as there is no need for the local-area layer to broadcast system information when system information already is provided by the wide-area layer. Furthermore, minimizing the transmission of always-on signals such as system information and cell-specific reference signals from the local-area layer enables higher energy efficiency, as the local-area access nodes in principle only need to be active in the subframes where there is a terminal to serve.

In Figure 21.1, an example of dual connectivity is illustrated where the terminal is connected to an anchor node in the wide-area layer for system information, basic radio-resource control (RRC) signaling and possibly low-rate/high-reliability user data, and to an assisting node in the local-area layer for large amounts of high-rate user data. Scheduling is done separately in the anchor and assisting nodes. Thus, as there are separate schedulers for the two connections and the data on the two have no time-critical relation, the latency requirements for the interconnection between the layers are relaxed.

21.1.1.4 WLAN interworking

Wireless local area networks (WLANs) based on the 802.11 family of standards are already used by many operators for offloading cellular networks and embracing capacity expansion into license-exempt spectrum bands. The LTE specifications provide support for WLAN interworking, including seamless as well as non-seamless mobility, at the core network level.

The selection whether to use WLAN or 3GPP access currently resides in the terminal and is implementation-specific. Existing terminals have no information about the network load

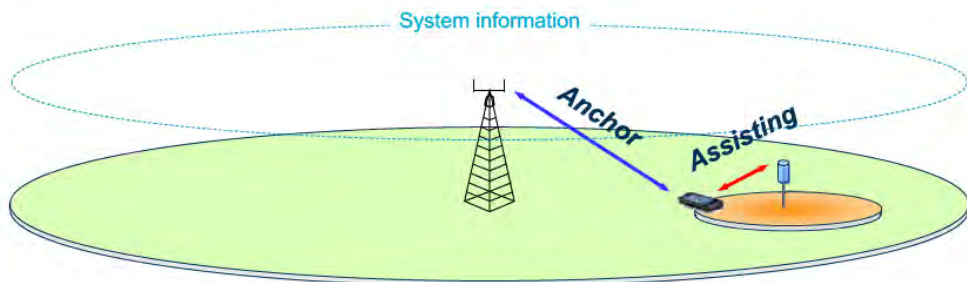


FIGURE 21.1

Dual connectivity

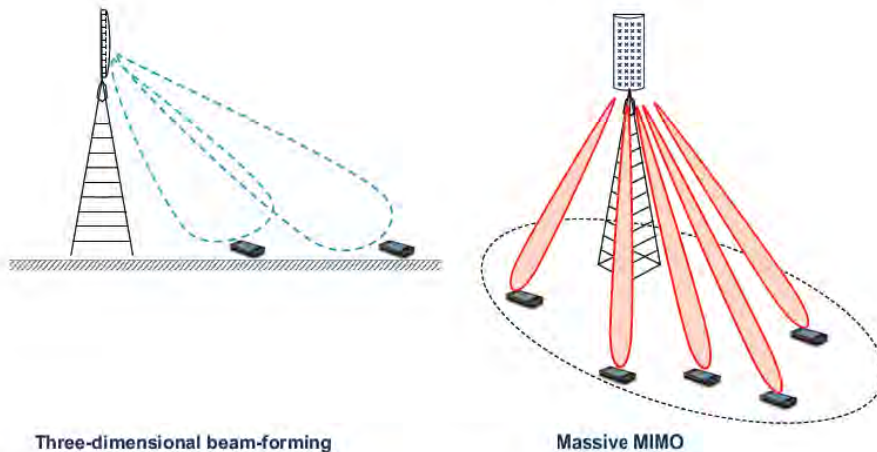
and typically connect to the WLAN whenever such a network is found, irrespective of which technology would be preferable from an end-user perspective in a particular scenario. Hence, existing terminals may switch to a WLAN even if it is heavily loaded and the LTE connection would provide significantly better end-user experience. Therefore, as part of the work on release 12, 3GPP will investigate signaling mechanisms to support radio-access-network control of the radio-access technology used by the terminal. An even tighter interaction between WLAN and LTE could also be envisioned in a later release, for example by using WLAN and LTE as the assisting and anchor connections, respectively, in a dual-connectivity scenario.

21.1.2 Enhanced multi-antenna/multi-point enhancements

Active-array-antenna systems are becoming increasingly common. Instead of having multiple passive antenna elements being fed by a single power amplifier, the RF components such as power amplifiers and transceivers are integrated with an array of antenna elements without increasing the overall size of the antenna enclosure. This approach provides several benefits compared to traditional deployments with passive antennas connected to transceivers through feeder cables. Not only are cable losses reduced, leading to improved performance and reduced energy consumption, but also the installation is simplified and the required equipment space is reduced. This can lead to multi-antenna deployments using the existing LTE functionality becoming increasingly common. However, to make multi-antenna techniques useful in practice, it is essential to ensure proper specification of relevant RF and electro-magnetic compatibility requirements. It is therefore important to continue the studies initiated in release 11 and to define these requirements as well as the corresponding test methodology as part of release 12. The spatial dimension is a key aspect of adaptive antenna systems and needs to be accounted for, which increases the complexity of the problem and possibly calls for some limited use of over-the-air testing. Prior to establishing the relevant performance requirements, the practical usefulness of defining more advanced features as discussed below is limited.

Multi-antenna and multi-point techniques in LTE will evolve not only in terms of defining RF requirements but also in terms of more advanced ways of exploiting multiple antennas. Examples of such concepts are three-dimensional beam-forming, massive MIMO and enhanced CoMP operation.

In three-dimensional beam-forming, beams are formed not only in the azimuth domain but also in the elevation domain, see [Figure 21.2](#). This can be exploited in several ways, including dynamic terminal-specific downtilt, multi-user MIMO, and vertical sectorization. To a large extent, LTE already includes support for three-dimensional beam-forming through non-codebook-based beam-forming as described in Chapter 10, but further enhancements can be considered based on performance evaluation in different scenarios. The performance of the different approaches is expected to depend heavily on channel and propagation models as well as on the actual deployment scenarios. Accurate channel modeling is therefore crucial for performance evaluations.

**FIGURE 21.2**

Examples of multi-antenna enhancements

Another more long-term example of advanced multi-antenna operation is massive MIMO where a large number of antenna elements are used to form very narrow beams. Massive MIMO is further discussed in Section 21.2.3.

Coordinated multipoint transmission/reception (CoMP) was one of the features introduced in release 11. As discussed in Chapter 15, the work in release 11 assumed that low-latency backhaul is available for the coordination between transmission points. However, due to the many open issues around CoMP, including the question of what schemes to use in what scenarios, it is expected that the work will continue in release 12. In particular, CoMP with relaxed backhaul requirements needs to be in focus to broaden the practical applicability of CoMP.

21.1.3 Machine-type communication

Up until today, mobile-communication technologies have been designed primarily with human-centric applications in mind. In case of LTE, the main focus on design was thus on developing techniques for providing high data rates and low latencies for mobile-broadband services such as file downloading and web browsing. However, as discussed in the introduction of this chapter, there is an expectation of a tremendous increase in the number of non-human-centric devices and associated “machine-type communication” (MTC). Actually, connected devices such as burglar alarms, power meters, and so on, are already a reality. Today, these devices are typically relying on second-generation GSM for connectivity, primarily due to its outstanding coverage and availability of very low-cost modems. However, there is a desire to at some stage move the main part of the machine communication to LTE.

Although LTE is capable of handling a wide range of machine-type communication scenarios already from the first release, several improvements are being discussed as part of the release-12 work, for example:

- Support low-cost and low-complexity device types to match low performance requirements (for example in peak data rates and delay) of certain MTC applications.
- Allow for very low energy consumptions for data transfers to ensure long battery life.
- Provide extended coverage options for MTC devices in challenging locations.
- Handle very large numbers of devices per cell.

Techniques to address low-cost MTC device types have been studied [115] by 3GPP and form the basis for the MTC enhancements in release 12. Of the techniques studied, the possibility for device-side half-duplex operation, reduced peak data rate requirements, and reduced bandwidth operation (less than 20 MHz) are particularly interesting. Additional simplifications may be achievable if single-antenna operation is allowed (in contrast to the current requirement on two-way receive diversity as a baseline) and if the maximum uplink transmit power is reduced. However, it has to be ensured that coverage and spectral efficiency are not significantly impacted.

