

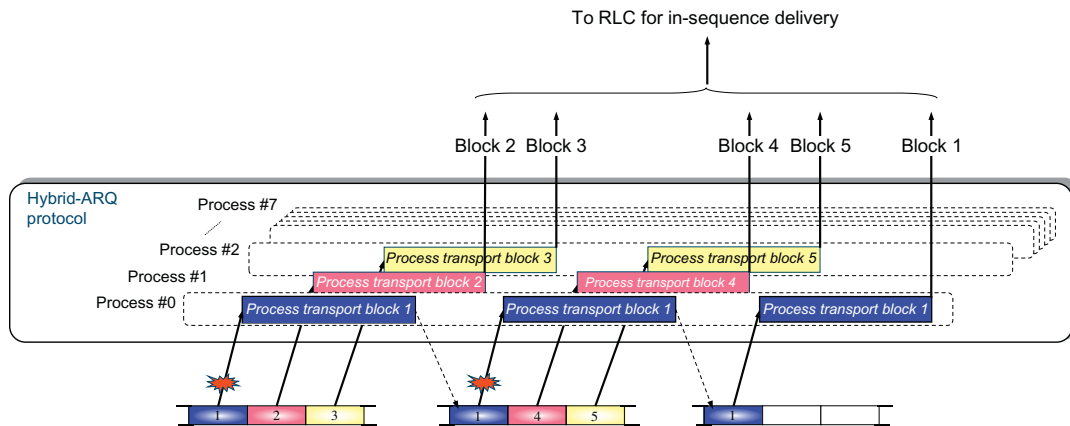


FIGURE 12.2

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

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. Clearly, 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 along with other downlink scheduling information on the PDCCH. 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 – that is, 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 parallel 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*,

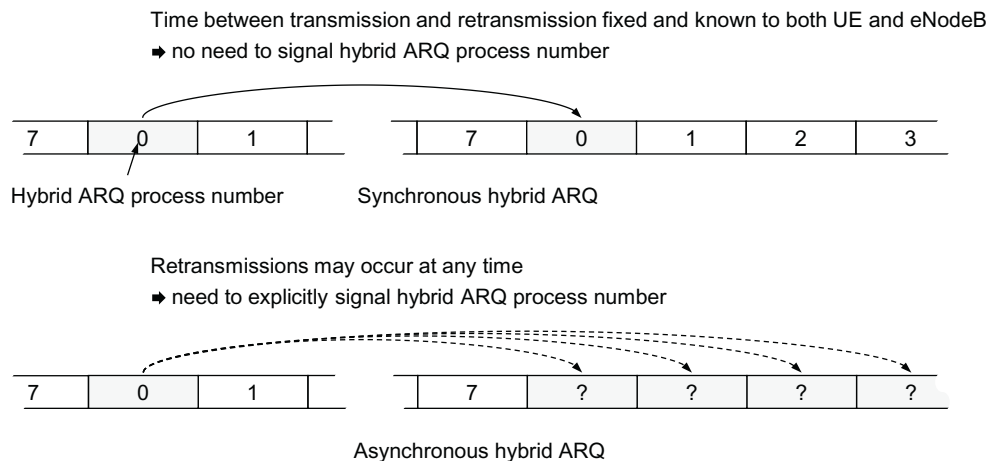


FIGURE 12.3

Synchronous and asynchronous hybrid ARQ.

in contrast, implies that the retransmission must occur at the same frequency resources and with the same transmission format as the initial transmission.

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.

Obviously, 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 is 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 is 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 obviously 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.

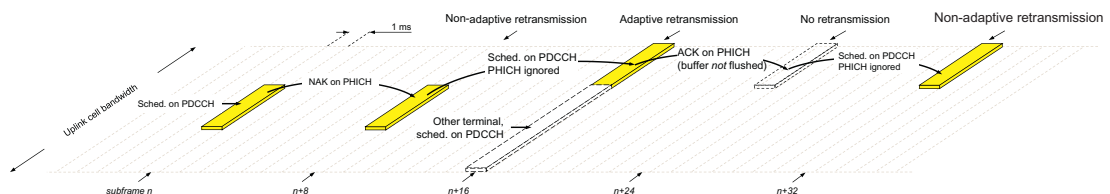


FIGURE 12.4

Non-adaptive and adaptive hybrid-ARQ operation.

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 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.

In order to minimize the overhead and not introduce any additional signaling in the uplink grants, the PHICH that the terminal will expect the hybrid-ARQ acknowledgement upon is derived from the number of the first resource block upon which the corresponding uplink PUSCH transmission occurred. This principle is also compatible with semi-persistently scheduled transmission (see Chapter 13) as well as retransmissions. In addition, the resources used for a particular PHICH further depend on the reference-signal phase rotation signaled as part of the uplink grant. In this way, multiple terminals scheduled on the same set of resources using multi-user MIMO will use different PHICH resources as their reference signals are assigned different phase rotations through the corresponding field in the uplink grant. For spatial multiplexing, where two PHICH resources are needed, the second PHICH uses the same principle as the first, but to ensure that different PHICHs are used for the two transport blocks, the resource for the second PHICH is derived not from the first but from the second resource block upon which the PUSCH was transmitted.¹

The above association is also used in the case of carrier aggregation. For the case when no cross-carrier scheduling is used and each uplink component carrier is scheduled on its corresponding downlink component carrier, this is straightforward and there is no risk of multiple uplink component carriers being associated with the same PHICH resource. With cross-carrier scheduling, on the other hand, transmissions on multiple uplink component carriers may need to be acknowledged on a single downlink component carrier, as illustrated in Figure 12.5. Avoiding PHICH collisions in this case is up to the scheduler by ensuring that different reference-signal phase rotations or different resource-block starting positions are used for the different uplink component carriers. For semi-persistent scheduling the reference-signal phase rotation is always set to zero, but since semi-persistent scheduling is supported on the primary component carrier only, there is no risk of collisions between component carriers.

In the case of multiple component carriers being used for the uplink data transmission in a subframe, there are one (in the absence of spatial multiplexing) or two (in the presence of spatial multiplexing)

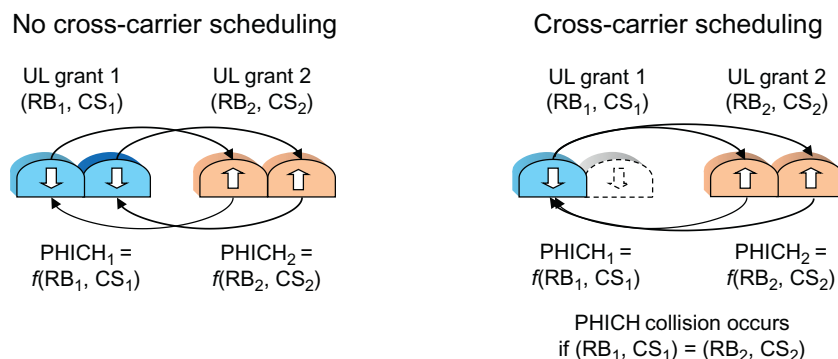


FIGURE 12.5

PHICH association.

¹In essence, this implies that uplink spatial multiplexing must use at least two resource blocks in the frequency domain.

PHICHs transmitted per uplink component carrier. As a general principle, LTE transmits the PHICH on the same component carrier that was used for the grant scheduling the corresponding uplink data transmission. Not only is this principle general in the sense that it can handle symmetric as well as asymmetric carrier aggregation scenarios, it is also beneficial from a terminal power consumption perspective as the terminal only need to monitor the component carriers it monitors for uplink scheduling grants (especially as the PDCCH may override the PHICH to support adaptive retransmissions as discussed above).

12.1.3 Hybrid-ARQ Timing

Clearly, 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 in both uplink and downlink for FDD. For TDD, the number of processes depends on the downlink–uplink allocation, as discussed below.

Starting with the FDD case, transmission of acknowledgements in the uplink in response to downlink data transmission is illustrated in Figure 12.6. 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

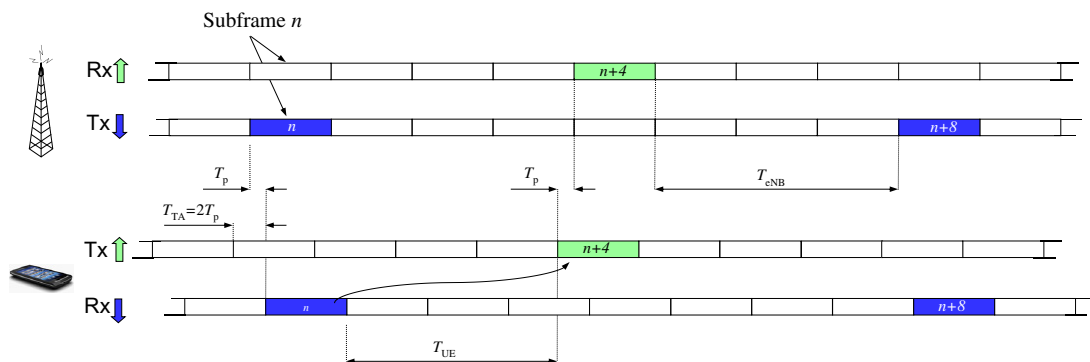


FIGURE 12.6

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

in Section 11.5). Upon reception of the hybrid-ARQ acknowledgement, the eNodeB can, if needed, retransmit the downlink data in subframe $n + 8$. Thus, eight hybrid-ARQ processes are used – that is, the hybrid-ARQ round-trip time is 8 ms.

The description in the previous paragraph, as well as the illustration in Figure 12.6, 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 – that is, 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 13.2.2) and allows a scheduling grant on the PDCCH to override the PHICH, as shown in Figure 12.4.

In Figure 12.6 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 thus 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.

Shifting to 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. Obviously, an uplink hybrid-ARQ acknowledgement can only be transmitted in an uplink subframe and a downlink acknowledgement only in a downlink subframe. The amount of processing time required in the terminal and the eNodeB remains the same though, as the same turbo decoders are used and the scheduling decisions are similar. 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 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 $0 + 7 = 7$.

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	–

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 12.7). 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 13.2.2. 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, in the absence of carrier aggregation, there is always a one-to-one relation between uplink and downlink subframes. Hence, 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.

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 and therefore the amount of PHICH groups may 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 12.7 (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.

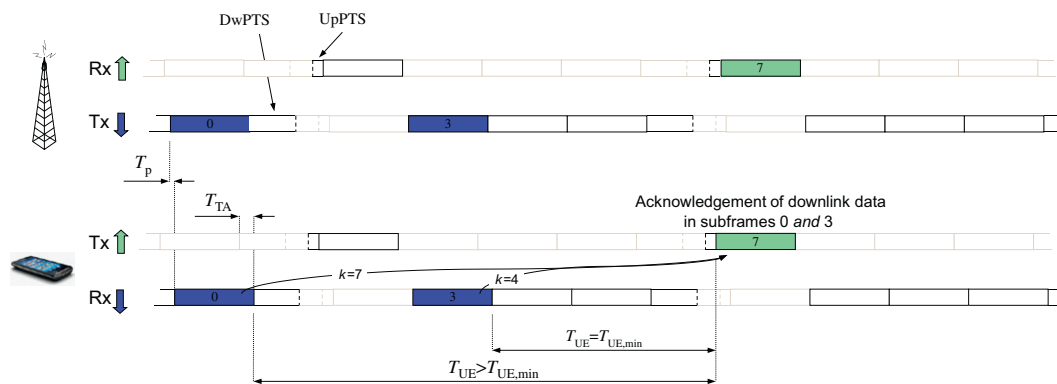


FIGURE 12.7

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

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 12.7 are correctly decoded will a positive acknowledgement be transmitted in uplink subframe 7.

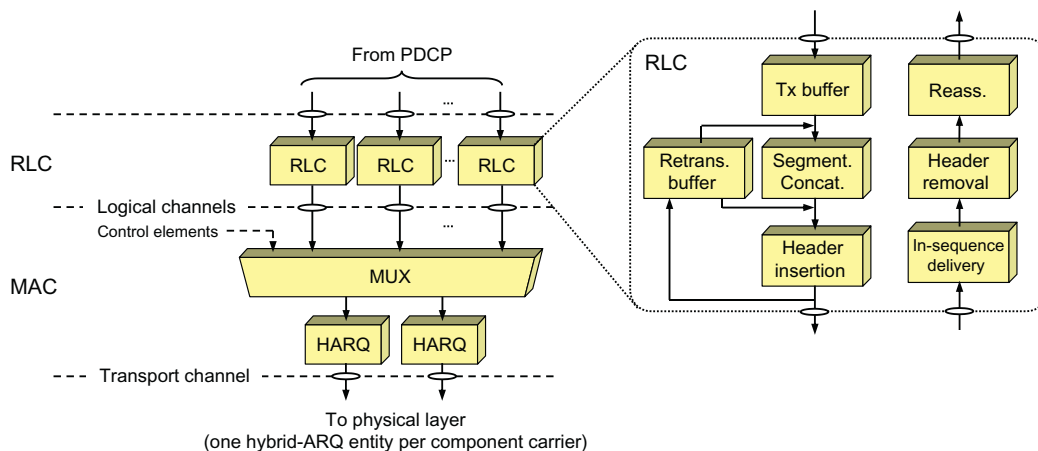
Combining acknowledgements related to multiple downlink transmissions into a single uplink message assumes that the terminal has not missed any of the scheduling assignments upon which the acknowledgement is based. Assume, as an example, that the eNodeB scheduled the terminal in two (subsequent) subframes, but the terminal missed the PDCCH transmission in the first of the two subframes and successfully decoded the data transmitted in the second subframe. Obviously, 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 10.4.4) in the scheduling assignment on the PDCCH 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.

TDD is not the only example when transmissions in multiple downlink subframes must be acknowledged in a single uplink subframe. Carrier aggregation in FDD with different numbers of uplink and downlink component carriers² in essence leads to the same problem. 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.

12.2 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.

²In LTE release 10, the number of downlink component carriers is always at least as large as the number of uplink component carriers (see Chapter 9).

**FIGURE 12.8**

MAC and RLC structure (single-terminal view).

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.

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 the case of carrier aggregation.

12.2.1 Segmentation, 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 were 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.

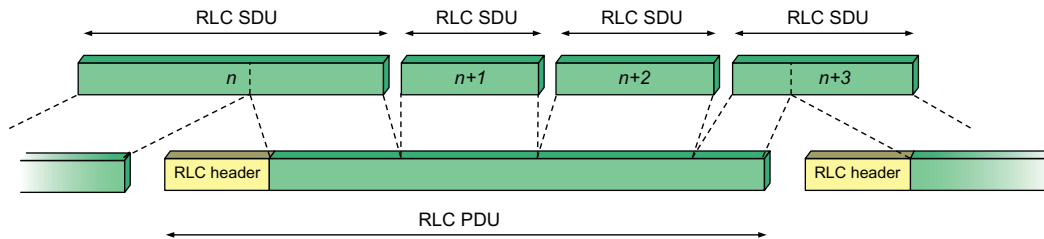


FIGURE 12.9

Generation of RLC PDUs from RLC SDUs.

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 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 possible 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 that 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 obviously needed 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 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 number have been delivered. Only when all PDUs with lower sequence number 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

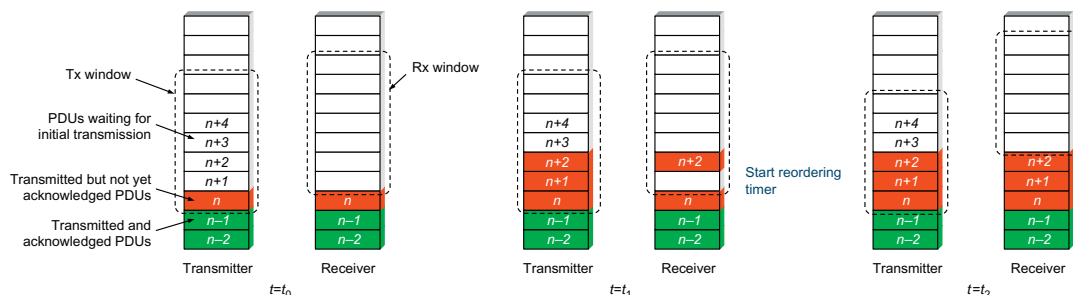


FIGURE 12.10

In-sequence delivery.

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 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 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 $n + 1$. Clearly, 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

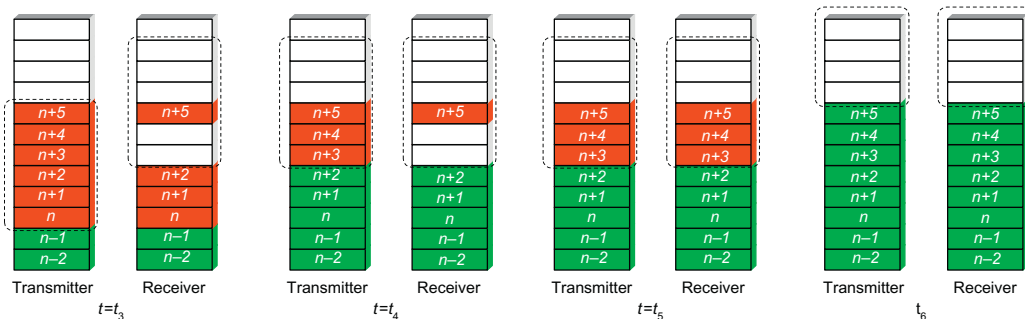


FIGURE 12.11

Retransmission of missing PDUs.

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 impact 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.

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.

Power Control, Scheduling, and Interference Handling

13

This chapter deals with some radio-resource-management issues, including uplink power control, downlink and uplink scheduling, and different means for inter-cell interference coordination to handle/avoid severe interference between different cells including cells of different layers in so-called *heterogeneous network deployments*.