In order to enable long battery life times, the energy consumptions of every data transfer of a device need to be reduced to a minimum. For devices with infrequent data transmission, energy consumption can be reduced if the DRX cycles (see Chapter 13) are significantly extended. This enables a device to make use of extended sleeping times while not transmitting data, which minimizes reading of control channels and mobility-related measurements. Furthermore, infrequent transmissions of small amounts of user data are typically associated with signaling procedures, for example to establish radio bearers. These signaling procedures are sometimes more expensive from a power-consumption perspective than the data transfer itself. Simplifying these procedures for infrequent small data transfers can therefore provide significant benefits.

In some use cases MTC devices may be located in challenging locations where LTE coverage is not available with existing network deployments. One example is smart meters in the basements of buildings. Options for coverage extensions are primarily repetition and energy accumulation of the appropriate signals, possibly complemented by beamforming.

Since the number of connected machines is expected to grow to very large numbers, mechanisms are needed to handle a large number of devices within a single cell. A load control scheme known as *enhanced access barring* has been standardized for release 11 to avoid RAN overload due to too many spontaneous access attempts. In general, signaling for every connected device can result in a very high control-plane load. Therefore, lightweight signaling procedures are desired which reduce the signaling load per device that is caused to the network. Note that a lightweight signaling scheme may at the same time improve device battery consumptions as described above.

21.1.4 Device-to-device communication

Direct device-to-device communications, sometimes called direct-mode communications, imply that wireless devices in the proximity of each other communicate in a peer-to-peer fashion without the user-plane involvement of a cellular infrastructure [116]. This has been studied in academia for some years, for example as part of the European research project WINNER [111]. Direct communication can, depending on the scenario, provide multiple benefits, both for the end-user in terms of very high data rates thanks to the proximity of the communicating devices, as well as for the overall system in terms of capacity gains from offloading traffic from the cellular infrastructure. Furthermore, communication may take place even in absence of infrastructure, which is an attractive property in public safety scenarios, for example when a first responder officer searches for people in need in a disaster situation without functioning infrastructure.

Direct device-to-device communications have so far mostly used unlicensed spectrum and operated independently of cellular infrastructure (lower left part of Figure 21.3). Examples are Bluetooth and WiFi Direct. However, recently there has been an increased interest in utilizing licensed (and in particular cellular) spectrum and more tightly integrated direct device-to-device communication with the cellular network (upper right part of Figure 21.3). In such a scheme, terminals in the proximity of each other can exchange information over a direct link rather than transmitting and receiving signals through a cellular base station [117,118]. The LTE network can, when present, assist with for example synchronization, peer discovery, and providing identity and security information. Preferably, the device-to-device

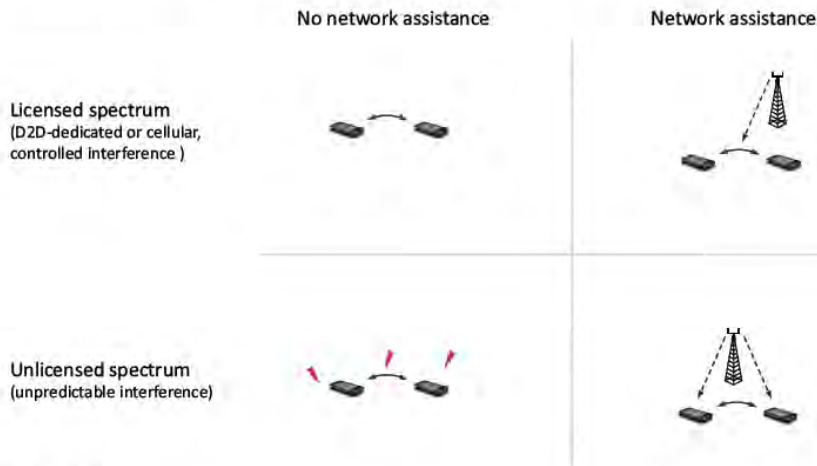


FIGURE 21.3

Categorization of device-to-device communication

communication scheme is designed such that the devices benefit from the network when present, but are still able to communicate, albeit possibly in a less efficient manner, in absence of network coverage.

Recognizing the potential advantages of direct device-to-device communication in a cellular network, 3GPP is currently investigating the feasibility of adding direct device-to-device communication to LTE. In particular, two use cases are considered:

- public safety, requiring communication between two devices both in presence and absence of network support, and
- peer discovery for commercial applications in presence of network support.

In both these cases, licensed spectrum will be used (upper part of [Figure 21.3](#)), primarily by allocating parts of the uplink resources for direct device-to-device usage.

Finally, it should be kept in mind that adding device-to-device communication to LTE impacts the overall architecture and is not a trivial task. For example, authorization and authentication, currently handled by the network, needs to be extended to handle the distributed situation occurring in a direct device-to-device scenario. Another example of an area requiring further thought is handling of direct communication between devices with subscriptions to different operators.

21.2 New radio access

This book has described the emergence of the LTE radio-access technology in 3GPP release 8 and the subsequent evolution of LTE through releases 9, 10, and 11. As discussed in the previous section, this evolution will continue, first with LTE release 12 and then onwards with future LTE releases, providing further enhanced system performance and further extended capabilities for the LTE radio-access technology. As part of this evolution, LTE will also continue to expand into new usage scenarios and use cases, such as local-area access and different types of “machine-type” communication.

However, at some stage completely new radio-access technology will undoubtedly emerge and initial discussions on such “beyond LTE” radio access have recently been initiated within the research community [119] as well as within ITU [120].

[Figure 21.4](#) illustrates the different “generations” of mobile communication, from 1st generation analog systems, via the 2nd and 3rd generation mobile-communication technologies, to 4th generation LTE. As the figure indicates, a new generation has entered the market roughly once every ten years. Solely based on this, one may anticipate that a new “beyond-LTE” radio-access solution may emerge and enter the market around 2020.

Up until now, each new generation of mobile communication has been designed to cover all use cases and all usage scenarios of earlier generations, including low to high mobility, wide-area coverage to very dense deployments, voice *and* data, and so on. One possible scenario is for this to continue beyond LTE.

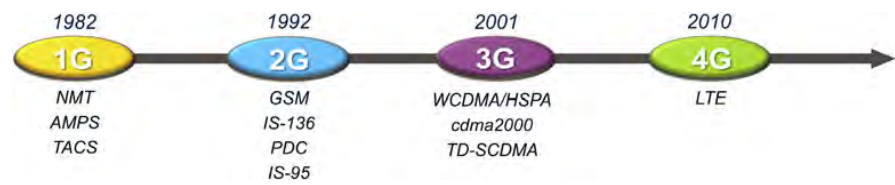


FIGURE 21.4
Generations of mobile communication

However, one could argue that for certain use cases and usage scenarios, for example wide-area mobile broadband, it is not likely that major additional benefits can be achieved with a completely new radio-access technology, compared to what can be achieved with the occurring evolution of LTE. It can also be argued that developing a single new radio-access technology targeting the full range of anticipated use cases and usage scenarios unavoidably leads to design compromises and partly sub-optimal solutions.

With this in mind one can envision an alternative future scenario as outlined in Figure 21.5. In this scenario, the overall radio-access solution of the future will consist of *a set of well-integrated radio-access technologies*, including the evolution of existing radio access and *complementary* new radio-access technologies targeting specific use cases and/or usage scenarios where more substantial benefits can be expected from introducing new technology, compared to what can be achieved with the evolution of existing radio access.

Note that in this future scenario the term “5G” would not refer to a single new radio-access technology. Rather, as illustrated in Figure 21.5, the term 5G would cover the *overall future radio-access solution* consisting of the *evolution of existing radio-access technologies* and *new complementary technologies*, all operating in a well-integrated way. Also note that the complementary new radio-access technologies could, in this scenario, very well consist of *multiple* technologies, each targeting and optimized towards a specific (group of) use case(s) and/or usage scenario(s).

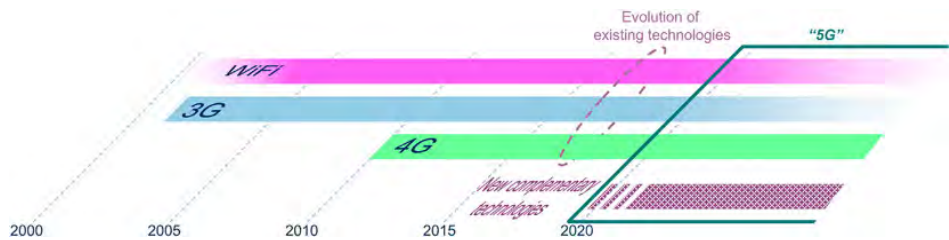


FIGURE 21.5
Future radio access

In the following, some key technology areas relevant for such beyond-LTE radio access are discussed.

21.2.1 New frequency ranges

The availability of sufficient amount of spectrum is key to future radio communication. Current mobile-communication systems operate in spectrum up to roughly 3.5 GHz. In order to further expand the overall traffic capacity of the mobile-communication networks one should extend the amount of available spectrum by considering new frequency ranges.

In the medium-time perspective, up until roughly 2020, it can be expected that existing mobile-communication technologies, primarily LTE, will expand their frequency-range of operation into higher bands but still below roughly 10 GHz. However, for the longer-term perspective—beyond 2020—one should consider further expansion into even higher frequency bands up to roughly 100 GHz and perhaps even beyond that.

The use of such very high frequencies for mobile communication would have the potential of opening up for a massive increase in the amount of available spectrum. It would also open up for the possibility for much wider transmission bandwidths in the order of several 100 MHz, or perhaps as wide as 1 GHz or even wider. Operation at such very high frequencies could also be an enabler for new technology solutions, such as the use of massive antenna configurations as further discussed below.

Still, there are many remaining questions regarding the use of such high frequencies for mobile communication, especially related to propagation. Although the higher path loss at higher frequencies can partly be compensated for by means of more advanced antenna solutions providing higher antenna gains, there are still question marks regarding issues such as the impact of shadowing and the possibility for outdoor-to-indoor propagation.

Despite this, the use of very high frequencies for mobile communication will undoubtedly be one of the more interesting possible evolution paths for future radio access.

21.2.2 New ways of assigning spectrum

There are currently two main approaches to spectrum assignment for wireless communication.

Cellular technologies operate in *dedicated licensed spectrum*. The local assignment, for example on a country basis, of a certain frequency band for exclusive use by a single operator, together with the strict specification of radio-transmission performance for that band, allows for high service quality and high system efficiency at high network loads. However, in very dense network deployments, which may constitute a substantial part of future wireless-communication networks, the traffic load per network node may often be relatively low. To assign a certain part of the spectrum for use by one specific operator may, in that case, not be the most efficient approach in terms of system efficiency. It may be more efficient if spectrum momentarily not used by one operator could be used by another operator.

In contrast, WiFi technologies operate in *unlicensed spectrum* with, often, multiple independent networks operating simultaneously and with no coordination within the same area. The possibility for simultaneous operation of multiple networks in the same spectrum may allow for more efficient overall spectrum utilization at low traffic loads, but makes it difficult to retain good service quality and high efficiency as the traffic load increases.

In the future one should also consider other means to assign spectrum, with an aim to maintain the high-quality service level of dedicated spectrum and, at the same time, enable high efficiency in terms of spectrum utilization regardless of the load.

Already today there are ongoing discussions on *Licensed Shared Access (LSA)* [122] as a means to increase the spectrum availability for mobile communication. In the context of mobile communication, LSA implies that spectrum assigned on a primary basis for a certain system can be used on a secondary basis for mobile communication as long as this secondary usage does not negatively impact the primary system. One typical case could be spectrum that is used only very infrequently and in a highly predictable manner by the primary system.

In the future one should also consider other means for spectrum sharing, including, for example, assigning certain spectrum jointly to a set of operators and providing tools as part of the radio-access technology to improve the performance of such sharing.

It should be noted that such new ways of assigning spectrum, including different means for spectrum sharing, should not be seen as a replacement for the currently dominating paradigm of licensed dedicated spectrum. Rather, different means for sharing spectrum should be considered a *complement* to dedicated licensed spectrum in order to further extend the overall amount of spectrum available for mobile communication.

21.2.3 Massive antenna configurations

It should be clear from earlier chapters that the use of different forms of multi-antenna transmission/reception is one important component of LTE. In the future it can be expected that advanced multi-antenna technologies will be of even higher importance, being a tool for further increase in the achievable data rates and even higher traffic capacity.

The introduction of *massive antenna configurations* with the number of controllable antenna elements increasing by orders of magnitude compared to today is one possible track. Proper use of such antenna configurations could be used to dramatically boost system capacity and achievable data rates, without requiring a corresponding densification of the network infrastructure.

Massive antenna configurations are especially of interest for operation in much higher frequency bands above 10 GHz (see also [Section 21.2.1](#)). Due to the smaller wave length at higher frequencies, truly massive antenna configurations consisting of a very large number of antenna elements can then be deployed without an unreasonably large physical size of the antenna configuration. For example, at 40 GHz an antenna array consisting of 512 cross-polarized elements with half-wavelength spacing could be packed into a roughly 10×10 cm area.

21.2.4 Ultra-dense deployments

Network densification is an ongoing process for currently existing cellular systems with inter-site distances as small as 200 m and even lower not being uncommon today. The interest in network densification is also reflected in the increased focus on heterogeneous deployments within 3GPP, a topic that was extensively discussed in Chapter 16.

The paradigm of *ultra-dense deployments* takes network densification even further, with network densities that may be orders of magnitude higher than the densest networks of today. Such deployments could, for example, be of interest in scenarios where extremely high data rates and/or extreme traffic capacity is required, for example in office environments and public places such as shopping malls, subway stations, or other dense urban environments with a very large number of users with very high individual traffic demands.

Ultra-dense deployments can primarily be anticipated to operate at high frequencies beyond 10 GHz. The primary motivation for the use of such high frequency ranges for ultra-dense deployments would be the possibility for large spectrum blocks allowing for large transmission bandwidths, something that would be needed to fully exploit the data-rate potential of ultra-dense deployments.

A key technical challenge of ultra-dense deployments is to ensure low-cost deployment and management despite the very large number of network nodes. Thus, support for different means for network *self-configuration* and *self-optimization* would be an important component for ultra-dense deployments.

Backhauling would be another key challenge for ultra-dense deployments. The possibility for wireless backhauling, including different means for multi-hop communication, should preferably be an integral part of future ultra-dense deployments. This could also include different forms of direct device-to-device communications.

21.3 Final thoughts

LTE has already emerged as the globally-accepted radio-access-technology of choice for providing mobile-broadband services for the next decade and beyond. The evolution of LTE will continue, further enhancing the LTE capabilities and extending it into a key enabler for new types of applications such as massive communication between machines. Further into the future new radio-access technologies will emerge complementing LTE for specific use cases and usage scenarios. Still, LTE will remain as a key component bringing us into the future networked society, where everyone and everything will be connected, and where there will be non-limiting access to information and possibility for data sharing anywhere and anytime for anyone and anything.