13.1 UPLINK POWER CONTROL

Uplink power control for LTE is the set of algorithms and tools by which the transmit power for different uplink physical channels and signals are controlled to ensure that they, if possible, are received at the cell site with appropriate power. This means that the transmission should be received with sufficient power to allow for proper demodulation of the corresponding information. At the same time, the transmit power should not be unnecessarily high as that would cause unnecessary interference to other cells.¹ The transmit power will thus depend on the channel properties, including the channel attenuation and the noise and interference level at the receiver side. Furthermore, in the case of DL-SCH transmission on PDSCH, if the received power is too low one can either increase the transmit power or reduce the data rate by use of rate control. Thus, in this case there is an intimate relation between power control and rate control.

How to set the transmit power for random access will be discussed in Chapter 14. Here we will mainly discuss the power-control mechanism for the PUCCH and PUSCH physical channels. We will also briefly discuss the power setting for sounding reference signals. Uplink *demodulation* reference signals are always transmitted together and time-multiplexed with PUSCH or PUCCH. The demodulation reference signals are then transmitted with the same power as the corresponding physical channel. This is also true in the case of uplink spatial multiplexing if the reference signal power is defined as the total power of all demodulation reference signals transmitted by the terminal. Expressed differently, the power of a single demodulation reference signal is equal to the corresponding *per-layer* PDSCH power.

Fundamentally, LTE uplink power control is a combination of an *open-loop* mechanism, implying that the terminal transmit power depends on estimates of the downlink path loss, and a *closed-loop* mechanism, implying that the network can, in addition, directly adjust the terminal transmit power by means of explicit *power-control commands* transmitted on the downlink. In practice, these power-control commands are determined based on prior network measurements of the received uplink power, thus the term “*closed loop*”.

¹In the case of uplink MU-MIMO there may also be interference to the cell itself.

13.1.1 Uplink Power Control – Some Basic Rules

Before going into the details of the power control algorithms for PUSCH and PUCCH, some basic rules for the power assignment to different physical channels will be discussed. These rules mainly deal with the presence of different transmit-power limitations and how these limitations impact the transmit-power setting for different physical channels. This is especially of interest in the case of the simultaneous transmission of multiple physical channels from the same terminal, a situation that may occur for LTE release 10 and beyond:

- With LTE release 10 there is the possibility for *carrier aggregation*, implying that multiple PUSCH may be transmitted in parallel on different component carriers.
- With LTE release 10 there is also the possibility for *simultaneous PUSCH/PUCCH transmission* on the same or different component carriers.

In principle, each physical channel is separately and independently power controlled. However, in the case of multiple physical channels to be transmitted in parallel from the same terminal, the total power to be transmitted for all physical channels may, in some cases, exceed the maximum terminal output power P_{TMAX} corresponding to the terminal power class. As will be seen below, the basic strategy is then to first ensure that transmission of any L1/L2 control signaling is assigned the power assumed to be needed for reliable transmission. The remaining available power is then assigned to the remaining physical channels.

For each uplink component carrier configured for a terminal there is also an associated and explicitly configured *maximum per-carrier transmit power* $P_{\text{CMAX},c}$, which may be different for different component carriers (indicated by the index c). Furthermore, although it obviously does not make sense for $P_{\text{CMAX},c}$ to exceed the maximum terminal output power P_{TMAX} , the sum of $P_{\text{CMAX},c}$ for all configured component carriers may very well, and typically will, exceed P_{TMAX} . The reason is that, in many cases, the terminal will not be scheduled for uplink transmission on all its configured component carriers and the terminal should also in that case be able to transmit with its maximum output power.

As will be seen in the next sections, the power control of each physical channel explicitly ensures that the total transmit power for a given component carrier does not exceed $P_{\text{CMAX},c}$ for that carrier. However, the separate power-control algorithms do not ensure that the total transmit power for all component carriers to be transmitted by the terminal does not exceed the maximum terminal output power P_{TMAX} . Rather, this is ensured by a subsequent *power scaling* applied to the physical channels to be transmitted. This power scaling is carried out in such a way that any L1/L2 control signaling has higher priority, compared to data (UL-SCH) transmission.

If PUCCH is to be transmitted in the subframe it is first assigned the power determined by its corresponding power-control algorithm, before any power is assigned to any PUSCH to be transmitted in parallel PUCCH. This ensures that L1/L2 control signaling on PUCCH is assigned the power assumed to be needed for reliable transmission before any power is assigned for data transmission.

If PUCCH is not transmitted in the subframe but L1/L2 control signaling is multiplexed on to PUSCH, the PUSCH carrying the L1/L2 control signaling is first assigned the power determined by its corresponding power-control algorithm, before any power is assigned to any other PUSCH to be transmitted in parallel. Once again, this ensures that L1/L2 control signaling is assigned the power assumed to be needed before any power is assigned for other PUSCH transmissions only carrying UL-SCH. Note that, in the case of transmission of multiple PUSCH in parallel (carrier aggregation),

at most one PUSCH may include L1/L2 control signaling. Also, there cannot be PUCCH transmission and L1/L2 control signaling multiplexed on to PUSCH in the same subframe. Thus, there will never be any conflict between the above rules.

If the remaining available transmit power is not sufficient to fulfill the power requirements of any remaining PUSCH to be transmitted, the powers of these remaining physical channels, which only carry UL-SCH, are scaled so that the total power for all physical channels to be transmitted does not exceed the maximum terminal output power.

Overall, the PUSCH power scaling, including the priority for PUSCH with L1/L2 control signaling, can thus be expressed as:

$$\sum_i w_c \cdot P_{\text{PUSCH},c} \leq P_{\text{TMAX}} - P_{\text{PUCCH}}, \quad (13.1)$$

where $P_{\text{PUSCH},c}$ is the transmit power for PUSCH on carrier c as determined by the power-control algorithm (before power scaling but including the per-carrier limitation $P_{\text{CMAX},c}$), P_{PUCCH} is the transmit power for PUCCH (which is zero if there is no PUCCH transmission in the subframe), and w_c is the power-scaling factor for PUSCH on carrier c ($w_c \leq 1$). For any PUSCH carrying L1/L2 control signaling the scaling factor w_c should be set to 1. For the remaining PUSCH, some scaling factors may be set to zero by decision of the terminal, in practice implying that the PUSCH, as well as the corresponding UL-SCH mapped to the PUSCH, are not transmitted. For the remaining PUSCH the scaling factors w_c are set to the same value less than or equal to 1 to ensure that the above inequality is fulfilled. Thus, all PUSCH that are actually transmitted are power scaled by the same factor.

After this overview of some general rules for the power setting of different terminals, especially for the case of multiple physical channels transmitted in parallel from the same terminal, the power control carried out separately for each physical channel will be described in more detail.

13.1.2 Power Control for PUCCH

For PUCCH, the appropriate received power is simply the power needed to achieve a desired – that is, a sufficiently low – error rate in the decoding of the L1/L2 control information transmitted on the PUCCH. However, it is then important to bear the following in mind:

- In general, decoding performance is not determined by the *received signal strength* but rather by the *received signal-to-interference-plus-noise ratio* (SINR). What is an appropriate received power thus depends on the interference level at the receiver side, an interference level that may differ between different deployments and which may also vary in time as, for example, the load of the network varies.
- As described in Chapter 11, there are different PUCCH formats which are used to carry different types of uplink L1/L2 control information (hybrid-ARQ acknowledgements, scheduling requests, channel-state reports, or combinations thereof). The different PUCCH formats thus carry different numbers of information bits per subframe and the information they carry may also have different error-rate requirements. The required received SINR may therefore differ between the different PUCCH formats, something that needs to be taken into account when setting the PUCCH transmit power in a given subframe.

Overall, power-control for PUCCH can be described by the following expression:

$$P_{\text{PUCCH}} = \min \{ P_{\text{CMAX},c}, P_{0,\text{PUCCH}} + PL_{\text{DL}} + \Delta_{\text{Format}} + \delta \}. \quad (13.2)$$

In the expression above, P_{PUCCH} is the PUCCH transmit power to use in a given subframe and PL_{DL} is the downlink path loss as estimated by the terminal. The “ $\min \{ P_{\text{CMAX},c}, \dots \}$ ” term ensures that the PUCCH transmit power as determined by the power control will not exceed the per-carrier maximum power $P_{\text{CMAX},c}$.

The parameter $P_{0,\text{PUCCH}}$ in expression (13.2) is a cell-specific parameter that is broadcast as part of the cell system information. Considering only the part $P_{0,\text{PUCCH}} + PL_{\text{DL}}$ in the PUCCH power-control expression and assuming that the (estimated) downlink path loss accurately reflects the true uplink path loss, it is obvious that $P_{0,\text{PUCCH}}$ can be seen as the *desired* or *target* received power. As discussed earlier, the required received power will depend on the uplink noise/interference level. From this point of view, the value of $P_{0,\text{PUCCH}}$ should take the interference level into account and thus vary in time as the interference level varies. However, in practice it is not feasible to have $P_{0,\text{PUCCH}}$ varying with the instantaneous interference level. One simple reason is that the terminal does not read the system information continuously and thus the terminal would anyway not have access to a fully up-to-date $P_{0,\text{PUCCH}}$ value. Another reason is that the uplink path-loss estimates derived from downlink measurements will anyway not be fully accurate, for example due to differences between the instantaneous downlink and uplink path loss, as well as due to measurement inaccuracies.

Thus, in practice, $P_{0,\text{PUCCH}}$ may reflect the average interference level, or perhaps only the relatively constant noise level. More rapid interference variations can then be taken care of by closed-loop power control, see below.

For the transmit power to reflect the typically different SINR requirements for different PUCCH formats, the PUCCH power-control expression includes the term Δ_{Format} , which adds a format-dependent power offset to the transmit power. The power offsets are defined such that a baseline PUCCH format, more exactly the format corresponding to the transmission of a single hybrid-ARQ acknowledgement (format 1 with BPSK modulation, as described in Section 11.4.1.1), has an offset equal to 0 dB, while the offsets for the remaining formats can be explicitly configured by the network. For example, PUCCH format 1 with QPSK modulation, carrying two simultaneous acknowledgements and used in the case of downlink spatial multiplexing, should have a power offset of roughly 3 dB, reflecting the fact that twice as much power is needed to communicate two acknowledgements instead of just a single acknowledgement.

Finally, it is possible for the network to directly adjust the PUCCH transmit power by providing the terminal with explicit power-control commands that adjust the term δ in the power-control expression above. These power-control commands are *accumulative* – that is, each received power-control command increases or decreases the term δ by a certain amount. The power-control commands for PUCCH can be provided to the terminal by two different means:

- As mentioned in Section 10.4, a power-control command is included in each downlink scheduling assignment – that is, the terminal receives a power-control command every time it is explicitly scheduled on the downlink. One reason for uplink PUCCH transmissions is the transmission of hybrid-ARQ acknowledgements as a response to downlink DL-SCH transmissions. Such downlink transmissions are typically associated with downlink scheduling assignments on PDCCH and

the corresponding power-control commands could thus be used to adjust the PUCCH transmit power prior to the transmission of the hybrid-ARQ acknowledgements.

- Power-control commands can also be provided on a special PDCCH that simultaneously provides power-control commands to multiple terminals (PDCCH using DCI format 3/3A; see Section 10.4.7). In practice, such power-control commands are then typically transmitted on a regular basis and can be used to adjust the PUCCH transmit power, for example prior to (periodic) uplink channel-state reports. They can also be used in the case of semi-persistent scheduling (see Section 13.2.3), in which case there may be uplink transmission of both PUSCH (UL-SCH) and PUCCH (L1/L2 control) without any explicit scheduling assignments/grants.

The power-control command carried within the uplink scheduling grant consists of two bits, corresponding to the four different update steps -1 , 0 , $+1$, or $+3$ dB. The same is true for the power-control command carried on the special PDCCH assigned for power control when this is configured to DCI format 3A. On the other hand, when the PDCCH is configured to use DCI format 3, each power-control command consists of a single bit, corresponding to the update steps -1 and $+1$ dB. In the latter case, twice as many terminals can be power controlled by a single PDCCH. One reason for including the possibility for 0 dB (no change of power) as one power-control step is that a power-control command is included in *every* downlink scheduling assignment and it is desirable not to have to update the PUCCH transmit power for each assignment.

13.1.3 Power Control for PUSCH

Power-control for PUSCH transmission can be described by the following expression:

$$P_{\text{PUSCH},c} = \min \{ P_{\text{CMAX},c} - P_{\text{PUCCH}}, P_{0,\text{PUSCH}} + \alpha \cdot PL_{\text{DL}} + 10 \cdot \log_{10}(M) + \Delta_{\text{MCS}} + \delta \}, \quad (13.3)$$

where M indicates the instantaneous PUSCH bandwidth measured in number of resource blocks and the term Δ_{MCS} is similar to the term Δ_{Format} in the expression for PUCCH power control – that is, it reflects the fact that different SINR is required for different modulation schemes and coding rates used for the PUSCH transmission.

The above expression is clearly similar to the power-control expression for PUCCH transmission, with some key differences:

- The use of “ $P_{\text{CMAX},c} - P_{\text{PUCCH}}$ ” reflects the fact that the transmit power available for PUSCH on a carrier is the maximum allowed per-carrier transmit power *after power has been assigned to any PUCCH transmission* on that carrier. This ensures priority of L1/L2 signaling on PUCCH over data transmission on PUSCH in the power assignment, as described in Section 13.1.1.
- The term $10 \cdot \log_{10}(M)$ reflects the fact that what is fundamentally controlled by the parameter $P_{0,\text{PUSCH}}$ is the power *per resource block*. For a larger resource assignment, a correspondingly higher received power and thus a correspondingly higher transmit power is needed.²

²One could also have included a corresponding term in the expression for PUCCH power control. However, as the PUCCH bandwidth always corresponds to one resource block, the term would always equal zero.

- The parameter α , which can take a value smaller than or equal to 1, allows for so-called *partial path-loss compensation*, as described below.

In general, the parameters $P_{0,\text{PUSCH}}$, α , and Δ_{MCS} can be different for the different component carriers configured for a terminal.

In the case of PUSCH transmission, the explicit power-control commands controlling the term δ above are included in the uplink scheduling grants, rather than in the downlink scheduling assignments. This makes sense as PUSCH transmissions are preceded by an uplink scheduling grant except for the case of semi-persistent scheduling. Similar to the power-control commands for PUCCH in the downlink scheduling assignment, the power-control commands for PUSCH are multi-level. Furthermore, also in the same way as for PUCCH power control, explicit power-control commands for PUSCH can be provided on the special PDCCH that simultaneously provides power-control commands to multiple terminals. These power-control commands can, for example, be used for the case of PUSCH transmission using semi-persistent scheduling.

Assuming α equal to 1, also referred to as *full path-loss compensation*, the PUSCH power-control expression becomes very similar to the corresponding expression for PUCCH. Thus, the network can select a modulation-and-coding scheme (MCS) and the power-control mechanism, including the term Δ_{MCS} , will ensure that the received SINR will match the SINR required for that modulation-and-coding scheme, *assuming that the terminal transmit power does not reach its maximum value*.

In the case of PUSCH transmission, it is also possible to “turn off” the Δ_{MCS} function by setting all Δ_{MCS} values to zero. In that case, the PUSCH received power will be matched to a certain MCS given by the selected value of $P_{0,\text{PUSCH}}$.

With the parameter α less than 1, the PUSCH power control operates with so-called *partial path-loss compensation* – that is, an increased path loss is not fully compensated for by a corresponding increase in the uplink transmit power. In that case, the received power, and thus the received SINR per resource block, will vary with the path loss and, consequently, the scheduled modulation-and-coding scheme should vary accordingly. Clearly, in the case of fractional path-loss compensation, the Δ_{MCS} function should be disabled. Otherwise, the terminal transmit power would be further reduced when the modulation-and-coding scheme is reduced to match the partial path-loss compensation.

Figure 13.1 illustrates the differences between full path-loss compensation ($\alpha = 1$) and partial path-loss compensation ($\alpha < 1$). As can be seen, with partial path-loss compensation, the terminal transmit power increases more slowly than the increase in path loss (left-hand figure) and, consequently, the received power, and thus also the received SINR, is reduced as the path loss increases (right-hand figure). To compensate for this, the modulation-and-coding scheme – that is, the PUSCH data rate – should be reduced as the path loss increases.

The potential benefit of partial path-loss compensation is a relatively lower transmit power for terminals closer to the cell border, implying less interference to other cells. At the same time, this also leads to a reduced data rate for these terminals. It should also be noted that a similar effect can be achieved with full path-loss compensation by having the scheduled modulation-and-coding scheme depend on the estimated downlink path loss, which can be derived from the power headroom report, and rely on Δ_{MCS} to reduce the relative terminal transmit power for terminals with higher path loss. However, an even better approach would then be to not only base the modulation-and-coding scheme selection on the path loss to the current cell, but also on the path loss to the neighboring interfered cells.

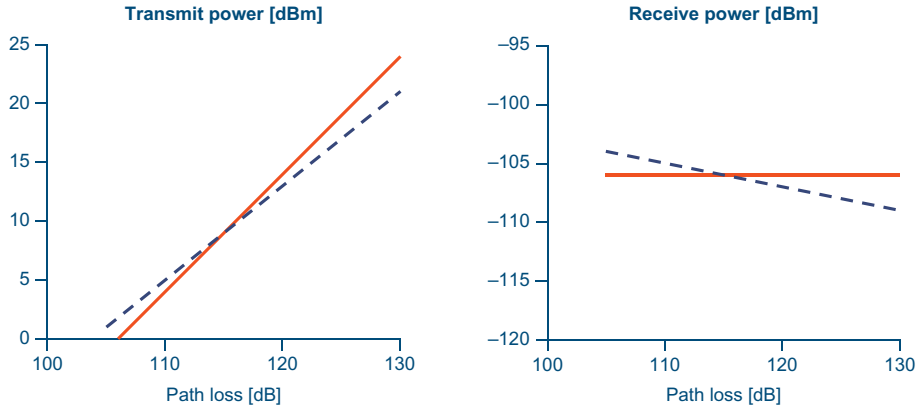


FIGURE 13.1

Full vs. partial path-loss compensation. Solid curve: full compensation ($\alpha = 1$). Dashed curve: partial compensation ($\alpha = 0.8$ in this example).

13.1.4 Power Control for SRS

The SRS transmit power basically follows that of the PUSCH, compensating for the exact bandwidth of the SRS transmission and with an additional power offset. Thus, the power control for SRS transmission can be described according to:

$$P_{\text{SRS}} = \min \{ P_{\text{CMAX},c}, P_{0,\text{PUSCH}} + \alpha \cdot PL_{\text{DL}} + 10 \cdot \log_{10}(M_{\text{SRS}}) + \delta + P_{\text{SRS}} \}, \quad (13.4)$$

where the parameters $P_{0,\text{PUSCH}}$, α , and δ are the same as for PUSCH power control, as discussed in Section 13.1.3. Furthermore, M_{SRS} is the bandwidth, expressed as number of resource blocks, of the SRS transmission and P_{SRS} is a configurable offset.

13.1.5 Power Headroom

To assist the scheduler in the selection of a combination of modulation-and-coding scheme and resource size M that does not lead to the terminal being power limited, the terminal can be configured to provide regular *power headroom* reports on its power usage (see also Section 13.2.2.2). According to Section 13.1.1 there is a separate transmit-power limitation for each component carrier. Thus, power headroom should be measured and reported separately for each component carrier.

There are two different types of power-headroom reports defined for LTE release 10, *Type 1* and *Type 2*. Type 1 reporting reflects the power headroom assuming PUSCH-only transmission on the carrier, while the Type-2 report assumes combined PUSCH and PUCCH transmission.

The Type-1 power headroom valid for a certain subframe, assuming that the terminal was really scheduled for PUSCH transmission in that subframe, is given by the following expression:

$$\text{Power Headroom} = P_{\text{CMAX},c} - (P_{0,\text{PUSCH}} + \alpha \cdot PL_{\text{DL}} + 10 \cdot \log_{10}(M) + \Delta_{\text{MCS}} + \delta), \quad (13.5)$$

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. 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, comparing with expression (13.3) 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. More exactly, a negative power headroom indicates that the per-carrier transmit power was limited by $P_{\text{CMAX},c}$ at the time of the power headroom reporting. 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 what are 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:

$$\text{Power Headroom} = P_{\text{CMAX},c} - (P_{0,\text{PUSCH}} + \alpha \cdot PL_{\text{DL}} + \delta). \quad (13.6)$$

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 $\Delta_{\text{MCS}} = 0\text{dB}$.

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 given by their corresponding power-control expressions (13.2) and (13.1) respectively, once again not taking into account any maximum per-carrier power when calculating the PUSCH and PUCCH transmit power.

Similar to 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 or PUCCH transmit power is calculated, assuming the smallest possible resource assignment ($M = 1$) and $\Delta_{\text{MCS}} = 0\text{dB}$ for PUSCH and $\Delta_{\text{Format}} = 0$ for PUCCH.

13.2 SCHEDULING AND RATE ADAPTATION

The purpose of the scheduler is to determine to/from which terminal(s) to transmit data and on which set of resource blocks. The scheduler is a key element and to a large degree determines the overall behavior of the system. The basic operation is so-called *dynamic* scheduling, where the eNodeB in each 1 ms TTI transmits scheduling information to the selected set of terminals, controlling the uplink and downlink transmission activity. The scheduling decisions are transmitted on the PDCCHs as described in Chapter 10. To reduce the control signaling overhead, there is also the possibility of *semi-persistent scheduling*. Semi-persistent scheduling will be described further in Section 13.2.3.

For carrier aggregation, each component carrier is independently scheduled with individual scheduling assignments/grants and one DL-SCH/UL-SCH per scheduled component carrier. 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.

The *downlink scheduler* is responsible for dynamically controlling the terminal(s) to transmit to and, for each of these terminals, the set of resource blocks upon which the terminal's DL-SCH (or

DL-SCHs in the case of carrier aggregation) is transmitted. Transport-format selection (selection of transport-block size, modulation-and-coding scheme, resource-block allocation, and antenna mapping) for each component carrier and logical channel multiplexing for downlink transmissions are controlled by the eNodeB, as illustrated in the left part of Figure 13.2.

The *uplink scheduler* serves a similar purpose, namely to dynamically control which terminals are to transmit on their UL-SCH (or UL-SCHs in the case of carrier aggregation) and on which uplink resources. The uplink scheduler is in complete control of the transport format the terminal will use, whereas 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 13.2, where the scheduler controls the transport format and the terminal controls the logical-channel multiplexing.

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

13.2.1 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.

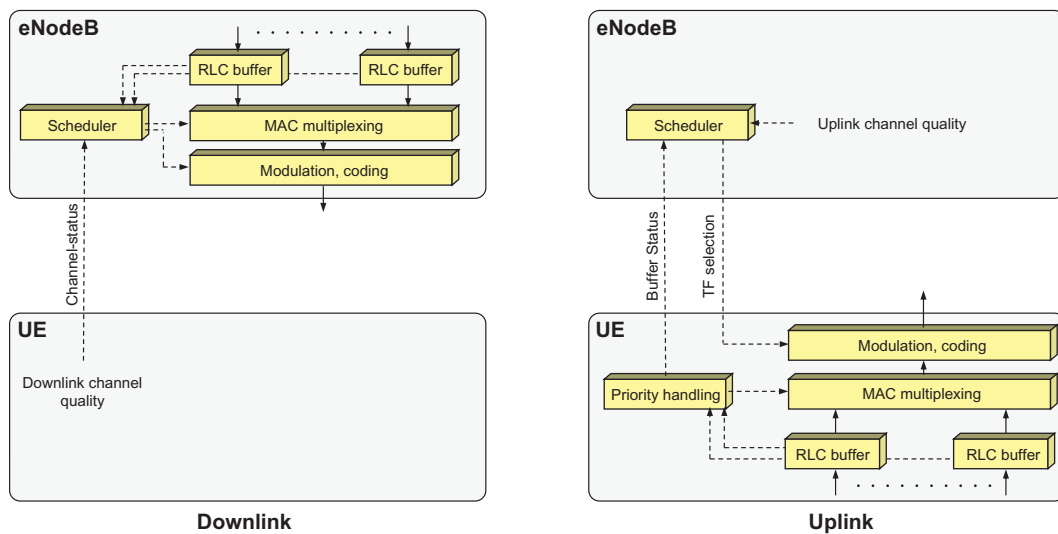


FIGURE 13.2

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

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 obviously 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 decision is communicated to each of the scheduled terminals through the downlink L1/L2 control signaling as described in Chapter 10, using one PDCCH per downlink assignment.

Each terminal monitors a set of PDCCHs as described in Chapter 10 for downlink scheduling assignments. A scheduling assignment is transmitted in the same subframe as the data. If a valid assignment matching the identity of the terminal is found, then the terminal receives and processes the transmitted signal as indicated in the assignment. Once the transport block is successfully decoded, the terminal will demultiplex the received data into the appropriate logical channels.

In the case of carrier aggregation, there is one PDCCH per component carrier. Furthermore, if cross-carrier scheduling is configured (see Chapter 10), the downlink assignment does not have to be transmitted on the same component carrier as the associated data, as information about the component carrier containing the associated data is included in the scheduling assignment in this case.

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 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 channel-state reports from the terminal, further described in Section 13.2.5, 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. Obviously 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.

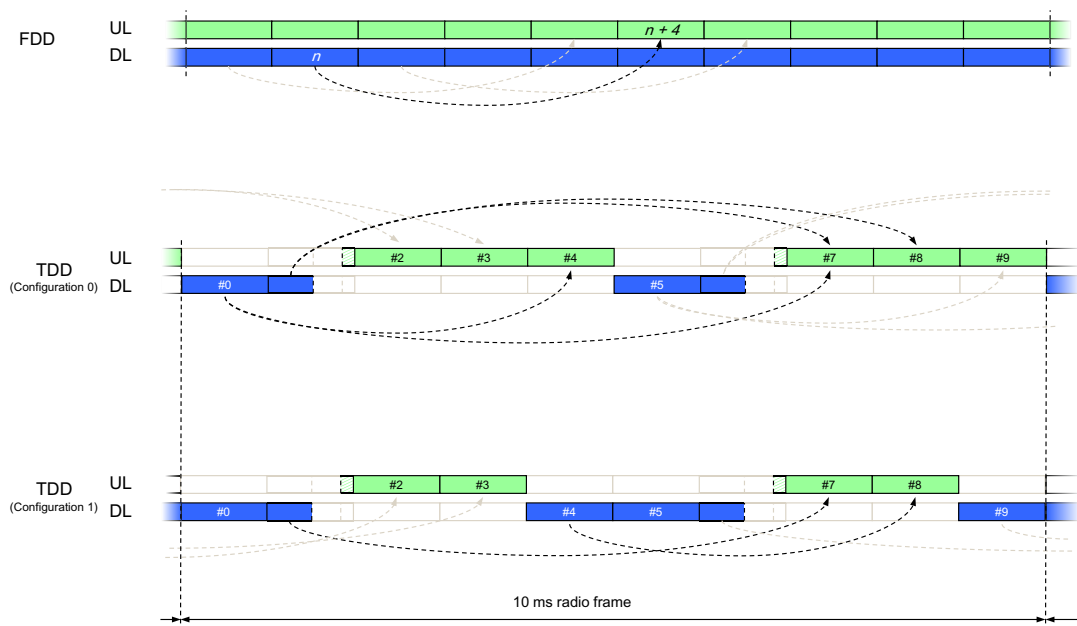
Downlink inter-cell interference coordination is also part of the implementation-specific scheduler strategy. A cell may signal to its neighboring cells the intention to transmit with a lower transmission power in the downlink on a set of resource blocks. This information can then be exploited by neighboring 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 Section 13.3.

13.2.2 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. As the scheduler knows the transport format the terminal will use when it is transmitting, there is no need for outband control signaling from the terminal to the eNodeB. This is beneficial from a coverage perspective, taking into account that the cost per bit of transmitting outband control information can be significantly higher than the cost of data transmission, as the control signaling needs to be received with higher reliability. It also 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 without a corresponding grant. 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.

The terminal monitors a set of PDCCHs as described in Chapter 10 for uplink scheduling grants. Upon detection of a valid uplink grant, the terminal will transmit its UL-SCH according to the information in the grant. Obviously, 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

**FIGURE 13.3**

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

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 13.3. 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.

For TDD, the situation is slightly more complicated 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 receipt 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 Chapter 10 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 13.3, 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.

Similarly to the downlink case, the uplink scheduler can exploit information about channel conditions, buffer status, and priorities of the different data flows, and, if some form of interference coordination is employed, the interference situation in neighboring cells. 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.

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 Section 13.3.

13.2.2.1 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 13.4.

13.2.2.2 Scheduling Requests and Buffer Status Reports

The scheduler needs knowledge about the amount of data awaiting transmission from the terminals to assign the proper amount of uplink resources. Obviously, 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

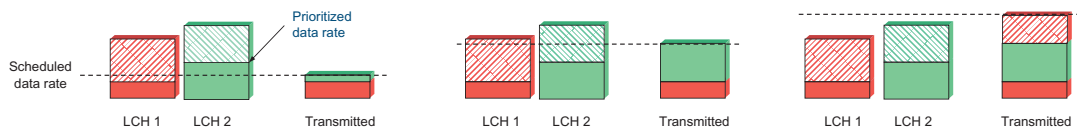


FIGURE 13.4

Prioritization of two logical channels for three different uplink grants.

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*.

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 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 13.5. 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. There is only a single scheduling-request bit, irrespective of the number of uplink component carriers the terminal is capable of. 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.

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

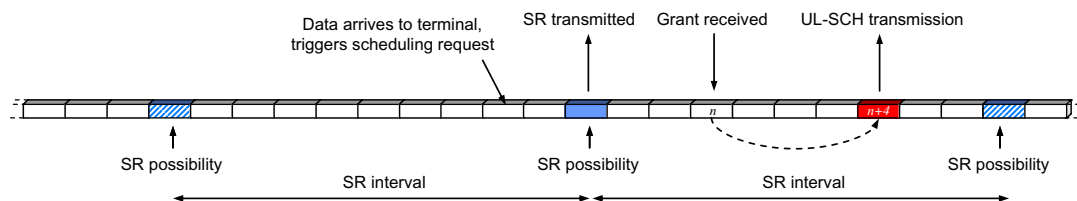


FIGURE 13.5

Scheduling-request transmission.

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.

Although the scheduling-request design for LTE relies on dedicated resources, a terminal that has not been allocated such resources obviously 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.

Terminals that already have a valid grant obviously 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 and the power availability is useful, as discussed above. 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 13.6.

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 channels in a logical-channel group. A buffer-status report represents one or all four logical-channel groups and 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.

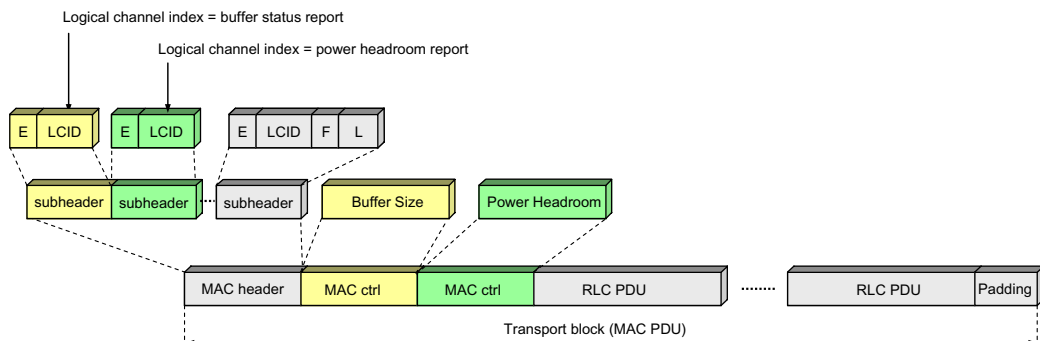


FIGURE 13.6

Signaling of buffer status and power-headroom reports.

- 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. Clearly it is better to exploit the available payload for useful scheduling information instead of padding if possible.

In addition to buffer status, the amount of transmission power available in each terminal is also relevant for the uplink scheduler. Obviously, 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 (as discussed in Section 13.1.5), is defined as the difference between the nominal maximum output power and the estimated output power for UL-SCH transmission. This quantity can be positive as well as negative (on a dB scale), where a negative value would indicate that the network has scheduled a higher data rate than the terminal can support given its current power availability. The power headroom depends on the power-control mechanism and thereby indirectly on factors such as the interference in the system and the distance to the base stations.

Information about the power headroom is fed back from the terminals 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. Type-1 reports are provided for all component carriers simultaneously, while Type-2 reports are provided for the primary component carrier only.

A power headroom report can be triggered for the following reasons:

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

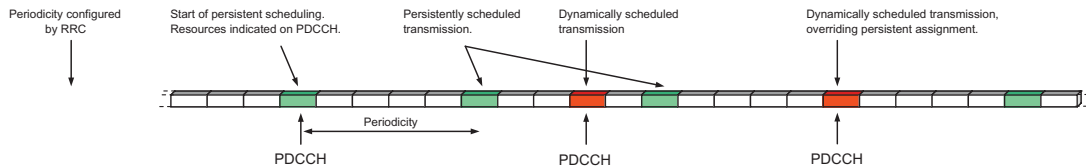
It is also possible to configure a prohibit timer to control the minimum time between two power-headroom reports and thereby the signaling load on the uplink.

13.2.3 Semi-Persistent Scheduling

The basis for uplink and downlinks scheduling is dynamic scheduling, as described in Sections 13.2.1 and 13.2.2. 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 a PDCCH in each subframe. In many situations, the overhead in terms of control signaling on the 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.

With semi-persistent scheduling, the terminal is provided with the scheduling decision on the 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 PDCCH using the semi-persistent C-RNTI.³ For example, for voice-over IP the scheduler can configure a periodicity

³Each terminal has two identities, the “normal” C-RNTI for dynamic scheduling and the semi-persistent C-RNTI for semi-persistent scheduling.

**FIGURE 13.7**

Example of semi-persistent scheduling.

of 20 ms for semi-persistent scheduling and, once a talk spurt starts, the semi-persistent pattern is triggered by the PDCCH.

After enabling semi-persistent scheduling, the terminal continues to monitor the PDCCH 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.

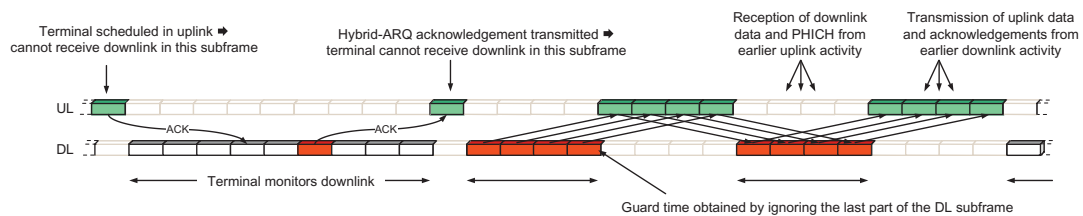
For the downlink, only initial transmissions use semi-persistent scheduling. Retransmissions are explicitly scheduled using a PDCCH assignment. This follows directly from the use of an asynchronous hybrid-ARQ protocol in the downlink. Uplink retransmissions, in contrast, can either follow the semi-persistently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

13.2.4 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 scheduled simultaneously 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

**FIGURE 13.8**

Example of half-duplex FDD terminal operation.

loss in uplink spectral efficiency of 20%. Clearly 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.