References

- [1] ITU-R, International mobile telecommunications-2000 (IMT-2000), Recommendation ITU-R M.687 2, February 1997.
- [2] ITU-R, Detailed specifications of the radio interfaces of international mobile telecommunications-2000 (IMT-2000), Recommendation ITU-R M.1457 11, February 2013.
- [3] ITU-R, Principles for the process of development of IMT-advanced. Resolution ITU-R 57, October 2007.
- [4] ITU-R, Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000, Recommendation ITU-R M.1645, June 2003.
- [5] ITU-R, Invitation for submission of proposals for candidate radio interface technologies for the terrestrial components of the radio interface(s) for IMT-advanced and invitation to participate in their subsequent evaluation, ITU-R SG5, Circular Letter 5/LCCE/2, March 2008.
- [6] ITU-R, ITU paves way for next-generation 4G mobile technologies; ITU-R IMT-advanced 4G standards to usher new era of mobile broadband communications, ITU Press Release, 21 October 2010.
- [7] ITU-R WP5D, Recommendation ITU-R M.2012. Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications Advanced (IMT-Advanced), January 2012.
- [8] ITU-R, Frequency arrangements for implementation of the terrestrial component of international mobile telecommunications-2000 (IMT-2000) in the bands 806 960 MHz, 1710 2025 MHz, 2110 2200 MHz and 2500 2690 MHz, Recommendation ITU-R M.1036-4, March 2012.
- [9] M. Olsson, S. Sultana, S. Rommer, L. Frid, C. Mulligan, SAE and the Evolved Packet Core—Driving the Mobile Broadband Revolution, Academic Press, 2009.
- [10] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN) (Release 7), 3GPP TR 25.913.
- [11] C.E. Shannon, A mathematical theory of communication, Bell System Tech. J 27 (July and October 1948) 379 423, 623 656.
- [12] J.G. Proakis, Digital Communications, McGraw-Hill, New York, 2001.
- [13] G. Bottomley, T. Ottosson, Y.-P. Eric Wang, A generalized RAKE receiver for interference suppression, IEEE J. Sel. Area Comm. 18 (8) (August 2000) 1536 1545.
- [14] WiMAX Forum, Mobile WiMAX—Part I: A technical overview and performance evaluation, White Paper, August 2006.
- [15] ETSI, Digital video broadcasting (DVB): Framing structure, channel coding and modulation for digital terrestrial television, ETS EN 300 744 v. 1.1.2.
- [16] NTT DoCoMo, et al., Views on OFDM parameter set for evolved UTRA downlink, Tdoc R1-050386, 3GPP TSG RAN WG 1, Athens, Greece, 9 13 May 2005.
- [17] L. Hanzo, M. Munster, B.J. Choi, and T. Keller, OFDM and MC CDMA for broadband multiuser communications, WLANs, and broadcasting, Wiley IEEE Press, Chichester, UK, ISBN 0-48085879.
- [18] R. van Nee, R. Prasad, OFDM for Wireless Multimedia Communications, Artech House Publishers, London, January 2000.
- [19] J. Tellado and J.M. Cioffi, PAR reduction in multi-carrier transmission systems, ANSI T1E1.4/97 367.
- [20] P.V. Etvalt, Peak to average power reduction for OFDM schemes by selective scrambling, Electron. Lett. 32 (21) (October 1996) 1963 1964.
- [21] W. Zirwas, Single frequency network concepts for cellular OFDM radio systems, International OFDM Workshop, Hamburg, Germany, September 2000.
- [22] A. Oppenheim, R.W. Schaffer, Digital Signal Processing, Prentice-Hall International, ISBN 0-13-214107-8 01.
- [23] S. Haykin, Adaptive Filter Theory, Prentice-Hall International, NJ, USA, 1986. ISBN 0-13-004052-5 025.

- [24] D. Falconer, S.L. Ariyavisitakul, A. Benyamin-Seeyar, B. Eidson, Frequency domain equalization for single-carrier broadband wireless systems, *IEEE Commun. Mag.* 40 (4) (April 2002) 58–66.
- [25] G. Forney, Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference, *IEEE T. Inform. Theory* IT-18 (May 1972) 363–378.
- [26] G. Forney, The viterbi algorithm, *Proceedings of the IEEE*, Piscataway, NJ, USA, Vol. 61, No. 3, March, 1973, 268–278.
- [27] Motorola, Comparison of PAR and Cubic Metric for Power De-rating, Tdoc R1-040642, 3GPP TSG-RAN WG1, May 2004.
- [28] U. Sorger, I. De Broeck, M. Schnell, IFDMA – A new spread-spectrum multiple-access scheme. *Proceedings of the ICC'98*, Atlanta, GA, USA, June 1998, pp. 1267–1272.
- [29] W. Lee, *Mobile Communications Engineering*, McGraw-Hill, New York, USA, ISBN 0-07-037039-7.
- [30] J. Karlsson, J. Heinegard, Interference rejection combining for GSM, *Proceedings of the 5th IEEE International Conference on Universal Personal Communications*, Cambridge, MA, USA, 1996, pp. 433–437.
- [31] L.C. Godara, Applications of antenna arrays to mobile communications, Part I: Beam-forming and direction-of-arrival considerations, *Proceedings of the IEEE*, Vol. 85, No. 7, July, 1997, 1029–1030.
- [32] L.C. Godara, Applications of antenna arrays to mobile communications, Part II: Beam-forming and direction-of-arrival considerations, *Proceedings of the IEEE*, Piscataway, Vol. 85, No. 7, July, 1997, 1031–1060.
- [33] A. Huebner, F. Schuehlein, M. Bossert, E. Costa, H. Haas, A simple space frequency coding scheme with cyclic delay diversity for OFDM, *Proceedings of the 5th European Personal Mobile Communications Conference*, Glasgow, Scotland, April 2000, pp. 106–110.
- [34] 3GPP, 3rd generation partnership project; technical specification group radio access network; Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD), 3GP TS 25.211.
- [35] V. Tarokh, N. Seshadri, A. Calderbank, Space time block codes from orthogonal design, *IEEE T. Inform. Theory* 45 (5) (July 1999) 1456–1467.
- [36] A. Hottinen, O. Tirkkonen, and R. Wichman, *Multi-Antenna Transceiver Techniques for 3G and Beyond*, John Wiley, Chichester, UK, ISBN 0470 84542 2.
- [37] C. Kambiz, L. Krasny, Capacity-achieving transmitter and receiver pairs for dispersive MISO channels, *IEEE T. Commun.* 42 (April 1994) 1431–1440.
- [38] R. Horn, C. Johnson, *Matrix analysis*, *Proceedings of the 36th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, November 2002.
- [39] K.J. Kim, Channel estimation and data detection algorithms for MIMO OFDM systems, *Proceedings of the 36th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, November 2002.
- [40] M.K. Varanasi, T. Guess, Optimum decision feedback multi-user equalization with successive decoding achieves the total capacity of the gaussian multiple-access channel, *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, Monterey, CA, November 1997.
- [41] S. Grant, J.F. Cheng, L. Krasny, K. Molnar, Y.-P.E. Wang, Per-antenna-rate-control (PARC) in frequency selective fading with SIC-GRACE receiver, *60th IEEE Vehicular Technology Conference*, Los Angeles, CA, USA, September 2004, 1458–1462.
- [42] S.T. Chung, A.J. Goldsmith, Degrees of freedom in adaptive modulation: A unified view, *IEEE T. Commun.* 49 (9) (September 2001) 1561–1571.
- [43] A.J. Goldsmith, P. Varaiya, Capacity of fading channels with channel side information, *IEEE T. Inform. Theory* 43 (November 1997) 1986–1992.
- [44] R. Knopp, P.A. Humblet, Information capacity and power control in single-cell multi-user communications, *Proceedings of the IEEE International Conference on Communications*, Seattle, WA, USA, Vol. 1, 1995, 331–335.