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 as, 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 receipt 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.

13.2.5 Channel-State Reporting

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 conditions – is a key feature of LTE. An important part of the support for downlink channel-dependent scheduling is *channel-state reports* provided by terminals to the network, reports on which the latter can base its scheduling decisions.

The channel-state reports 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 downlink transmission to the terminal. RI only needs to be reported by terminals that are configured to be in one of the spatial-multiplexing transmission modes. There is at most one RI reported, valid across the full bandwidth – that is, the RI is frequency non-selective. Frequency-dependent transmission rank would be impossible to utilize since all layers are transmitted on the same set of resource blocks in LTE.
- *Precoder matrix indication* (PMI), indicating which of the precoder matrices (see Chapter 10) should preferably be used for the downlink transmission. The reported precoder matrix is determined assuming 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. Furthermore, the network can restrict the set of matrices from which the terminal should select the recommended precoder, so-called codebook subset restriction, to avoid reporting precoders that are not useful in the antenna setup used.
- *Channel-quality indication* (CQI), representing the highest modulation-and-coding scheme that, if used, would mean PDSCH transmissions (using the recommended RI and PMI) were received with a block-error rate 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 implementation in the terminal. Also, basing the feedback reports on CQI instead of signal-to-noise ratio also 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 a channel-state report.

Together, a combination of the RI, PMI, and CQI forms a channel-state report. Exactly what is included in a channel-state report depends on the reporting mode the terminal is configured to be in. As mentioned earlier, 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 modes 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 reports, what a terminal delivers to the network are not explicit reports of the downlink channel state. Rather, what the terminal delivers are *recommendations* on the transmission rank and precoding matrix to use, together with an indication of the highest possible modulation-and-coding scheme that the network preferably should not exceed. 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.

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 need to account for other users. Furthermore, the amount of data awaiting transmission in the eNodeB also needs to be accounted for. Obviously, 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 sufficient number of resource blocks can be allocated to the terminal in question.

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, then the precoding is frequency non-selective. The network may also choose to override the transmission rank only, in which case the terminal assumes that a subset of the columns in each of the recommended precoder matrices is used.

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

- Aperiodic channel-state 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 channel-state report is always delivered using the PUSCH – that is, on a dynamically assigned resource.
- Periodic channel-state reports are configured by the network to be delivered with a certain periodicity, possibly as often as once every 2ms, 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 PUSCH⁴ if the terminal has a valid uplink grant and is anyway to transmit on the PUSCH.

Aperiodic and periodic reports, despite both providing estimates on the channel conditions, are quite different in terms of their detailed contents and the 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 PUCCH used for the periodic reports. Furthermore, as aperiodic reports are transmitted on a per-need basis only, the overhead from these reports are 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.

⁴In release 10, a terminal can be configured for simultaneous PUSCH and PUCCH transmission, in which case the periodic channel-state reports can remain on the PUCCH.

Three aperiodic reporting modes are supported in LTE, where each mode has several submodes depending on the configuration:

- 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 obviously 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 transmission mode configured, the PMI and RI are also provided as part of this type of report.
- Configured reports, where the network configured the set of sub-bands the terminal should generate reports for. The terminal reports one wideband CQI reflecting the channel quality across the full downlink carrier bandwidth and one CQI per configured 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 transmission 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](#).

Periodic reports are configured by the network to be delivered with a certain periodicity. The limited, compared to PUSCH, payload supported on the PUCCH also 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 points. 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 are supported in LTE, again with different submodes possible:

- 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

Table 13.1 Possible Aperiodic 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-forming			•			•	•
7	Single-layer trans., DM-RS				•			•
8	Dual-layer trans., DM-RS			•	•		•	•
9	Multi-layer trans., DM-RS			•	•		•	•

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 report 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.

The different aperiodic 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).

Channel-state reports, irrespective of whether they are aperiodic or periodic, need a known reference signal as input to the CSI computation. The cell-specific reference signals can be used for this purpose, and is the only possibility in release 8/9, but in release 10 channel-state reports can also be based on the CSI-RS (see Chapter 10 for a discussion on CSI-RS). Which reference signal to base the channel-state reports upon is linked to the transmission mode. For transmission modes already supported in release 8/9, the cell-specific reference signals are used, while for transmission mode 9, introduced in release 10, the CSI-RS is used.

The discussion above also holds for carrier aggregation, although with some modifications and enhancements as channel-state reports for multiple downlink component carriers are needed.

For aperiodic reporting, the two-bit⁵ CSI request in the downlink control signaling allows for three different types 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 reports for the downlink component carrier associated with the uplink component carrier for which 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.

For periodic reporting, the basic principle in the case of carrier aggregation is to configure the reporting cycles such that the CSI reports for the different component carriers are not transmitted simultaneously.

13.2.6 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. Clearly, 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).

⁵Recall from Chapter 10 that the CSI request field expanded from one to two bits in the case of carrier aggregation.

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-forming		•			•			
7	Single-layer trans., DM-RS	•			•				
8	Dual-layer trans., DM-RS	•	•		•	•			
9	Multi-layer trans., DM-RS	•	•		•	•			

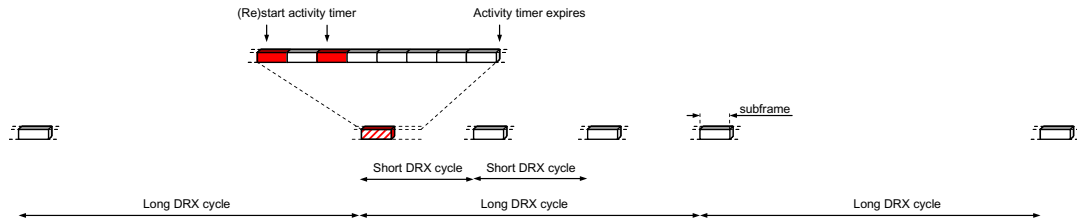
**FIGURE 13.9**

Illustration of DRX operation.

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 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. Obviously, 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.

13.3 INTER-CELL INTERFERENCE COORDINATION

Like all modern mobile-communication technologies, LTE can be deployed with one-cell frequency reuse. Fundamentally, this implies that the LTE transmission structure has been designed so that reliable transmission is also possible for the low signal-to-interference ratios (SIR) that may occur in a reuse-one deployment when the same time–frequency resource is used in neighboring cells (SIR as low as -5 dB or even somewhat lower in the worst case⁶). This is especially true for transmission of critical information such as system information and L1/L2 control signaling. It should be noted that data transmission on DL-SCH can always be made sufficiently reliable by selecting a sufficiently low instantaneous data rate in combination with hybrid-ARQ retransmissions.

Still, a one-cell frequency reuse typically implies relatively large SIR variations over the cell area. As a consequence, the data rates that can be offered to the end-user may also vary substantially, with only relatively low data rates being available at the “cell edge”.

If one was *only* interested in maximizing the data rates that could be offered to users at the cell edge – that is, maximizing the “worst-case-user” quality – a reuse larger than one could actually be preferred. For cell-edge users receiving high interference from neighboring cells, the negative impact of reduced bandwidth availability due to frequency reuse larger than one could be more than compensated for by the higher SIR and corresponding higher achievable data rate per MHz, leading to overall higher achievable data rates. However, the overall system efficiency would be degraded as the

⁶This assumes a relatively homogeneous deployment. In *heterogeneous network deployments* with large differences in the cell output powers, the SIR could be even lower and special means may need to be taken to ensure reliable transmission, as discussed in Section 13.4.

majority of users are not at the cell edge and thus have a relatively good SIR even with one-cell reuse. For such users, any further SIR improvement due to larger reuse would typically not be able to compensate for the reduced bandwidth availability for at least two reasons:

- The relative interference reduction and corresponding increase in SIR would not be as large as for users on the cell edge.
- The achievable data rates do not vary linearly with the SIR and, especially for high SIR, a further SIR improvement may give a relatively small increase in achievable data rate per MHz, as was discussed in Chapter 2.

Furthermore, the assumption of low SIR at the cell edge in the case of one-cell reuse is based on the assumption that there really are transmissions ongoing in the neighboring cells. There are always time instances when there is no or at least limited transmission ongoing in a cell, especially at low-load conditions. The cell-edge SIR may then be relatively good even with a one-cell reuse and a higher reuse with a corresponding reduction in the bandwidth available per cell would not be beneficial even from a cell-edge-user point of view.

Thus, the basic mode of operation should be one-cell reuse, giving each cell access to the over-all available spectrum. However, system performance, and especially the quality for cell-edge users, could be further enhanced if one could at least partly coordinate the scheduling between neighboring cells. The basic principle of such *inter-cell interference coordination* (ICIC) would be to, if possible, avoid high-power transmission on time–frequency resources on which cell-edge users are scheduled in neighboring cells, users that would otherwise experience high interference and correspondingly low data rates. This kind of “selective” interference avoidance would benefit the cell-edge-user quality and could also enhance overall system performance.

The above discussion implicitly assumed downlink transmission with terminals at the cell edge being interfered by downlink transmissions from other cells. The concept of inter-cell interference coordination is equally applicable 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 cell border causing more interference to neighboring cells. The location of the transmitting terminal is still important though, as a terminal closer to the cell site can raise its transmission power to compensate for high interference coming from terminals in neighboring cells, something that may not be possible for cell-edge terminals. Thus, the fundamental goal of uplink inter-cell interference is the same as for the downlink – that is, to coordinate the, in this case, uplink scheduling between cells to avoid simultaneous transmissions from terminals at the cell border in neighboring cells causing severe interference to each other.

In the case of scheduling located at a higher-level node above the eNodeB, coordinated scheduling between cells of different eNodeB would, at least conceptually, be straightforward. However, in LTE there is no higher-level node and scheduling is carried out locally at the eNodeB – that is, in practice at the cell site.⁷ Thus, the best that can be done in the radio-access specifications is to introduce messages that convey information about the scheduling strategy between neighboring eNodeBs using the

⁷As discussed below, in case of so-called centralized RAN (C-RAN) this may change somewhat.

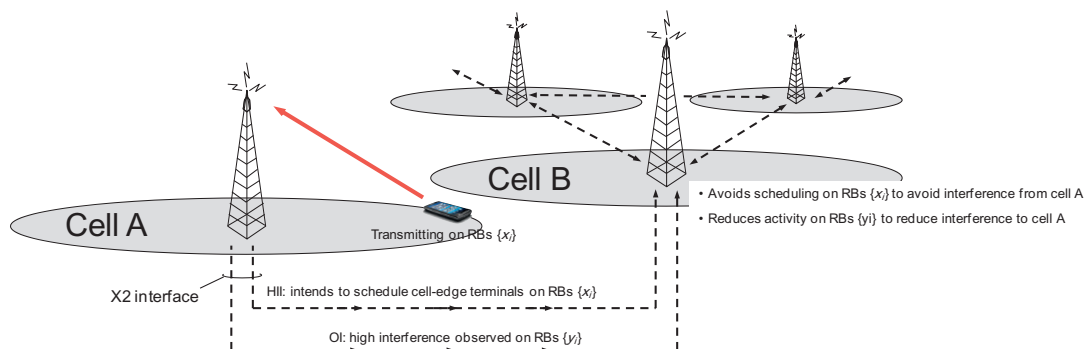
**FIGURE 13.10**

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

X2 interface. An eNodeB can then use the information provided by a neighboring eNodeB as input to its own scheduling process. For LTE, a number of such ICIC-related X2 messages have been defined.

To assist uplink interference coordination, two messages are defined, the *High Interference Indicator* (HII) and the *Overload Indicator* (OI), see also Figure 13.10.

The *High Interference Indicator* provides information about the set of resource blocks within which an eNodeB is likely to schedule transmissions from cell-edge terminals – that is, resource blocks on which a neighboring cell can expect higher interference. Although nothing is explicitly specified on how an eNodeB should react to the HII (or any other ICIC-related 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) is defined to support ICIC operation (see Figure 13.11). 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.

One kind of deployment that has recently received some interest is the *Centralized RAN* (C-RAN) [73], where the baseband processing of the eNodeBs is located in a central office, geographically separate from the actual cell sites. In such a scenario it is more straightforward to coordinate the scheduling between multiple geographically separate cells, either by introducing inter-eNodeB coordination within the central office or by simply deploying a massive eNodeB able to handle all the

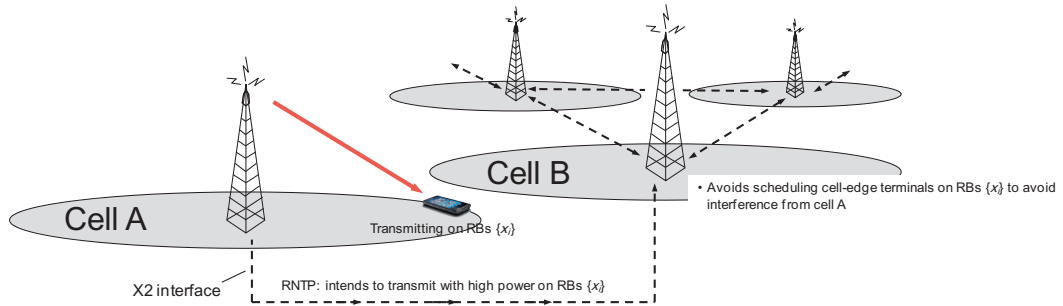


FIGURE 13.11

Illustration of downlink ICIC based on RNTP X2 signaling.

cells connected to the central office. As an eNodeB is anyway just a logical node, in terms of physical realization these two approaches are essentially the same thing.

13.4 HETEROGENEOUS NETWORK DEPLOYMENTS

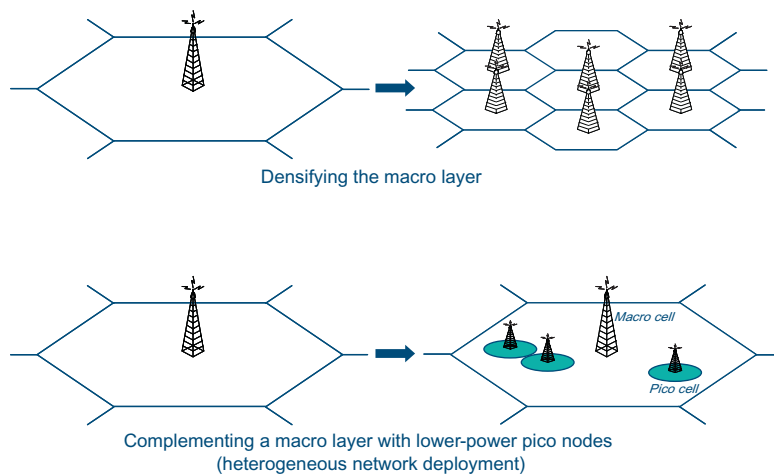
The continuous increase in traffic within mobile-broadband systems and an equally continuous increase in terms of the data rates requested by end-users will impact how cellular networks are deployed in the future. In general, providing very high system capacity (traffic per m^2) and very high per-user data rates will require a densification of the radio-access network – that is, the deployment of additional network nodes. By increasing the number of cells, the traffic per m^2 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 general densification of the macro-cell layer – that is, reducing the coverage area of each cell and increasing the total number of macro-cell sites⁸ – as illustrated in the upper part of Figure 13.12, 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 often less than a few hundred meters in many cases.

An alternative or complement to a uniform densification of the macro-cell layer is to deploy additional lower-power nodes under the coverage area of a macro cell, as illustrated in the lower part of Figure 13.12. In such a *heterogeneous* or *multi-layered* network deployment, the underlaid *pico-cell* layer does not need to provide full-area coverage. Rather, pico sites 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 cell.

Another example of heterogeneous network deployment is the complementary use of so-called *home-eNodeBs*, also often referred to as *femto* base stations. A home-eNodeB corresponds to a small low-power base station deployed by the end-user, typically within the home, and connecting to the operator network using the end-user's wireline broadband connection.

⁸A macro cell is herein defined as a high-power cell with its antennas typically located above rooftop level.

**FIGURE 13.12**

Network densification to enhance system capacity and support higher data rates.

A home-eNodeB is often associated with a so-called *Closed Subscriber Group* (CSG), with only users that are members of the CSG being allowed to access the home-eNodeB. Thus, users not being members of the CSG have to access the radio-access network via the overlaid macro-cell layer even when in close proximity to a home-eNodeB. As discussed further below, this causes additional interference problems with home-eNodeB deployments, beyond those of ordinary heterogeneous network deployments.

13.4.1 Interference Handling in a Heterogeneous Deployment

In itself, the use of heterogeneous network deployments in mobile-communication systems – that is, complementing a macro layer with lower-power pico nodes to increase traffic capacity and/or achievable data rates in specific areas – is nothing new and has been used for a relatively long time in, for example, GSM networks. Different sets of carrier frequencies have then typically been used in the different cell layers, thereby avoiding strong interference between the layers.

However, for a wideband radio-access technology such as LTE, using different carrier frequencies for different cell layers may lead to an undesirable spectrum fragmentation. As an example, for an operator having access to 20 MHz of spectrum, a static frequency separation between two cell 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 obviously reduce the maximum achievable data rates within each cell layer. Also, assigning a substantial part of the overall available spectrum to a cell layer with relatively low traffic may lead to inefficient spectrum utilization. Thus, with a wideband high-data-rate system such as LTE, it should preferably be possible to deploy a multi-layer network structure with the same spectrum being available in the different cell layers.

However, the simultaneous use of the same spectrum in different cell layers will obviously imply interference between the layers. Due to the difference in transmit power between the nodes of a

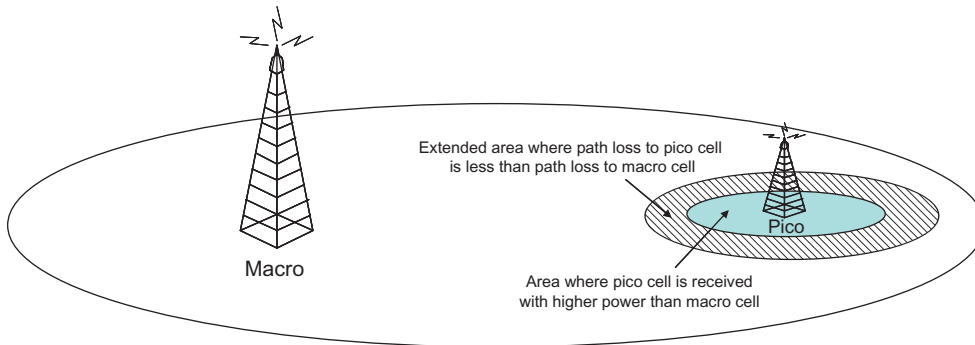
**FIGURE 13.13**

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

heterogeneous network deployment, such *inter-layer interference* may be more severe compared to inter-cell interference between cells of the same layer.

The characteristics of the inter-layer interference in a heterogeneous network deployment will also depend on the exact cell-selection strategy being used. Conventional cell selection is typically based on terminal measurements of the received power of some downlink signal, more specifically the cell-specific reference signals in the case of LTE. However, in a heterogeneous network deployment with cells with substantially different transmit power, including different power of the reference signals, selecting the cell that is received with the highest power implies that the terminal may often select the higher-power macro cell even if the path loss to a pico cell is significantly smaller. This will obviously 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 cell with the highest received power in a heterogeneous network deployment. Although the high-power macro cell is received with higher power, this is at least partly due to the higher macro-cell transmit power. In that case, transmission from the macro cell is associated with a higher “cost” in terms of interference to other cells. Expressed alternatively, a transmission from the macro cell will prohibit the use of the same physical resource in *any* of the underlaid pico cells.

Alternatively, at the other extreme, cell selection could be based on estimates of the (uplink) path loss. In practice this can be achieved by applying a cell-specific offset to the received-power measurements used in conventional cell selection, an offset that would compensate for the difference in cell transmit power.⁹ Such a cell-selection strategy would extend the area in which the pico-cell is selected, as illustrated in Figure 13.13. It is therefore also sometimes referred to as *range extension*.

Selecting the cell to which the path loss is the smallest – that is, applying range extension – 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 in other pico cells, thereby also improving downlink system efficiency.

⁹Such offsets or *cell-selection biasing* is already supported by LTE.

However, due to the difference in transmit power between the cells of the different cell layers, there is an area (illustrated by the dashed region in Figure 13.13) where the pico cell is selected while, at the same time, the downlink transmission from the macro cell is received with substantially higher power than the actual desired downlink transmission from the pico cell. Within this area, there is thus potential for severe downlink inter-cell interference from the macro cell to pico-cell terminals, interference that may require special means to handle.

In the following we will assume that the transmissions from the macro cell and its underlaid pico cells are relatively time aligned, implying, for example, that the macro-cell PDCCH (PDSCH) transmissions will interfere with PDCCH (PDSCH) transmissions in the pico cell and vice versa.¹⁰

The interference from PDSCH transmissions in the macro cell to lower-power PDSCH transmissions within a pico cell can be relatively straightforwardly handled by scheduling coordination between the cells according to the same principles as inter-cell interference coordination between cells within a layer, as described in Section 13.3. As an example, the overlaid macro cell could simply avoid high-power PDSCH transmission in resource blocks in which a terminal in the high-interference region of a pico cell 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 macro cell and the underlaid pico cells can be coordinated. It should also be noted that, for a pico cell on the border between two macro cells, it may be necessary to coordinate scheduling simultaneously with both macro cells.

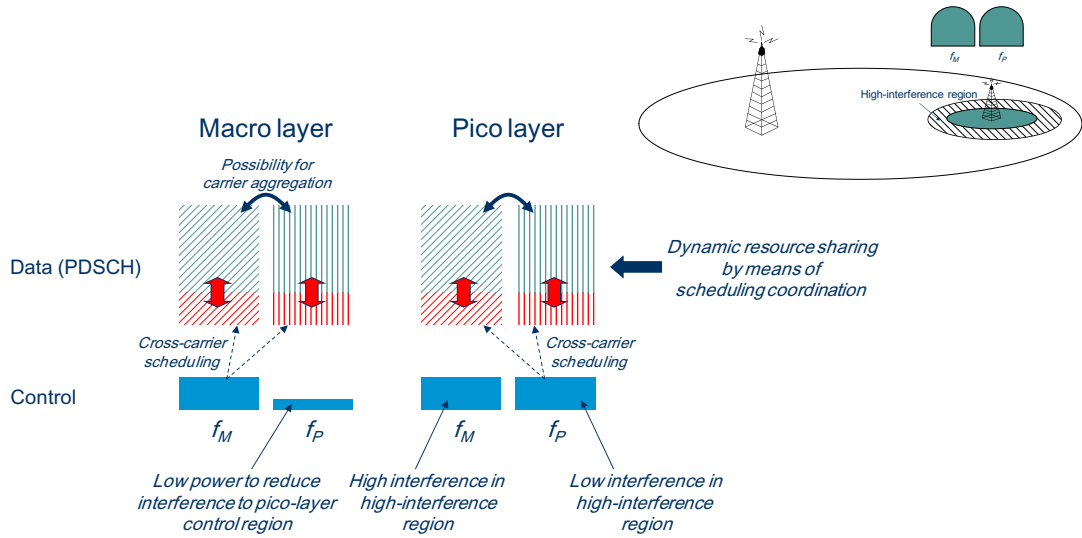
Less obvious is how to handle interference due to the macro-cell transmissions that cannot be dynamically scheduled, such as the L1/L2 control signaling (PDCCH, PCFICH, and PHICH). Within a cell layer, for example between two macro cells, interference between such transmissions is not a critical issue 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. However, in a heterogeneous network deployment with extensive range expansion, the downlink interference from the macro cell can be more than 15 dB above the signal received from the pico cell, corresponding to an SIR below -15 dB. In such low SIR it will not be possible to correctly decode even the very robust control channels. Instead, means are needed to avoid macro-to-pico interference for the L1/L2 control channels.

One possible approach to handle the extended interference in a heterogeneous network deployment with range-extended pico cells is to use carrier aggregation in combination with cross-carrier scheduling, as described in Section 10.4.6. The basic principle of such an approach is illustrated in Figure 13.14 for the case of two layers (macro and pico) and two downlink carriers (f_M and f_P).

In terms of data (PDSCH) transmission, both carriers are available in both cell layers and interference between the layers is handled by “conventional” inter-cell interference coordination, as discussed above. As already mentioned, such interference coordination can be more or less dynamic depending on the time scale on which the macro cell and its underlaid pico cells 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.

On the other hand, in terms of L1/L2 control signaling there is at least partly a more semi-static frequency separation between the layers. More specifically, the macro cell should avoid high-power

¹⁰This also assumes the same size of the control (PDCCH) region for macro and pico cells.

**FIGURE 13.14**

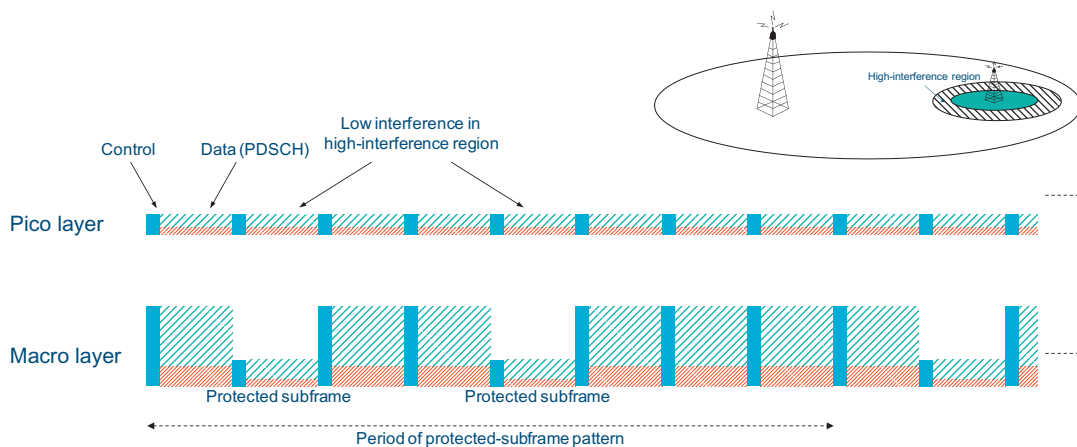
Carrier-aggregation approach to interference avoidance in a heterogeneous network deployment.

transmission within the control region on carrier f_P . In this way, interference to the control region of an underlaid pico cell is reduced on this carrier and the pico cells can use the carrier for control signaling to terminals in the high-interference region. Due to the possibility for cross-carrier scheduling, even if the macro cell only transmits control signaling, including downlink scheduling assignments, on carrier f_M , DL-SCH transmission on PDSCH can still be scheduled on both carriers as well as an aggregation of these. The same is true for the pico cell; even if the pico cell can only use carrier f_P for transmission of scheduling assignments to terminals in the high-interference region, DL-SCH transmissions can still be scheduled on both carriers.

It should be noted that, for terminals not in the high interference region, the pico cell could also use carrier f_M for L1/L2 control signaling. Similarly, the macro cell could use also carrier f_P for control signaling, assuming a reduced transmit power is used. Thus, the macro cell could use carrier f_P for lower-power control signaling, for example for terminals close to the macro-cell site.

The main drawback with the above-described carrier-aggregation-based approach to interference handling in a heterogeneous network deployment is that it requires terminal support for carrier aggregation in order to ensure full flexibility in spectrum usage. For terminals not capable of carrier aggregation, for example all pre-release-10 terminals, the frequency separation of control signaling between the layers implies a corresponding frequency separation in terms of data transmission as, for such terminals, scheduling assignments on carrier f_M can only schedule PDSCH transmissions on the same carrier. Thus, for such terminals there would be spectrum fragmentation.

If the carrier-aggregation approach described above cannot be used, one can instead apply more conventional interference coordination between the different layers, similar to interference coordination within one cell layer as described in Section 13.3, but extended to also cover interference between control-channel transmissions of the different cell layers.

**FIGURE 13.15**

Single-carrier approach to interference avoidance in a heterogeneous network deployment.

With such an approach, the same carrier is used for transmission in both the macro layer and the underlaid pico cells. However, the power of macro-cell transmissions is restricted in some subframes. In contrast to conventional ICIC as described above, this restriction would not only apply to data (PDSCH) transmission, but also to the control region, as illustrated in Figure 13.15. In these *protected subframes* pico-cell terminals will thus experience less macro-cell interference for both data and control, and the pico cell can also use these subframes for transmission to terminals in the high-interference region. In the example illustrated in Figure 13.15, two of eight subframes are configured as protected subframes.

To support this second approach to interference handling in a heterogeneous network, signaling of *protected-subframe patterns* – that is, information about the set of protected subframes – is supported between eNodeBs of different cell layers. Note that the set of protected subframes could be more or less dynamic, once again depending on the time scale on which the macro cell and the underlaid pico cells can be coordinated.

It should be noted that the macro cell must not necessarily completely avoid control-signaling transmission 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 not to 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 high-interference region of the pico cell could be kept at an acceptable level. However, the signaling of protected-subframe patterns is also defined so 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.

Clearly, the interference experienced by pico-cell terminals 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 subframes. Thus, as part of the enhanced support for heterogeneous network deployments, it is possible to configure a terminal with different *CSI-measurement subsets*, confining the terminal CSI measurements to subsets of the full set of subframes with terminals reporting CSI for each subset separately. The corresponding CSI reports should then preferably reflect the interference level in protected and non-protected subframes respectively. However, as mentioned above, the set of protected subframes may vary dynamically and it may not be feasible to update the CSI measurement sets accordingly. Thus, in practice, the CSI-measurement set corresponding to protected subframes may consist of only a subset of the protected subframes, a set that should be relatively static. Similarly, the CSI-measurement set corresponding to non-protected subframes may consist of only a subset of the non-protected subframes.

It should also be noted that, in some cases, a pico cell could be located at the border between, and suffer interference from, two macro cells. If the macro cells have different 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.

13.4.2 Interference Coordination in the Case of Home-eNodeB

In the case of home-eNodeB with CSG there are additional interference issues due to the fact that a terminal can be very close to the home-eNodeB and still have to communicate with the overlaid macro cell. Such a terminal may then be severely interfered on the downlink by any home-eNodeB transmission and may also cause severe uplink interference to the home-eNodeB. In principle, this can 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, possibly extended by the carrier-aggregation approach.

A key difference in this case, though, is 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.

A further complicating factor in a home-eNodeB scenario is that there are obvious limitations in the coordination between the home-eNodeB and an overlaid macro layer due to limited backhaul capabilities of a home-eNodeB. In essence, there is no X2 interface to an Home-eNodeB and, in practice, the configuration of protected /non-protected subframes must be more or less static, something that could obviously affect the overall system efficiency.

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.

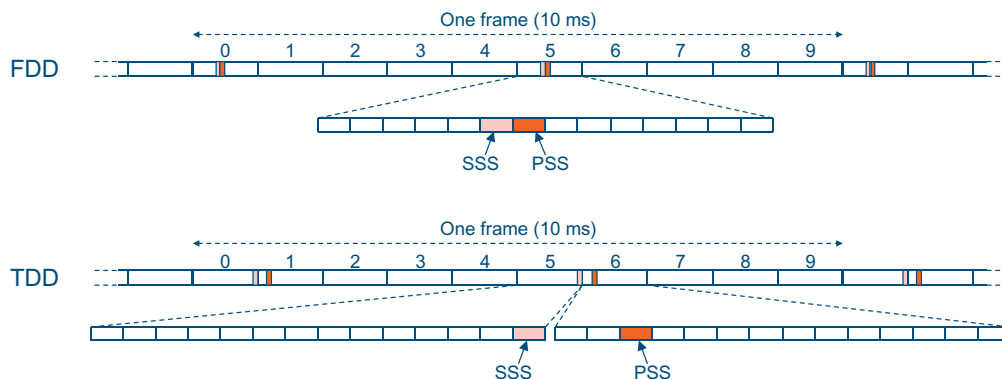
14.1.1 Overview of LTE Cell Search

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 parts:

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

As already mentioned (e.g. in Chapter 10), there are 504 different physical-layer cell identities defined for LTE, where each cell identity corresponds to one specific downlink reference-signal sequence. The set of physical-layer cell identities is further divided into 168 *cell-identity groups*, with three cell identities within each group.

**FIGURE 14.1**

Time-domain position of PSS and SSS for FDD and TTD.

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 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.

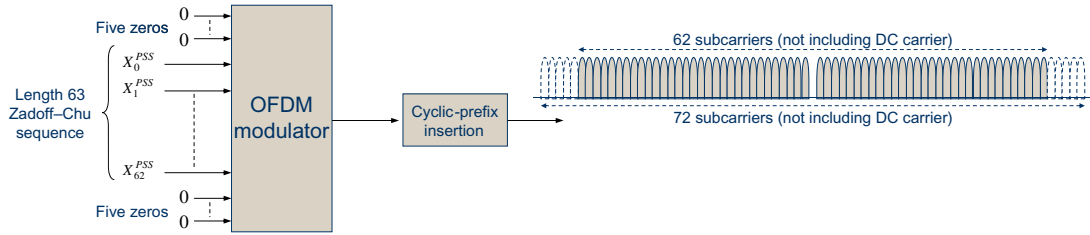
There are several reasons for the difference in the positions of the synchronization signals for FDD and TDD. For example, the difference allows for the detection of the duplex scheme used on a carrier if this is not known in advance.

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.
- 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); and
- the cell-identity group (168 alternatives).

**FIGURE 14.2**

Definition and structure of PSS.

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.

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 (SSS_1 in subframe 0 and SSS_2 in subframe 5) are different, implying that, from the detection of a single SSS, the terminal can determine whether SSS_1 or SSS_2 has been detected and thus determine frame timing.

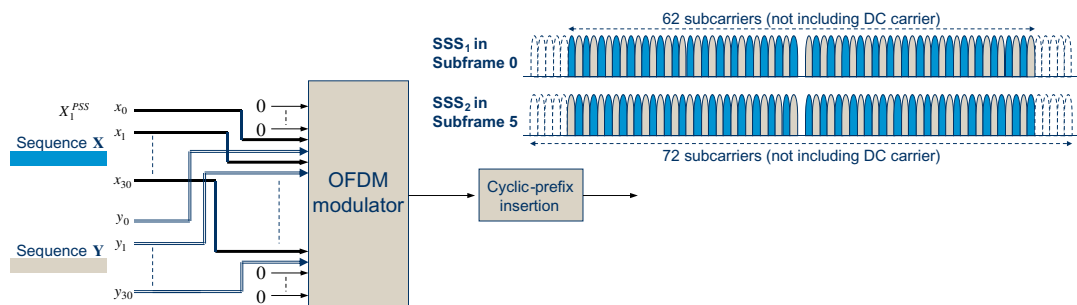
Once the terminal has acquired frame timing and the physical-layer cell identity, it has identified the cell-specific reference signal. The behavior is slightly different depending on whether it is an initial cell search or cell search for the purpose of neighboring cell measurements:

- In the case of initial cell search – that is, 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 – that is, 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,¹ 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 extended with five zeros at the edges and mapped to the center 73 subcarriers (center six resource blocks) as illustrated in Figure 14.2. It should be noted though 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).