- [45] D. Tse, Optimal power allocation over parallel gaussian broadcast channels, Proceedings of the International Symposium on Information Theory, Ulm, Germany, June 1997, p. 7.
- [46] M.L. Honig and U. Madhow, Hybrid intra-cell TDMA/inter-cell CDMA with inter-cell interference suppression for wireless networks, Proceedings of the IEEE Vehicular Technology Conference, Secaucus, NJ, USA, 1993, pp. 309–312.
- [47] S. Ramakrishna, J.M. Holtzman, A scheme for throughput maximization in a dual-class CDMA system, IEEE J. Sel. Area Comm. 16 (6) (1998) 830–844.
- [48] S.-J. Oh, K.M. Wasserman, Optimality of greedy power control and variable spreading gain in multi-class CDMA mobile networks, Proceedings of the AMC/IEEE Mobi Comp, Seattle, WA, USA, 1999. pp. 102–112.
- [49] J.M. Holtzman, CDMA forward link waterfilling power control, Proceedings of the IEEE Vehicular Technology Conference, Tokyo, Japan, Vol. 3, May, 2000, 1663–1667.
- [50] J.M. Holtzman, Asymptotic analysis of proportional fair algorithm, Proceedings of the IEEE Conference on Personal Indoor and Mobile Radio Communications, San Diego, CA, USA, Vol. 2, 2001, 33–37.
- [51] P. Viswanath, D. Tse, R. Laroia, Opportunistic Beam-forming using dumb antennas, IEEE T. Inform. Theory 48 (6) (2002) 1277–1294.
- [52] S. Lin and D. Costello, Error Control Coding, Prentice-Hall, Upper Saddle River, NJ, USA, June 2004.
- [53] C. Schlegel, Trellis and Turbo Coding, Wiley IEEE Press, Chichester, UK, March 2004.
- [54] J.M. Wozencraft, M. Horstein, Digitalised Communication Over Two-way Channels, Fourth London Symposium on Information Theory, London, UK, September 1960.
- [55] D. Chase, Code combining—a maximum-likelihood decoding approach for combining and arbitrary number of noisy packets, IEEE T. Commun. 33 (May 1985) 385–393.
- [56] M.B. Pursley, S.D. Sandberg, Incremental-redundancy transmission for meteor-burst communications, IEEE T. Commun. 39 (May 1991) 689–702.
- [57] S.B. Wicker, M. Bartz, Type-I hybrid ARQ protocols using punctured MDS codes, IEEE T. Commun. 42 (April 1994) 1431–1440.
- [58] J.-F. Cheng, Coding performance of hybrid ARQ schemes, IEEE T. Commun. 54 (June 2006) 1017–1029.
- [59] P. Frenger, S. Parkvall, and E. Dahlman, Performance comparison of HARQ with chase combining and incremental redundancy for HSDPA, Proceedings of the IEEE Vehicular Technology Conference, Atlantic City, NJ, USA, October 2001, pp. 1829–1833.
- [60] J. Hagenauer, Rate-compatible punctured convolutional codes (RCPC codes) and their applications, IEEE T. Commun. 36 (April 1988) 389–400.
- [61] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Physical Channels and Modulation (Release 8), 3GPP TS 36.211.
- [62] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Multiplexing and Channel Coding (Release 8), 3GPP TS 36.212.
- [63] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Physical Layer Procedures (Release 8), 3GPP TS 36.213.
- [64] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Physical Layer—Measurements (Release 8), 3GPP TS 36.214.
- [65] ITU-R, Requirements related to technical performance for IMT-Advanced radio interface(s), Report ITU-R M.2134, 2008.
- [66] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced) (Release 9), 3GPP TR 36.913.
- [67] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Evolved universal terrestrial radio access (E-UTRA); User Equipment (UE) Radio Access Capabilities, 3GPP TS 36.306.

- [68] IETF, Robust header compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and Uncompressed, RFC 3095.
- [69] J. Sun, O.Y. Takeshita, Interleavers for turbo codes using permutation polynomials over integer rings, *IEEE T. Inform. Theory* 51 (1) (January 2005) 101–119.
- [70] O.Y. Takeshita, On maximum contention-free interleavers and permutation polynomials over integer rings, *IEEE T. Inform. Theory* 52 (3) (March 2006) 1249–1253.
- [71] D.C. Chu, Polyphase codes with good periodic correlation properties, *IEEE T. Inform. Theory* 18 (4) (July 1972) 531–532.
- [72] J. Padhye, V. Firoiu, D.F. Towsley, J.F. Kurose, Modelling TCP reno performance: A simple model and its empirical validation, *ACM/IEEE T. Network.* 8 (2) (2000) 133–145.
- [73] CRAN International Workshop, CRANdRoad Towards Green Radio Access Network, China Mobile, 23 April 2010.
- [74] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); User equipment (UE) conformance specification; Radio Transmission and Reception (Part 1, 2 and 3), 3GPP TS 36.521.
- [75] 3GPP, Evolved universal terrestrial radio access (E-UTRA); Physical layer for relaying operation, 3GPP TS 36.216.
- [76] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Evolved universal terrestrial radio access (E-UTRA); User Equipment (UE) radio transmission and reception, 3GPP TS 36.101.
- [77] APT Wireless Forum, APT Report on Harmonised Frequency Arrangements for the Band 698–806 MHz, APT/AWF/REP-14, September 2010 Edition.
- [78] 3GPP, 3rd generation partnership project; Technical specification group radio access network; UMTS-LTE 3500 MHz Work Item Technical Report (Release 10), 3GPP TR 37.801.
- [79] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Feasibility Study for Evolved Universal Terrestrial Radio Access (UTRA) and Universal Terrestrial Radio Access Network (UTRAN) (Release 7), 3GPP TR 25.912.
- [80] 3GPP, E-UTRA, UTRA and GSM/EDGE; Multi-Standard Radio (MSR) Base Station (BS) Radio Transmission and Reception, 3GPP TR 37.104.
- [81] 3GPP, E-UTRA, UTRA and GSM/EDGE; Multi-Standard Radio (MSR) Base Station (BS) Conformance Testing, 3GPP TR 37.141.
- [82] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Evolved universal terrestrial radio access (E-UTRA); User Equipment (UE) Radio Transmission and Reception, 3GPP TR 36.803.
- [83] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Evolved universal terrestrial radio access (E-UTRA); Base Station (BS) Radio Transmission and Reception, 3GPP TR 36.804.
- [84] 3GPP, Evolved universal terrestrial radio access (E-UTRA); User Equipment (UE) Radio transmission and reception, 3GPP TR 36.807.
- [85] 3GPP, Evolved universal terrestrial radio access (E-UTRA); Carrier Aggregation Base Station (BS) Radio transmission and reception, 3GPP TR 36.808.
- [86] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Evolved universal terrestrial radio access (E-UTRA); Base Station (BS) Radio transmission and reception, 3GPP TS 36.104.
- [87] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Evolved universal terrestrial radio access (E-UTRA); Base Station (BS) conformance testing, 3GPP TS 36.141.

- [88] FCC, Title 47 of the Code of Federal Regulations (CFR), Federal Communications Commission.
- [89] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Evolved universal terrestrial radio access (E-UTRA); Radio Frequency (RF) system scenarios, 3GPP TR 36.942.
- [90] ITU-R, Unwanted Emissions in the Spurious Domain, Recommendation ITU-R SM.329 10, February 2003.
- [91] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Radio Frequency (RF) system scenarios, 3GPP TR 25.942.
- [92] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Feasibility study for further advancements for E-UTRA (LTE-Advanced) (Release 9), 3GPP TR 36.912.
- [93] ITU-R, Guidelines for Evaluation of Radio Interface Technologies for IMT-Advanced, Report ITU-R M.2135 1, December 2009.
- [94] A. Furuskär, Performance evaluations of LTE-advanced—the 3GPP ITU proposal, 12th International Symposium on Wireless Personal Multimedia Communications (WPMC 2009), Invited Paper.
- [95] E. Dahlman, H. Ekström, A. Furuskär, Y. Jading, J. Karlsson, M. Lundevall, S. Parkvall, The 3G long-term evolution—radio interface concepts and performance evaluation, 63rd Vehicular Technology Conference, VTC 2006—Spring, Vol. 1, pp. 137–141, IEEE, 2006.
- [96] E. Dahlman, S. Parkvall, J. Sköld, P. Beming, 3G Evolution—HSPA and LTE for Mobile Broadband, second ed., Academic Press, 2008.
- [97] WINNER, IST-WINNER II Deliverable 1.1.2 v. 1.2. WINNER II Channel Models, IST-WINNER2, Tech. Rep., 2007.
- [98] 3GPP, 3rd generation partnership project; Technical specification group radio access network; Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations (Release 9), 3GPP TR 25.996.
- [99] Alcatel Shanghai Bell, CATT, CMCC, RITT, Spectrum Communications, TD-TECH, ZTE, Introduce TD-SCDMA Industry Standard in CCSA to 3GPP, Document R4-071394, 3GPP TSG-RAN WG4 meeting #44, Athens, Greece, August 2007.
- [100] 3GPP, 3rd generation partnership project; Technical specification group GSM/EDGE radio access network; Feasibility Study for Evolved GSM/EDGE Radio Access Network (GERAN) (Release 7), 3GPP TR 45.912.
- [101] H. Axelsson, P. Bjärkén, P. de Bruin, S. Eriksson, H. Persson, GSM/EDGE continued evolution, Ericsson Rev. (01) (2006) 20—29 Telefonaktiebolaget LM Ericsson, Stockholm, Sweden.
- [102] 3GPP, Updated New WID on Higher Uplink Performance for GERAN Evolution (HUGE), Tdoc GP-061901, 3GPP TSG GERAN #31, Denver, USA, 4–8 September 2006.
- [103] 3GPP, Circuit Switched Voice Capacity Evolution for GSM/EDGE Radio Access Network (GERAN), 3GPP TR 45.914.
- [104] N. Bhushan, C. Lott, P. Black, R. Attar, Y.-C. Jou, M. Fan, et al., CDMA2000 1xEV-DO revision A: A physical layer and MAC layer overview, IEEE Commun. Mag. (February 2006) 75–87.
- [105] M.W. Thelander, The 3G Evolution: Taking CDMA2000 into the Next Decade, Signal Research Group, LLC, White Paper developed for the CDMA Development Group, (October 2005).
- [106] R. Attar, Evolution of CDMA2000 cellular networks: Multicarrier EV-DO, IEEE Commun. Mag. (March 2006) 46–53.
- [107] M. Alendal, Operators need an ecosystem to support 50 billion connections, Ericsson Bus. Rev. (3) (2010).
- [108] M. Nekovee, A survey of cognitive radio access to TV white spaces, Int. J. Digi. Multimed. Broadcast. (2010) 1–11.
- [109] J. Mitola, Cognitive Radio—An Integrated Agent Architecture for Software Defined Radio, Royal Institute of Technology (KTH), Sweden, (May 2000).
- [110] S. Haykin, Cognitive radio: Brain-empowered wireless communications, IEEE J. Sel. Area. Comm. 23 (2) (February 2005) 201–220.