¹As discussed in Chapter 9, the specifications use the terms primary and secondary cells instead of primary and secondary component carriers.

**FIGURE 14.3**

Definition and structure of SSS.

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 SSS (SSS_1 in subframe 0 and SSS_2 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 SSS_2 is different from the set of values applicable for SSS_1 to allow for frame-timing detection from the reception of a single SSS.

The structure of the two SSS is illustrated in Figure 14.3. SSS_1 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, SSS_2 is based on exactly the same two sequences as SSS_1 . 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 SSS_1 has then been selected so that a swapping of the two sequences in the frequency domain is not a valid combination for SSS_1 . Thus, the above requirements are fulfilled:

- The set of valid combinations of X and Y for SSS_1 (as well as for SSS_2) are 168, allowing for detection of the physical-layer cell identity.
- As the sequences X and Y are swapped between SSS_1 and SSS_2 , 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 and uplink power control, 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 in Chapter 8, BCCH can be mapped to both BCH and DL-SCH depending on the exact BCCH information.

14.2.1 MIB 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:

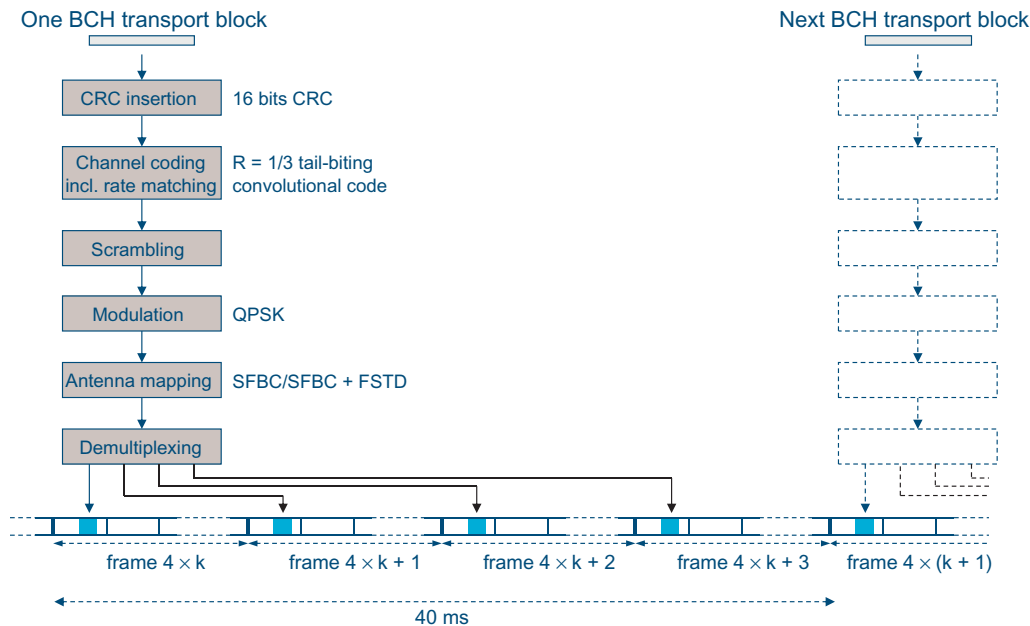
- *Information about the downlink cell bandwidth.* Four bits are available within the MIB to indicate the downlink bandwidth. Thus, up to 16 different bandwidths, measured in number of resource blocks, can be defined for each frequency band.
- *Information about the PHICH configuration of the cell.* 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 which, in turn, is needed to receive DL-SCH. 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

**FIGURE 14.4**

Channel coding and subframe mapping for the BCH transport channel.

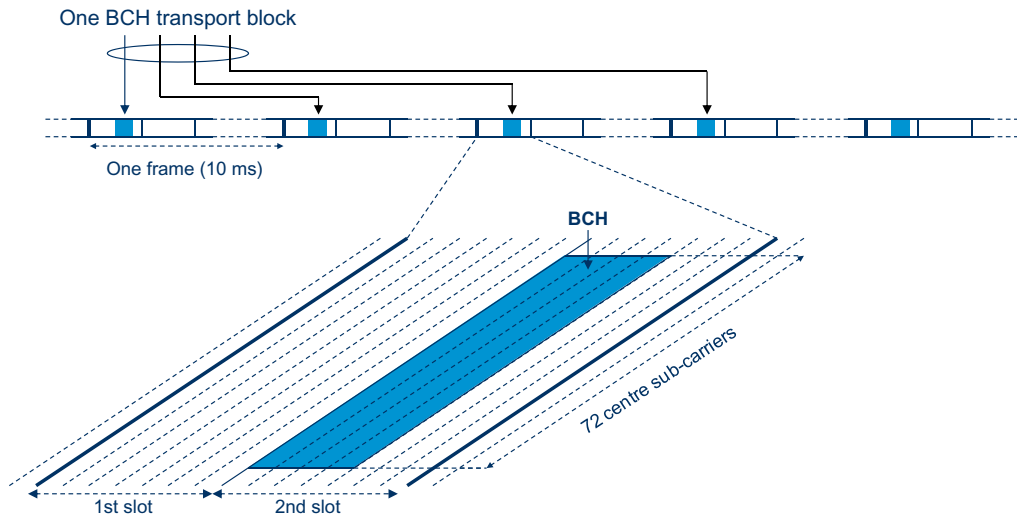
code used for all other transport channels, is the small size of the BCH transport block. With such small blocks, tail-biting convolutional coding actually outperforms Turbo coding. The channel coding is followed by rate matching, in practice repetition of the coded bits, and bit-level scrambling. QPSK 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 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.² 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,

²Not including the DC carrier.

**FIGURE 14.5**

Detailed resource mapping for the BCH transport channel.

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.

Clearly, 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 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 40ms timing or, equivalently, the two least significant bits of the SFN.³ This is the reason why these bits do not need to be explicitly included in the MIB.

³BCH 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.

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 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 15).

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-1,⁴ 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 obviously 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.

⁴Strictly 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.

**FIGURE 14.6**

Example of mapping of SIBs to SIs.

- 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 160ms. SIB3 and SIB4 are multiplexed into SI-3 with a transmission period of 320ms, whereas SIB5, which also requires a transmission period of 320ms, is mapped to a separate SI (SI-4). Finally, SIB6, SIB7, and SIB8 are multiplexed into SI-5 with a transmission period of 640ms. Information about the detailed SIB-to-SI mapping, as well as the transmission period of the different SIs, is provided in SIB1.

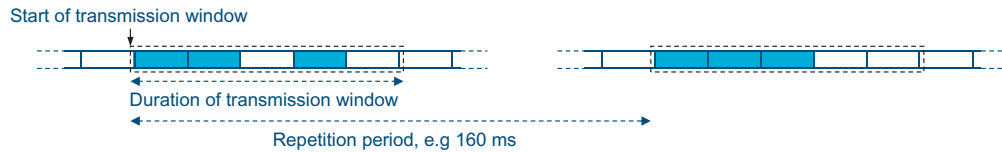
Regarding the more detailed transmission of the different system-information messages there is a difference between the transmission of SI-1, corresponding to SIB1, and the transmission of the remaining SIs.

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 the 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

**FIGURE 14.7**

Transmission window for the transmission of an SI.

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.

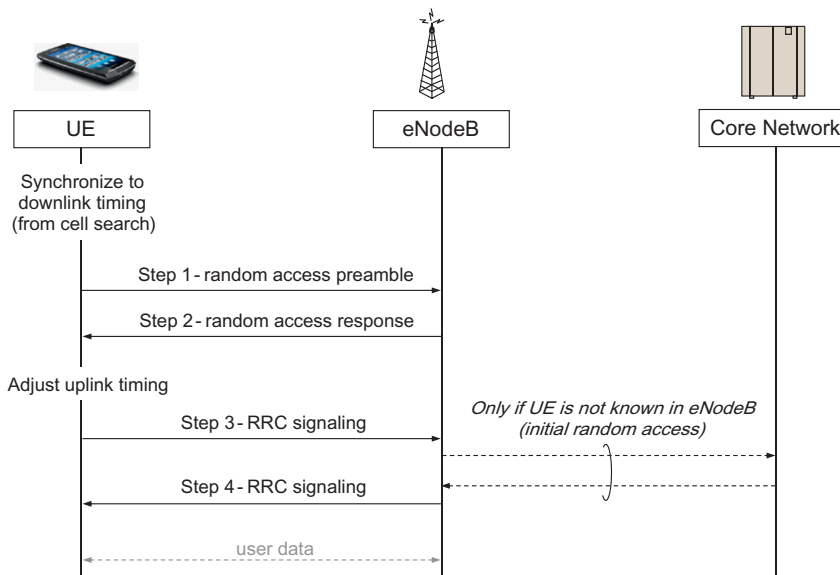
Strictly speaking the single transport block containing the SI is not transmitted over multiple subframes. Rather, the subsequent SI transmissions are seen as autonomous hybrid-ARQ retransmissions of the first SI transmission – that is, retransmissions taking place without any explicit feedback signaling 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;

**FIGURE 14.8**

Overview of the random-access procedure.

- 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).

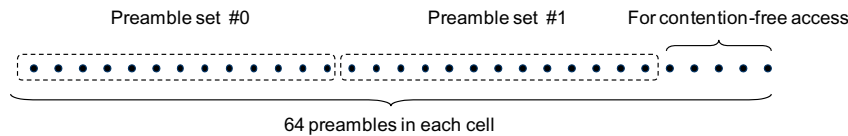
Acquisition of uplink timing is a main objective for all the cases above; when establishing an initial radio link (that is, when moving from RRC_IDLE to RRC_CONNECTED), the random-access procedure also serves the purpose of assigning a unique identity, the C-RNTI, to the terminal.

A terminal may perform random access on its primary component carrier only.⁵ Either a contention-based or a contention-free scheme can be used.

Contention-based random access uses a 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,

⁵The primary component carrier is UE specific, as already discussed; hence, from an eNodeB perspective, random access may occur on multiple component carriers.

**FIGURE 14.9**

Preamble subsets.

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.

Contention-free random access can only be used for re-establishing uplink synchronization upon downlink data arrival, handover, and positioning. Only the first two steps of the procedure above are used as there is no need for contention resolution in a contention-free scheme.

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 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

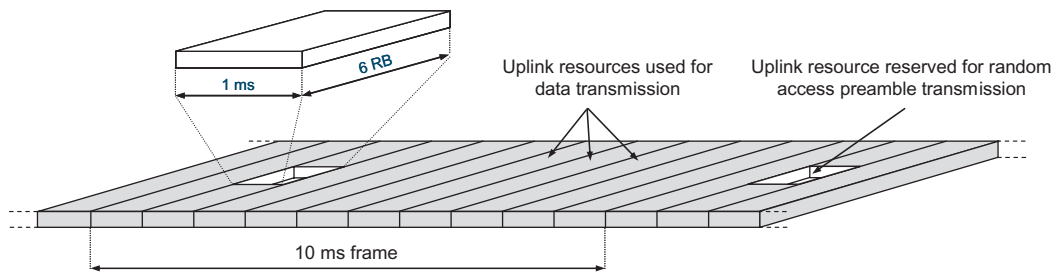


FIGURE 14.10

Principal illustration of random-access preamble transmission.

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 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 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

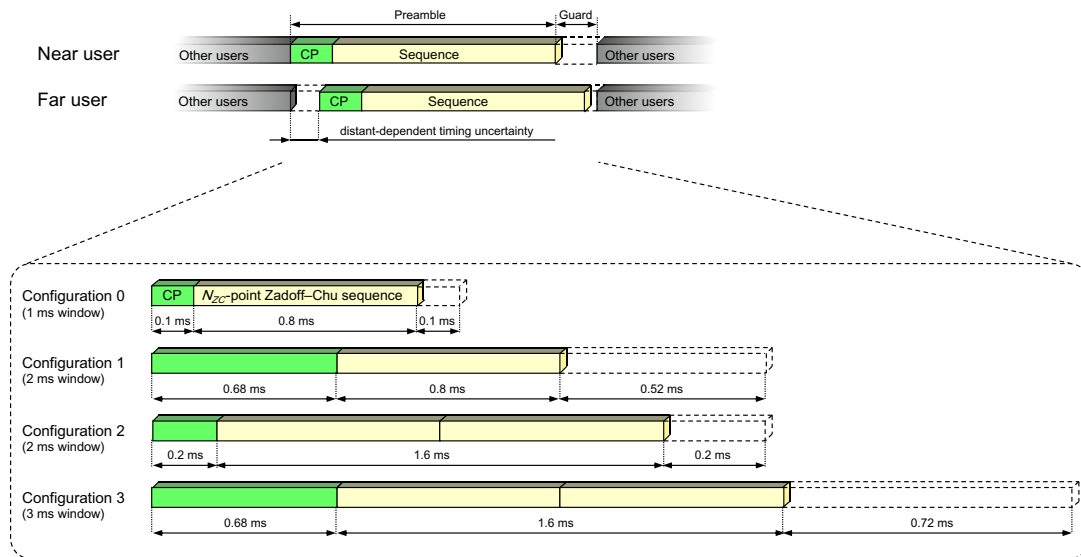


FIGURE 14.11

Different preamble formats.

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 timing⁶ 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

⁶The start of an uplink frame at the terminal is defined relative to the start of a downlink frame received at the terminal.

of the guard period. With a preamble sequence length of approximately 0.8ms, there is 0.1ms cyclic prefix and 0.1ms guard time. This allows for cell sizes up to 15km 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. 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.

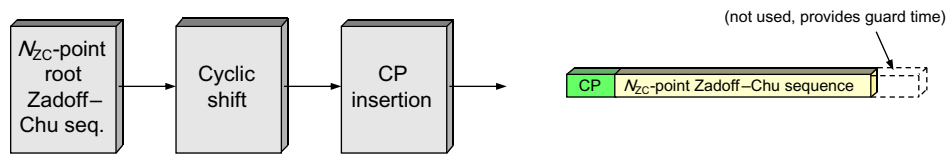
The LTE random-access mechanism allows power ramping where the actual PRACH transmission power is increased for each unsuccessful random-access attempt. For the first attempt, the PRACH transmission power is set to the initial PRACH power. In most cases, this is sufficient for the random-access attempts 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, $X_{ZC}^{(u)}(k)$, $[N_{ZC}/N_{CS}]$ cyclically shifted⁷ 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

⁷The 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.

**FIGURE 14.12**

Random-access preamble generation.

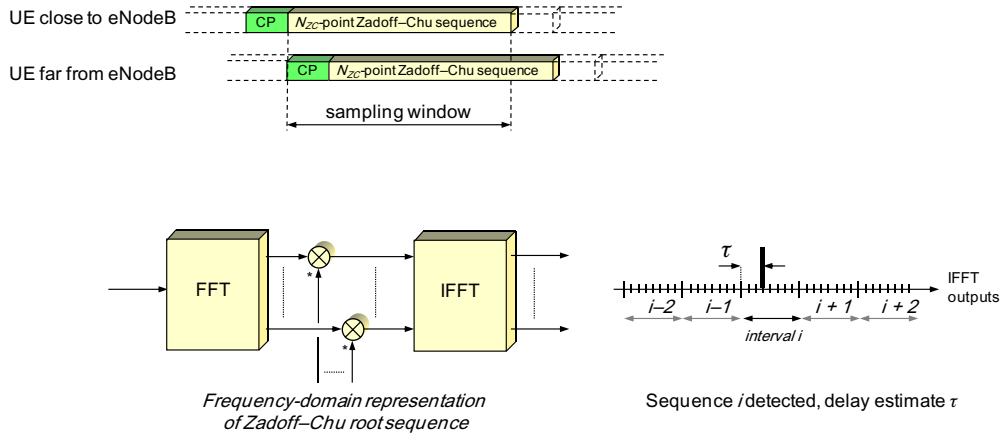
in Figure 14.12. Although the figure illustrates generation in the time domain, frequency-domain generation can equally well be used in an implementation.

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 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 restriction 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 broadcast as part of the system information in the cell.

**FIGURE 14.13**

Random-access preamble detection in the frequency domain.

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 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. 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 sufficiently 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. 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 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.⁸ 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.

UE-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, in the fourth step, each terminal receiving the downlink message 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 fourth step and the identity transmitted as part of the third step will declare the random-access procedure successful. If the terminal has not yet been assigned a C-RNTI, the TC-RNTI from the second step is promoted to the C-RNTI; otherwise the terminal keeps its already assigned C-RNTI.

The contention-resolution message is transmitted on the DL-SCH, using the temporary identity from the second step for addressing the terminal on the L1/L2 control channel. Since uplink synchronization has already been established, hybrid ARQ is applied to the downlink signaling in this step. Terminals with a match between the identity they transmitted in the third step and the message received in the fourth step will also transmit a hybrid-ARQ acknowledgement in the uplink.

Terminals that 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. Obviously, 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.

⁸The terminal identity is included as a MAC control element on the UL-SCH.

Table 14.1 Paging Cycles and Paging Subframes									
		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

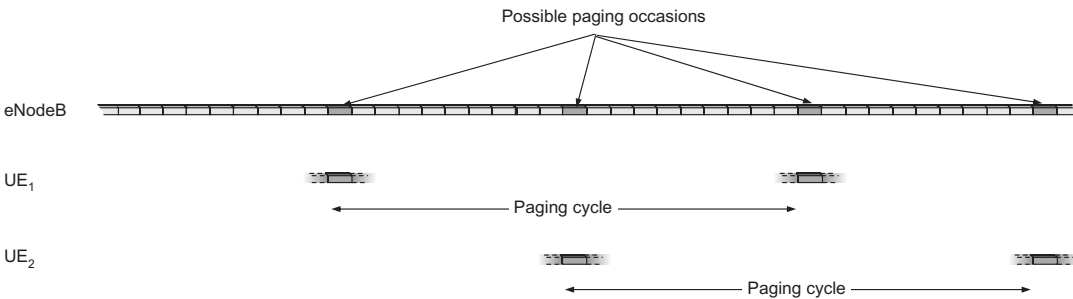


FIGURE 14.14

Illustration of paging cycles.

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. Obviously, 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 UE-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 UE-specific paging cycle. The identity used is the so-called IMSI, an identity coupled to the 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 sub-frame 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.

Multimedia Broadcast/Multicast Services

15

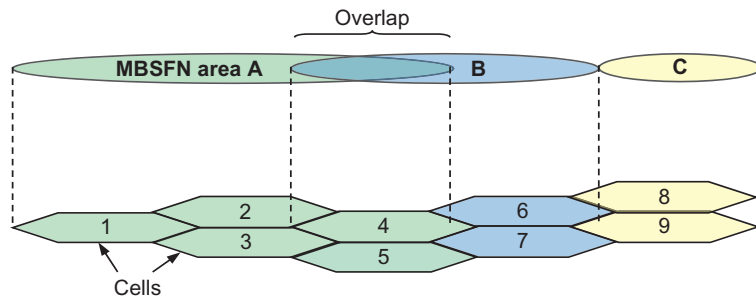
In the past, cellular systems have mostly focused on transmission of data intended for a single user and not on multicast/broadcast services. Broadcast networks, exemplified by the radio and TV broadcasting networks, have on the other hand focused on covering very large areas with the same content and 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, 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.

**FIGURE 15.1**

Example of MBSFN areas.

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.

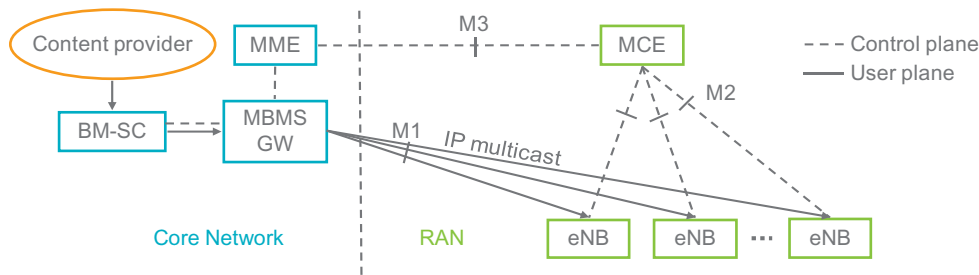
Providing for power-efficient reception in the terminal in essence implies that the structure of the overall transmission should be such that data for a service-of-interest is provided in short high-data-rate bursts rather than longer low-data-rate bursts. This allows the terminal to occasionally wake 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.

15.1 ARCHITECTURE

An *MBSFN area* is a specific area where one or several cells transmit the same content. For example, in Figure 15.1, 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 15.1, 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 obviously 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 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 15.2, the MCE¹ can control multiple eNodeBs, each handling one or more cells.

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.

**FIGURE 15.2**

LTE MBMS architecture.

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.

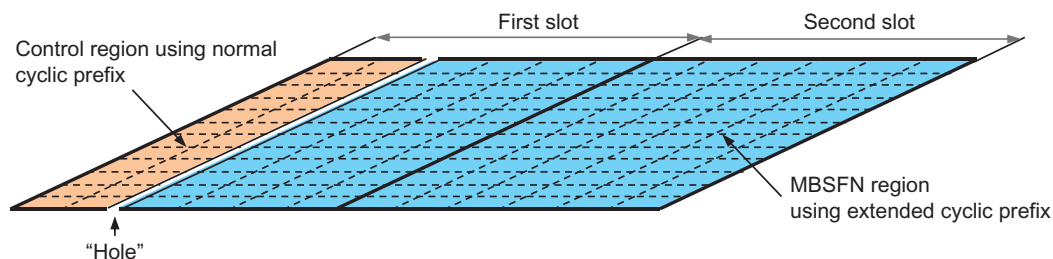
15.2 OVERALL CHANNEL STRUCTURE AND PHYSICAL-LAYER PROCESSING

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. Obviously, 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.

**FIGURE 15.3**

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

One or several MTCHs and, if applicable,² one MCCH are multiplexed at the MAC layer to form an MCH transport channel. As already 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:

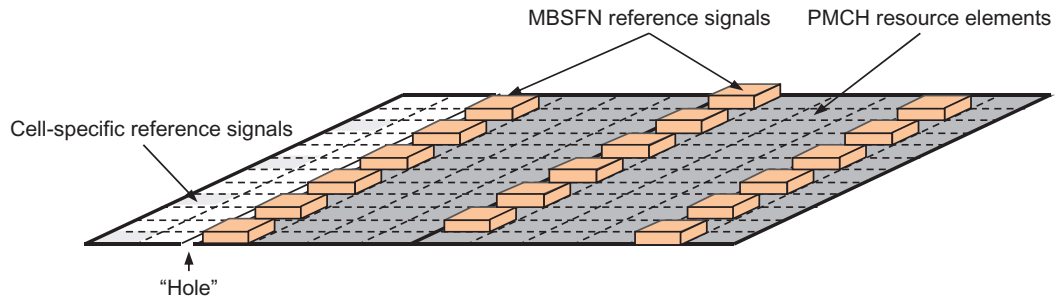
- In the case of MBSFN transmission, the same data is to be transmitted with the same transport format using the same physical resource from multiple cells typically belonging to different eNodeBs. Thus, the MCH transport format and resource allocation cannot be dynamically adjusted by the eNodeB. As 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.

The MCH is mapped to the PMCH physical channel and transmitted in MBSFN subframes, illustrated in Figure 15.3. 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.

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.

³As 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 15.4**

Reference-signal structure for PMCH reception.

As already 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 an MBSFN subframe, as illustrated in Figure 15.3. 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 15.4. 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.

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 15.4, 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.

15.3 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 15.5.

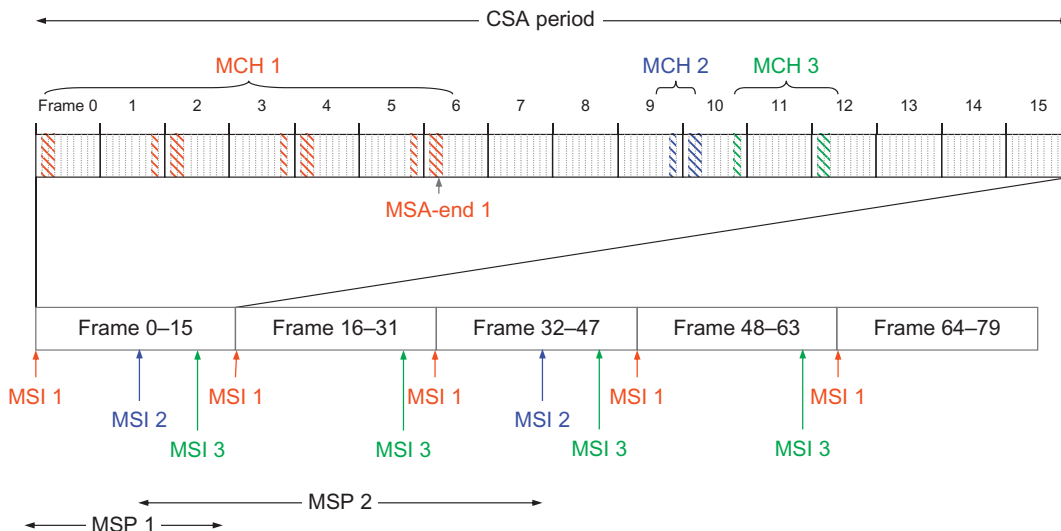


FIGURE 15.5

Example of scheduling of MBMS services.

Obviously, 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 16. Furthermore, the allocation of MBSFN subframes for MCH transmission should obviously 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 15.5), while the remaining subframes are not used for MBMS 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.

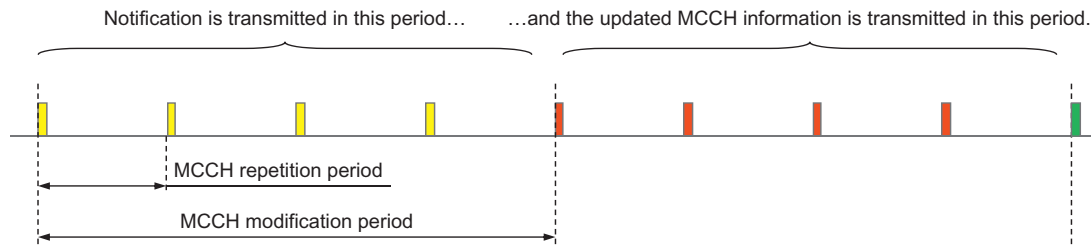
The fact that the transport format is signaled as part of the MCCH implies that the MCH transport format may differ between MCHs but must remain constant across subframes used for the same MCH. The only exception is subframes used for the MCCH and MSI, where the MCCH-specific transport format, signaled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

In the example in Figure 15.5, 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 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,

**FIGURE 15.6**

MCCH transmission schedule.

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 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 is changed, which can only be done at the beginning of a new modification period, as shown in Figure 15.6, 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.

Relaying

16

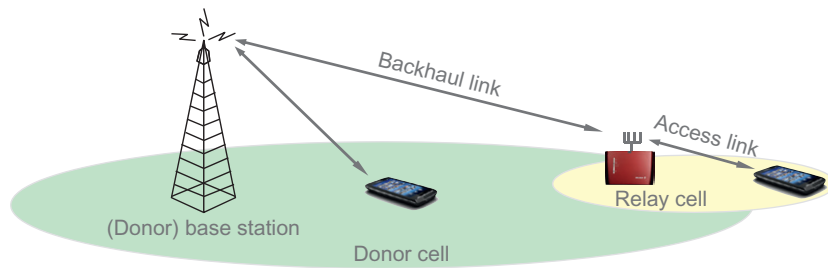
The possibility of a terminal communicating with the network, and the data rate that can be used, depends on several factors, the path loss between the terminal and the base station being one. The link performance of LTE is already quite close to the Shannon limit and from a pure link-budget perspective, 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 are 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.

**FIGURE 16.1**

Access and backhaul links.

16.1 RELAYS IN LTE

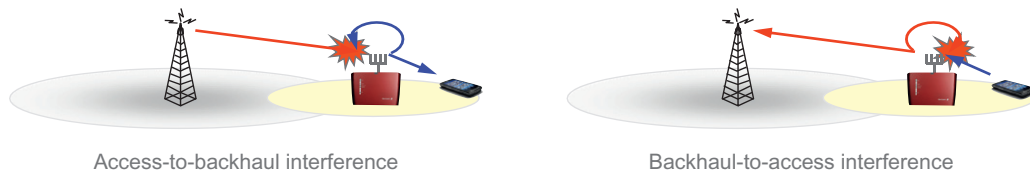
LTE release 10 introduces support for a decode-and-forward relaying scheme (repeaters require no additional standardization support other than RF requirements and are available already in release 8). A basic requirement in the development of LTE relaying solutions was that the relay should be transparent 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 release-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 16.1.

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 100dB, 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 16.2. 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 *out-band* and *in-band* 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

**FIGURE 16.2**

Interference between access and backhaul links.

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 in the following. 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.

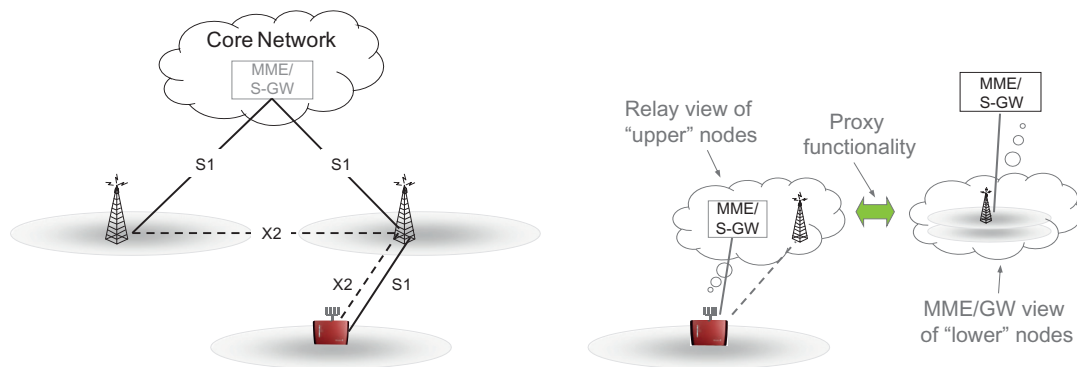
16.2 OVERALL ARCHITECTURE

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 16.3](#). 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

**FIGURE 16.3**

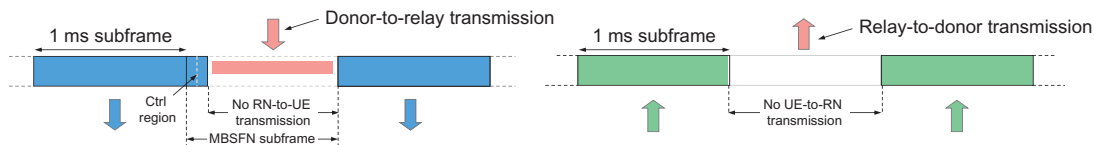
LTE relaying architecture.

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.

16.3 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).

An obvious 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 release-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.

**FIGURE 16.4**

Multiplexing between access and backhaul links.

Fortunately, already 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 16.4, 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, is introduced in release 10.

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].

16.3.1 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 4ms after a downlink subframe. This is chosen to match the access-link hybrid-ARQ timing relations, where an uplink subframe follows 4ms 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.

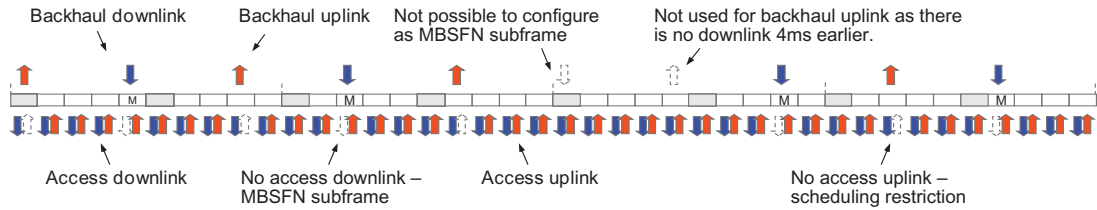
In TDD, the relay node may not be able to receive hybrid-ARQ feedback on PUCCH in uplink subframes used for transmission on the backhaul link. One possibility is to restrict the downlink scheduler such that no terminals transmit PUCCH in uplink subframes the relay cannot receive. However, such a restriction 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.

16.3.2 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 8ms 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 10ms structure while the hybrid-ARQ timing follows an 8ms periodicity, there is an inherent mismatch between the two. Hence, as illustrated in Figure 16.5, some backhaul subframes may be spaced 16ms apart, as subframes 0, 4, 5, and 9 cannot be configured as MBSFN subframes (see Chapter 9). Uplink backhaul subframes follow 4ms after a downlink backhaul subframe, following the principle discussed in the previous paragraph.

For TDD, there is an inherent 10ms component in the hybrid-ARQ timing relations, which matches the 10ms 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 MBSFN subframes. For configuration 5, there

**FIGURE 16.5**

Example of backhaul configuration for FDD.

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 16.1.

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, obviously 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 16.6, taking into account 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.

16.3.3 Backhaul Downlink Control Signaling

The gap during which the relay can receive transmissions from the donor cell is, as seen in Figure 16.4, 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, is introduced in release 10.

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, which is beneficial, for example, in heterogeneous deployments.

Table 16.1 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	8	9
0	1	1:1					D				U	
1						U						D
2							D				U	D
3						U	D					D
4	2	2:2				U	D				U	D
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	D
11						U				D		D
12	4	3:1				U				D	D	D
13						U						D
14						U				D		D
15						U					D	D
16	6	3:1				U				D	D	D
17						U	D			D	D	D
18							U					D

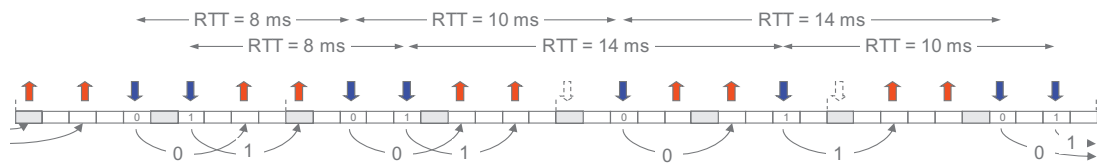
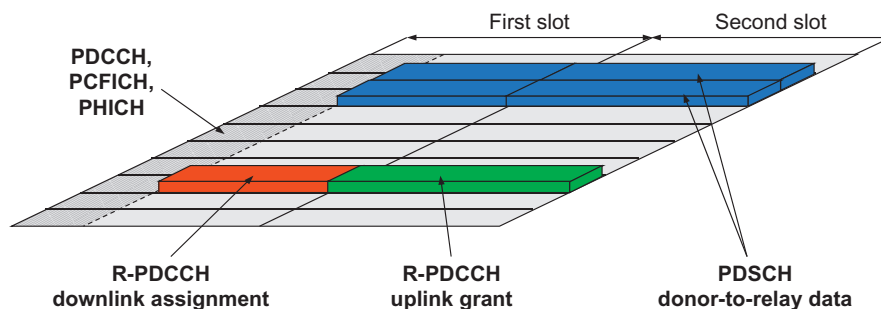


FIGURE 16.6

Example of hybrid-ARQ process numbering for FDD.