- [111] Wireless World Initiative New Radio, Eurescom, 2006, <https://www.ist-winner.org>.
- [112] Ericsson, “Ericsson Mobility Report,” June 2013, <http://www.ericsson.com/res/docs/2013/ericsson-mobility-report-june-2013.pdf>.
- [113] 3GPP, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Relay radio transmission and reception,” 3GPP TS 36.116.
- [114] 3GPP RWS-120045, “Summary of 3GPP TSG-RAN Workshop on Release 12 and Onward,” http://3gpp.org/ftp/workshop/2012-06-11_12_RAN_REL12/Docs/RWS-120045.zip.
- [115] 3GPP TR 36.888, “Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE”
- [116] Fitzek FHP. “Cellular Controlled Short-Range Communication for Cooperative P2P Networking,” Wireless World Research Forum (WWRF) 17 WG5, 2010.
- [117] Doppler K, Rinne M, Wijting C, Riberio CB, Hugl K. Device-to-Device Communication as an Underlay to LTE-Advanced Networks,” IEEE Communications Magazine (2009), Vol. 7(No. 12):42–9.
- [118] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklós, Z. Turányi, Design Aspects of Network Assisted D2D Communications, IEEE Communications Magazine (March 2012), 180–177.
- [119] METIS project, [METIS] METIS project, <https://www.metis2020.com/>.
- [120] International Telecommunication Union, “Framework and overall objectives of future development of IMT for 2020 and beyond”
- [121] International Telecommunication Union, “Future technology trends of terrestrial IMT systems”
- [122] J. Khun-Jush, P. Bender, B. Deschamps, and M. Gundlach, “Licensed shared access as complementary approach to meet spectrum demands: Benefits for next generation cellular systems”, ETSI Workshop on Reconfigurable Radio Systems, Cannes, France, 2012.
- [123] METIS project, “Scenarios, Requirements and KPIs for 5G Mobile and Wireless System”
- [124] Ericsson, “5G Radio Access—Research and Vision,” Ericsson White Paper 284 23 3204, June 2013, <http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf>

Index

Note: Page numbers with “f” denote figures; “t” tables.

0–9

16QAM modulation, 21–24
3G, 3–14
3D beam forming. *See* three-dimensional beam-forming
3GPP, 12–14
4G, 14–15
5G, 495
64QAM modulation, 21–24

A

Access Stratum (AS), 122
ACIR, 461–462
Acknowledged mode (for RLC), 316
Acknowledgement Resource Indicator (ARI), 227, 279
Acknowledgement Resource Offset (ARO), 228, 278–279
ACLR, 461–463
ACS, 461–462
Adaptive Modulation and Coding (AMC), 87
Additional Maximum Power Reduction (A-MPR), 454, 457
Advanced Receivers, 117
Almost Blank Subframes (ABS), 395
A-MPR. *See* Additional Maximum Power Reduction
Antenna port, 147–148
Automatic Repeat Request (ARQ), 97–101
 in LTE, 299–314
 for LTE relaying, 418–421

B

Band categories, 449
Bandwidth utilization, 18
Base-station classes, 455–456
BCCH. *See* Broadcast Control Channel
BCH. *See* Broadcast Channel,
 transmission of, 351–354
Beam forming:
 classical, 73–75
 in LTE. *See* Pre-coding, LTE
Blind decoding of PDCCH/EPDCCH, 218–223
Blocking, receiver, 465–466
 narrowband, 466
 for Multi-band base stations, 470
 for relays, 471
BM-SC. *See* Broadcast Multicast Service Center
Broadcast, 46–48. *See also* MBMS
Broadcast Control Channel (BCCH), 128

Broadcast Channel (BCH), 129
Broadcast Multicast Service Center (BM-SC), 405
Buffer status report, 327–328

C

Capability Set (CS), 449–450
Carrier aggregation:
 control signaling for, 237–240
 overview, 113, 117, 154–158
 RF characteristics, 441–445
Carrier Indicator, 226, 234
Carrier raster, 158–159
CCCH. *See* Common Control Channel
CCE. *See* Control Channel Element
CDD. *See* Cyclic Delay Diversity
Cell identity, 347–348
Cell-identity group, 348
Cell search, 347–351
Cell-specific reference signal. *See* reference-signal,
 cell-specific
Channel capacity, 17
Channel coding, LTE:
 downlink, 164–165
 uplink, 242
Channel rank, 190
Channel State Information (CSI), 106, 267,
 337–346
Channel-state report, 106, 137, 340
Channel-Quality Indicator (CQI), 286, 337–338
Chase combining, 98–101
Closed Subscriber Group (CSG), 401
CN. *See* Core Network
Code block, 163
Coexistence:
 requirements, 438
 with TD-SCDMA, 153–154
Co-location requirements, 438
Common Control Channel (CCCH), 128
Common Subframe Allocation (CSA), 409–410
CoMP. *See* Coordinated Multipoint transmission
Component carrier, 155–158
Contention resolution, 360, 368–369
Control-channel element, 210
Control region, 145, 199
Coordinated beam-forming, 380

Coordinated link adaptation, 376–379
 Coordinated power control, 380
 Coordinated scheduling, 380
 Coordinated multipoint transmission, 371
 Core Network (CN), 121–122
 CQI. *See* Channel-Quality Indicator
 CRC. *See* Cyclic Redundancy Check
 C-RNTI, 138
 in relation to random access, 367–369
 Cross-carrier scheduling, 237–240, 313, 330–332
 CSG. *See* Closed Subscriber Group
 CSI. *See* Channel-State Information
 CSI-RS. *See* reference signal, for channel-state information
 Zero-power, 183
 CSI-IM, 344
 Cubic metric, 59
 Cyclic-Delay Diversity (CDD), 71
 Cyclic prefix:
 extended, 143
 normal, 143
 OFDM, 34–37, 43
 Random-access preamble, 362–365
 single-frequency network, 43
 Cyclic Redundancy Check (CRC), 97
 for BCH, 352
 for EPDCCH, 214
 for DL-SCH, 163
 for PDCCH, 209
 for UL-SCH, 242

D

Data region, 145, 199
 DC-subcarrier, 145
 DCCH. *See* Dedicated Control Channel
 DCI. *See* Downlink Control Information
 DCI format, 207–208
 blind decoding of, 218–223
 downlink related, 223–225
 uplink related, 233–234
 Dedicated Control Channel (DCCH), 128
 Dedicated Traffic Channel (DTCH), 129
 Delay diversity, 70–71
 Demodulation reference signal. *See* reference-signal
 demodulation
 Device-to-device communication, 493–494
 DFT-spread OFDM (DFTS-OFDM), 56–62
 localized vs. distributed, 61–62
 receiver, 59–60
 user multiplexing, 60–61
 in LTE, 105–106, 243
 Discontinuous Reception (DRX), 335–337
 DL-SCH. *See* Downlink Shared Channel

Donor cell, 414
 Downlink Control Information (DCI), 136, 207
 Downlink Shared Channel (DL-SCH), 129
 processing, 161–171
 DRX. *See* Discontinuous Reception
 DTCH. *See* Dedicated Traffic Channel
 DTX. *See* Discontinuous Transmission
 Dual connectivity, 488–489
 Dynamic TDD. *See* TDD dynamic
 DwPTS, 151–153
 Dynamic Point Selection, 381
 Dynamic range requirement, 465

E

ECCE. *See* Enhanced Control Channel Element
 eNB. *See* eNodeB
 Enhanced Control Channel Element, 215
 Enhanced Physical Downlink Control Channel, 136, 200, 213
 DM-RS for, 179
 Enhanced Resource Element Group, 215
 eNodeB, 123
 EPC. *See* Evolved Packet Core
 EPDCCH. *See* Enhanced Physical Downlink Control Channel
 EPS. *See* Evolved Packet System
 Equalization, 26, 49–55
 Decision-feedback, 54
 Frequency-domain, 52–54
 Time-domain, 49–51
 EREG. *See* Enhanced Resource Element Group
 EVM requirement, 457–458
 Evolved Packet Core, 121–123
 Evolved Packet System, 121

F

FDD. *See* Frequency Division Duplex
 Femto base station. *See* Home-eNodeB
 Frame, LTE, 142
 Frequency bands, 431–438
 numbering of, 433–434
 Frequency Division Duplex,
 Half-duplex. *See* Half-duplex FDD
 in LTE, 109–110, 149–150
 spectrum, 4, 8
 Frequency hopping, uplink, 245–248
 Frequency-Switched Transmit Diversity, 187–188
 FSTD. *See* Frequency-Switched Transmit Diversity

G

GSM, 3
 migration from, 38
 multi-standard radio base-station for, 447–450
 Guard period for TDD, 152

H

Half-duplex FDD, 149–150
 in relation to scheduling, 333–335
 HARQ. *See* Hybrid ARQ
 Heterogeneous network deployments, 114–115, 387–401
 Higher-order modulation, 21–24
 High Interference Indicator (HII), 374
 High Speed Packet Access (HSPA), 1, 9, 14
 HII. *See* High Interference Indicator
 Home-eNodeB, 401
 Home Subscriber Service (HSS), 112
 HSPA. *See* High Speed Packet Access
 HSS. *See* Home Subscriber Service
 Hybrid ARQ, 97–101
 adaptive, 302
 asynchronous, 302
 control signalling for, 203–207, 277–284, 285–287
 in LTE, 108, 133–134, 299–314
 non-adaptive, 302–303
 process, 301
 relaying, 418–421
 synchronous, 302
 with soft combining, 134–135

I

ICIC. *See* Inter-cell interference coordination
 ICS, 465–466
 IMT-2000, 3–8
 IMT-Advanced, 4–8, 13–15, 112
 Incremental redundancy (IR), 98–101, 301–302
 In-sequence delivery, 316–317
 Inter-cell interference coordination (ICIC), 107–108, 374–375
 in closed subscriber groups, 401
 in heterogeneous deployments, 388–392
 Interference coordination. *See* Inter-cell Interference Coordination
 Interference Rejection Combining (IRC), 67
 Intermodulation:
 receiver, 466–467
 transmitter, 464
 Internet Protocol (IP), 9
 IR. *See* Incremental Redundancy
 ITU, 3–8
 IMT. *See also* IMT-Advanced

J

Joint Reception, 385
 Joint Transmission, 384

L

Latency, 474–475
 Layer mapping, LTE, 189–190, 261–262

Licensed Shared Access (LSA), 497
 Link adaptation, 85–87, 127. *See also* rate control
 Logical channels, 126, 128–129
 LSA. *See* Licensed Shared Access
 LTE-Advanced, 113. *See also* Release 10, IMT-Advanced

M

MAC, 128–135
 control element, 131
 Machine-type Communication, 491–492
 Master Information Block (MIB), 351–354
 MBMS, 403–411
 MBMS Control Channel (MCCH), 128, 406–411
 MBMS Traffic Channel (MTCH), 129, 406, 409
 MB-MSR. *See* Multi-band Base Station
 MBSFN, 112, 142, 145–147, 404–411
 MBSFN area, 404
 MBSFN subframe, 145–147, 407–408, 417
 MCCH. *See* MBMS Control Channel
 MCH. *See* Multicast Channel
 MCH Scheduling Information (MSI), 409
 MCH Scheduling Period (MSP), 409
 MCH subframe allocation (MSA), 409
 MIB. *See* Master Information Block
 MIMO. *See* spatial multiplexing
 MME. *See* Mobility Management Entity
 Mobility Management Entity (MME), 112
 MSR. *See* Multi-Standard Radio
 MTCH. *See* MBMS Traffic Channel
 Multi-band Base Station, 467–470
 Multi-carrier transmission, 26–28
 for LTE. *See* carrier aggregation
 Multicast, 46–48. *See also* MBMS
 Multicast/Broadcast over Single-Frequency Network.
 See MBSFN
 Multicast Channel (MCH), 129–130, 406–408
 Multi-cell/multicast Coordination Entity (MCE), 404
 Multi-point coordination/transmission. *See* Coordinated
 Multipoint transmission
 Multi-Standard Radio (MSR) base stations, 447–450

N

NAS. *See* Non-Access Stratum
 New-data indicator, 227, 235, 301–305
 Non-Access Stratum (NAS), 122
 Non-contiguous Carrier Aggregation, 114, 155, 441–442

O

OCC. *See* Orthogonal Cover Codes
 OFDM. *See* Orthogonal Frequency Division Multiplex
 OFDMA, 44
 OI. *See* Overload Indicator
 Operating bands, 432–438

Operating band unwanted emissions (UEM), 452, 459–460
 Orthogonal Cover Codes (OCC):
 for downlink demodulation reference signals, 176, 179
 for CSI reference signals, 181
 for uplink reference signals, 254
 Orthogonal Frequency Division Multiplex (OFDM), 29–48
 demodulation, 31–33
 FFT/IFFT implementation, 33–34
 in LTE, 141–145
 Out-of-band:
 blocking, 465–466
 emissions, 458–463
 Overload Indicator (OI), 374–375

P

Packet Data Network Gateway (P-GW), 122
 Packet Data Convergence Protocol (PDCP), 124–125
 Paging, 369–370
 Paging Channel (PCH), 129
 Paging Control Channel (PCCH), 128
 Partial path-loss compensation. *See* path-loss compensation
 Path-loss compensation, 293–295
 PBCH. *See* Physical Broadcast Channel
 PCCH. *See* Paging Control Channel
 PCFICH. *See* Physical Control Format Indicator Channel
 PCH. *See* Paging Channel
 PCRF. *See* Policy and Charging Rules Function
 PDCCH. *See* Physical Downlink Control Channel
 PDCP. *See* Packet Data Convergence Protocol
 PDSCH. *See* Physical Downlink Shared Channel
 PDU. *See* Protocol Data Unit
 P-GW. *See* Packet Data Network Gateway
 PHICH. *See* Physical Hybrid-ARQ Indicator Channel
 Physical Broadcast Channel (PBCH), 136
 Physical channel, 136
 Physical Control Format Indicator Channel (PCFICH), 136, 200–203
 Physical Downlink Control Channel (PDCCH), 136, 207–212
 Physical Downlink Shared Channel (PDSCH), 136, 161–171
 Physical Hybrid-ARQ Indicator Channel, 136, 203–207
 Physical-layer cell identity, 173, 347
 Physical Multicast Channel (PMCH), 136, 407
 Physical Random Access Channel (PRACH), 136, 360
 Physical Resource Block (PRB):
 downlink, 144–145, 167–171
 uplink, 144–145, 244
 Physical Uplink Control Channel (PUCCH), 137, 267–284
 transmit diversity for, 266
 Physical Uplink Shared Channel (PUSCH), 137, 241–248
 PMCH. *See* Physical Multicast Channel
 PMI. *See* Precoding-Matrix Indicator
 Policy and Charging Rules Function (PCRF), 122

Positioning, 112
 Power control:
 for PRACH, 363–364
 for PUCCH, 291–293
 for PUSCH, 293–295
 for SRS, 295
 in LTE uplink, 288–295
 Power Headroom, 328–329
 PRACH. *See* Physical Random Access Channel
 PRB. *See* Physical Resource Block
 Pre-coding, LTE:
 codebook-based, 189–193
 closed-loop, 191–192
 non-codebook-based, 193–195
 open-loop, 192–193
 uplink, 261–265
 Precoding-Matrix Indicator (PMI), 191, 286, 337
 Protocol Data Unit (PDU), 126
 PUCCH. *See* Physical Uplink Control Channel
 PUSCH. *See* Physical Uplink Shared Channel

Q

QCL. *See* Quasi-co-located
 QoS, 122
 QPP, 164–165
 QPSK, 21–22
 Quasi-co-located, 148, 183

R

RACH. *See* Random Access Channel
 Radio access network, 103, 123
 Radio bearer, 124
 Radio frame. *See* Frame
 Radio Link Control (RLC), 127–128, 299, 314–315
 Radio Resource Control (RRC), 137–139
 Radio Resource Management (RRM), 123
 RAKE, 50
 RAN. *See* Radio access network
 Random access, 358–369
 preamble, 360–366
 Random Access Channel (RACH), 130
 Range expansion, 390
 Rank indicator, 286, 338–339
 Rate adaptation, 106. *See also* link adaptation
 Rate control, 85, 94. *See also* rate adaptation
 Rate matching:
 in downlink, 165–166
 in uplink, 242
 Receive diversity, 65, 108
 Redundancy version, 99–100, 165, 226, 235, 242, 303
 Reference sensitivity:
 for LTE BS, 464

- for LTE terminal, 464–465
 - Reference signal:
 - cell-specific, 172–175
 - CRS. *See* reference signal, cell specific
 - CSI. *See* reference signal, for channel-state information
 - demodulation downlink, 175–179
 - demodulation uplink, 249–257
 - for channel-state information, 179–183
 - for MBSFN, 172, 407–408
 - for EPDCCH, 179
 - for positioning, 172
 - for sounding. *See* sounding reference signal
 - for uplink demodulation, 249–257
 - UE-specific. *See* reference signal, demodulation downlink
 - Reference-signal sequence:
 - phase-rotated/“cyclically shifted”, 252–254
 - time-frequency structure, 249–251
 - Reference symbol, 38, 172
 - frequency shifted, 173
 - Relaying, 413–429
 - access link, 414
 - amplify-and-forward. *See* repeater
 - architecture, 415–416
 - backhaul link, 414
 - control signalling. *See* Relay Physical Downlink Control Channel
 - decode-and-forward, 413
 - hybrid-ARQ operation, 418–421
 - inband, 415
 - outband, 415
 - REG. *See* Resource Element Group
 - Relay Physical Downlink Control Channel (R-PDCCH), 421–425
 - Relative Narrowband Transmit Power (RNTP), 375
 - Release 8 (LTE), 103–111
 - Release 9 (LTE), 111–112
 - Release 10 (LTE), 112–115. *See also* LTE-Advanced
 - Release 11 (LTE), 115–117
 - Release 12 (LTE), 486
 - Reordering, 316–317
 - Repeater, 413
 - Resource-allocation type:
 - in downlink, 228–230
 - in uplink, 237
 - Resource block, 144–145
 - in downlink, 167–171
 - in uplink, 244–248
 - signaling of, 228–230, 237
 - Resource-block pair, 144
 - Resource element, 144
 - Resource element group, 201
 - RF bandwidth, 446
 - for carrier aggregation, 441
 - for MSR base stations, 449
 - for multi-band base stations, 467–468
 - RF characteristics, 431–471
 - RI. *See* Rank Indicator
 - RLC. *See* Radio Link Control
 - RNTP. *See* Relative Narrowband Transmit Power
 - R-PDCCH. *See* Relay Physical Downlink Control Channel
 - RRC. *See* Radio Resource Control
 - RRC_IDLE, 138–139
 - RRC_CONNECTED, 138–139
 - RRM. *See* Radio Resource Management
- ## S
- S1 interface, 123
 - SAE, 10, 103
 - Scheduling:
 - assignment, 207, 223–233
 - channel-dependent, 85, 87–96
 - downlink, 87–91, 322–323
 - frequency-domain, 94, 106–107
 - grant, 207, 233–237
 - greedy filling, 93
 - half-duplex FDD, 333–335
 - max-C/I, 88–91
 - proportional fair, 91
 - request, 325–327, 358
 - round robin, 90–91
 - semi-persistent, 332–333
 - uplink, 92–94, 324–329
 - SDU. *See* Service Data Unit
 - Search space, 218–223
 - Selectivity:
 - adjacent channel. *See* ACS
 - in-channel. *See* ICS
 - Self-backhauling. *See* Relay
 - Service Data Unit (SDU), 126
 - Serving Gateway (S-GW), 122
 - SFBC. *See* Space-Frequency (Block) Coding
 - SFN. *See* Single-frequency network
 - SFN. *See* System Frame Number
 - S-GW. *See* Serving Gateway
 - Shared-channel transmission, 106, 132
 - SI. *See* System Information message
 - SIB. *See* System Information Block
 - SIC. *See* Successive Interference cancellation
 - Single-frequency network (SFN), 48. *See also* MBSFN
 - SI-RNTI. *See* System Information RNTI
 - Slot, 142–144
 - Sounding reference signal, 257–261
 - Space-Frequency Block Coding (SFBC), 73, 187
 - Space Time (Block) Coding (ST(B)C), 72

Spatial multiplexing, 77–83
 multi-codeword based, 82
 pre-coder based, 80–81
 single-codeword based, 82
 in LTE. *See* Pre-coding (LTE)
 Special subframe, 149
 Spectral efficiency:
 performance, 476–477
 requirement, 480–481
 Spectrum sharing, 497
 Spectrum flexibility, 109–111
 Spurious emissions, 452–453, 458, 463, 470
 Standardization, 10–15
 Subcarrier spacing:
 in general, 29–31, 41–42
 LTE, 141
 reduced, 142
 Subframe, 142
 Successive Interference Cancellation (SIC), 82
 Symbol quadruplet, 201
 Synchronization signal, 348–351
 System Frame Number (SFN), 143, 352
 System Information, 351
 System Information Block (SIB), 351, 354–355
 System Information message (SI), 356
 System Information RNTI (SI-RNTI), 355
 System performance, 473–475
 System throughput, 475

T

TDD. *See* Time Division Duplex
 TD-SCDMA, 13, 110, 153–154
 Three-dimensional beam-forming, 490–491
 Time Division Duplex:
 dynamic, 488
 in LTE, 109–110, 150–153
 spectrum, 4, 8
 Timing advance, 295–297
 for relaying, 425–426
 Transmission mode, 184–186
 Transmission rank, 190
 Transmission Time Interval (TTI), 129
 Transmit Diversity, 70–73
 for PUCCH, 266
 LTE downlink, 186–189
 BCH, 352

Transparent mode (for RLC), 316
 Transport block, 129
 signaling of size, 230–233
 Transport channels, 129
 Transport format, 129
 Transport format selection, 129
 TTI. *See* Transmission Time Interval

U

UCI. *See* Uplink Control Information
 UE categories, 118
 UEM, 459–460. *See also* Operating band unwanted emissions
 UL-SCH. *See* Uplink Shared Channel
 Ultra-dense deployments, 498
 Unacknowledged mode (for RLC), 316
 Unwanted emissions, 441
 Uplink Control Information (UCI), 136
 Uplink-Downlink separation, 488
 Uplink Shared Channel (UL-SCH), 130
 UpPTS, 152
 User Equipment (UE), 118
 categories, 118
 User throughput, 90, 266, 475
 UTRA FDD, 5, 13, 432, 449. *See also* WCDMA
 UTRA TDD, 13, 153, 449

V

Virtual Resource Block (VRB), 168, 230
 distributed vs. localized, 169
 downlink, 168
 uplink, 245
 VRB. *See* Virtual Resource Block

W

WCDMA, 1, 4, 14
 WLAN:
 interworking with, 489–490

X

X2 Interface, 108, 123–124

Z

Zadoff-Chu sequences, 251–252, 364–366
 Zero-power CSI-RS. *See* CSI-RS, Zero Power

This page intentionally left blank

This page intentionally left blank