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 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.

**FIGURE 16.7**

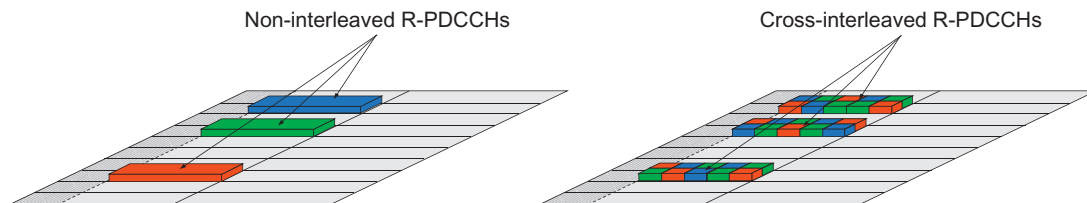
Example of R-PDCCH transmission.

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 16.7). 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, for obvious reasons, different. Two different mapping methods, illustrated in Figure 16.8, are supported:

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

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

**FIGURE 16.8**

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

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. Obviously, 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 also 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 16.9 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 overhead is less of an issue, the

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

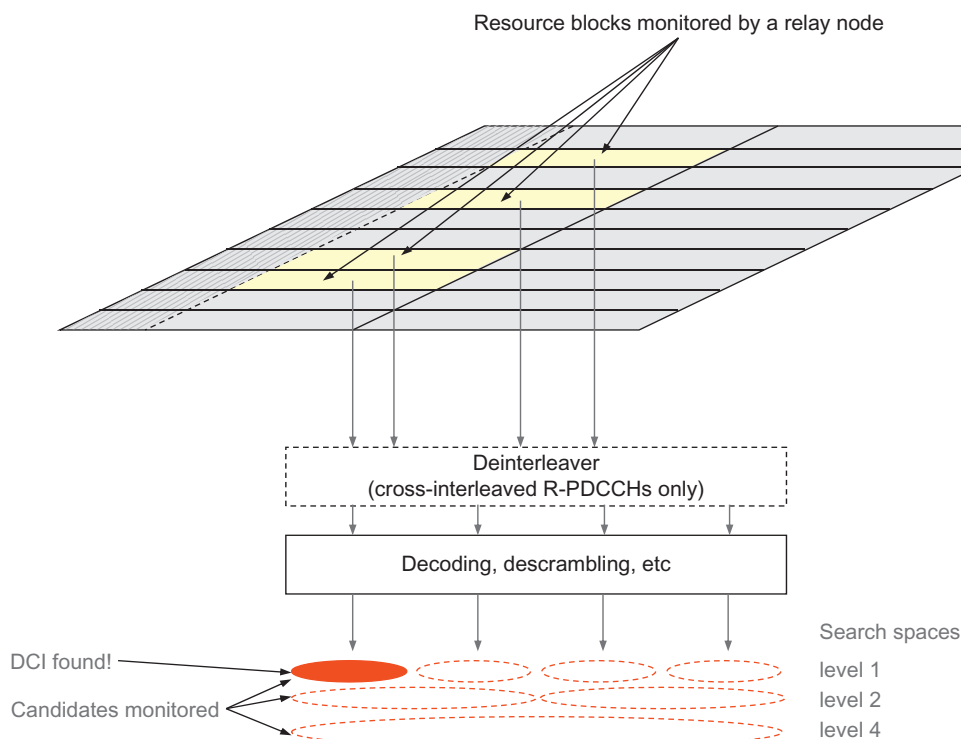
**FIGURE 16.9**

Illustration of the principle of R-PDCCH monitoring.

PHICH was omitted from the backhaul in order to simplify the overall design. Retransmissions are still supported through the use of the R-PDCCH.

16.3.4 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 use beam-forming also 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 16.2.

Table 16.2 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.

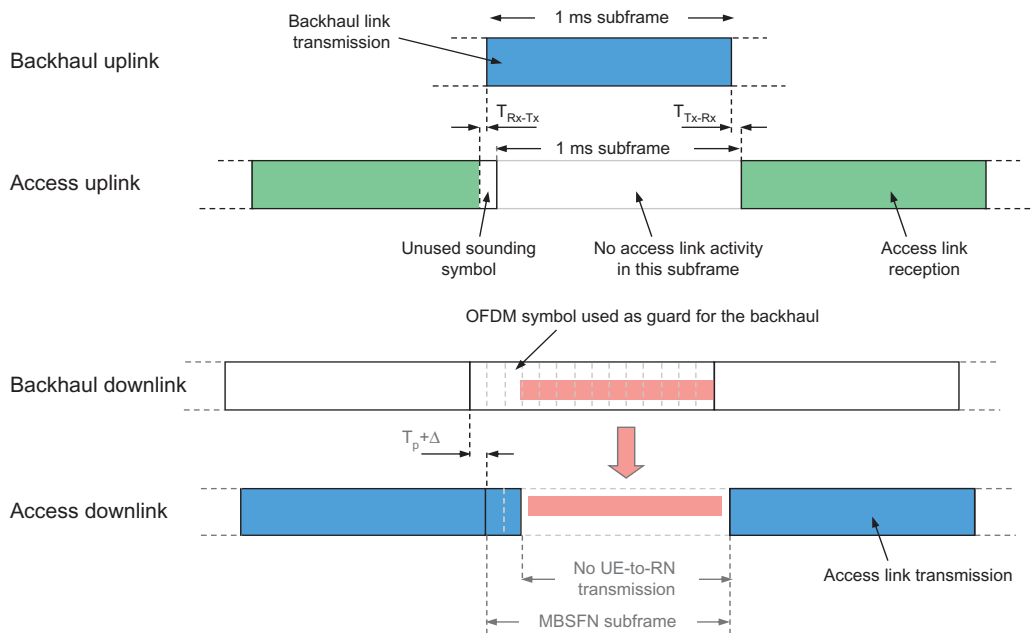
16.3.5 Backhaul–Access Link Timing

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 16.10. 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 16.11. 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.

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

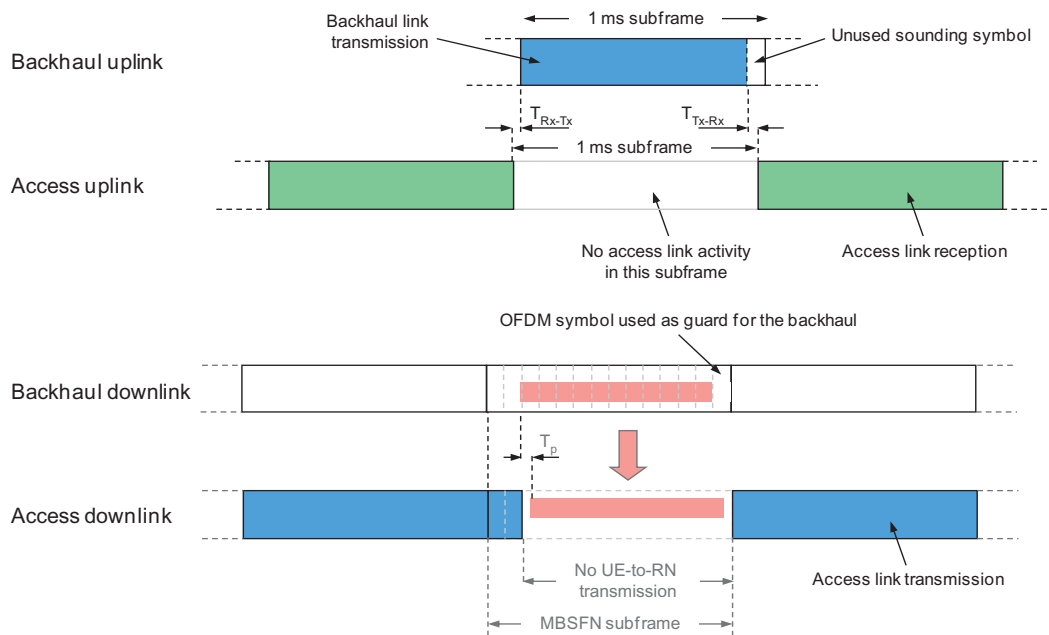
**FIGURE 16.10**

Backhaul timing relations when the relay cell timing is derived from the backhaul timing.

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 16.10. 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 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 obviously 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

**FIGURE 16.11**

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

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 16.11.

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 16.11. 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 16.12 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 obviously 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.

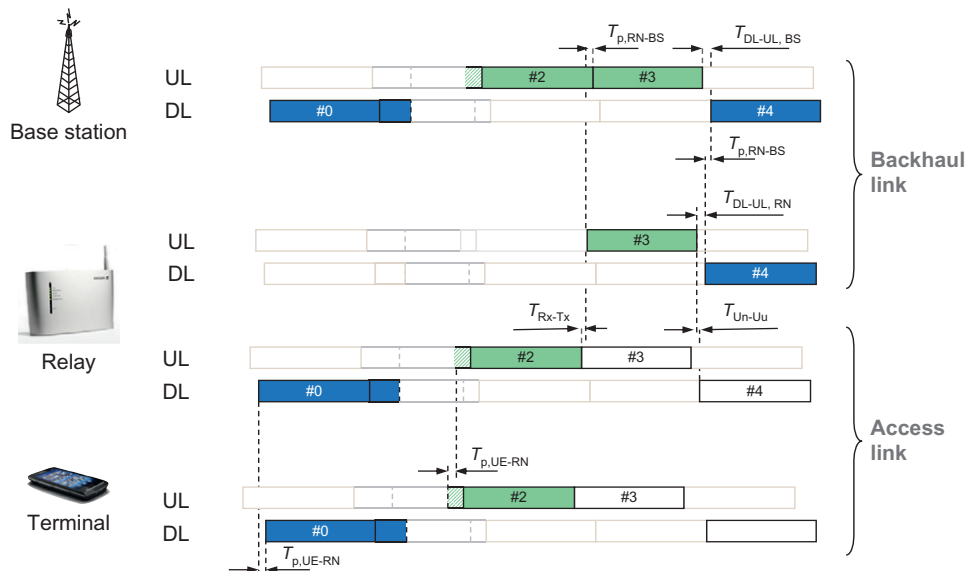


FIGURE 16.12

Example of uplink timing relation for TDD.

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.

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 positions is small and does not justify the additional specification and testing complexity.

Spectrum and RF Characteristics

17

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. Most of these bands were already defined for operation with WCDMA/HSPA, and LTE is the next technology to be deployed in those bands. 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.

17.1 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, quad-band GSM terminals are common, capable of operating in the 850, 900, 1800, and 1900 MHz bands. 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 are 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.

17.1.1 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 3 G 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 2 G cellular systems, especially in the Americas. The first deployment of 3 G in 2001–2002 by Japan and Europe were 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 2 G mobile systems at 806–960 and 1710–1885 MHz, and “new” 3 G spectrum in the bands at 2500–2690 MHz. The identification of bands assigned for 2 G was also recognition of the evolution of existing 2 G mobile systems into 3 G. 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.

The somewhat diverging arrangement between regions of the frequency bands assigned to 3 G means that there is not one single band that can be used for 3 G 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 3 G.

17.1.2 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, as discussed in the previous chapters.

Release 8 of the 3GPP specifications for LTE includes 19 frequency bands for FDD and nine for TDD. The paired bands for FDD operation are numbered from 1 to 21 [76], as shown in [Table 17.1](#), while the unpaired bands for TDD operation are numbered from 33 to 41, as shown in [Table 17.2](#). 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 17.1 and 17.2](#), 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. At the same time, a high degree of commonality between the bands is desired to enable global roaming. The set of bands have first been 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 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 3 G operation in the GSM1800 band. The unpaired *Bands 35, 36, and 37* are also defined for the PCS1900 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 3 G 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.

Table 17.1 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
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	USA
13	777–787	746–756	USA
14	788–798	758–768	USA
17	704–716	734–746	USA
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

Table 17.2 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	USA

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 1500MHz 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.

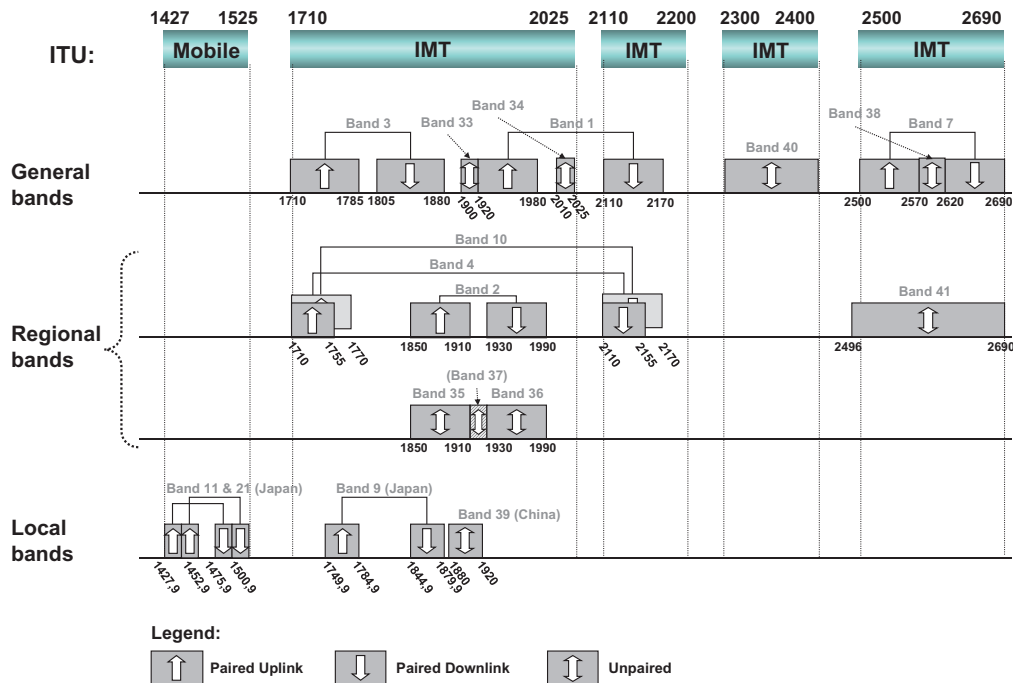


FIGURE 17.1

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

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 17.2, several bands are defined for FDD operation in this range. Band 8 uses the same band plan as GSM900. Bands 5, 18, and 19 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 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.

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. Another regional band for the digital dividend is *Band 20* that is defined in Europe.

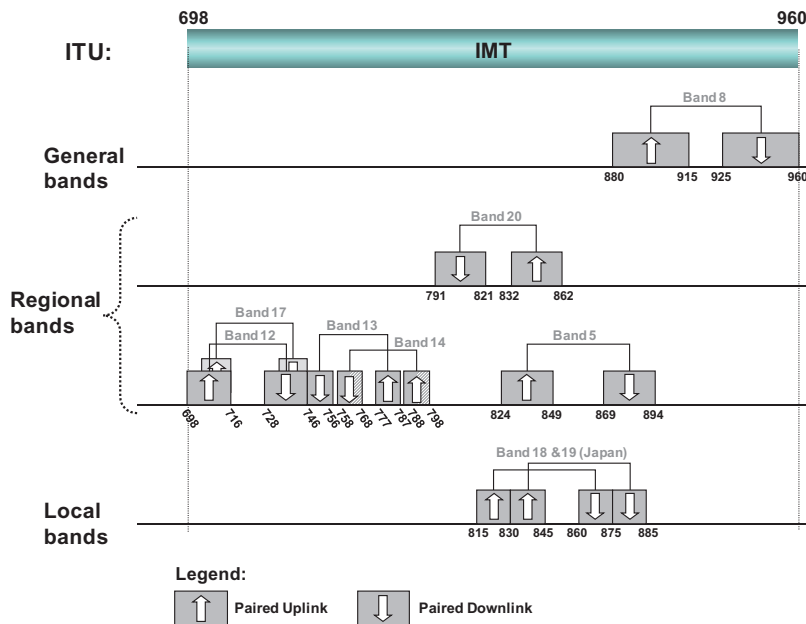


FIGURE 17.2

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

17.1.3 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 bands were defined by WRC'07 that will be 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.
- 698–806 MHz was allocated to mobile service and identified to 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.
- 2300–2400 MHz was identified for IMT on a worldwide basis in all three regions.
- 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.

Additional bands for IMT are not on the agenda for WRC'12, but are likely to be treated at WRC'16. For the frequency ranges below 1 GHz identified at WRC-07, 3GPP has already specified several operating bands, as shown in Figure 17.2. The bands with the widest use are Bands 5 and 8, while most of the other bands have limited regional 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 USA and Band 20 in Europe for the digital dividend. There is a

recent development in the Asia-Pacific area, where a harmonized band plan is agreed [77] for the digital dividend at 698–806 MHz, consisting either of a 2×45 MHz paired band or a ~ 100 MHz unpaired band.

Work in 3GPP is ongoing for the frequency band 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. Licensing of 3.6–3.8 GHz for Wireless Access is more limited. There is a European spectrum decision for 3.4–3.8 GHz with “flexible usage modes” for deployment of fixed, nomadic, and mobile networks. Frequency arrangements considered in the decision include FDD use with 100 MHz block offset between paired blocks and/or TDD use. In Japan, not only 3.4–3.6 GHz but also 3.6–4.2 GHz will be available to terrestrial mobile services such as IMT to use after 2010. The band 3.4–3.6 GHz has also been licensed for wireless access in Latin America. 3GPP is specifying both LTE TDD and LTE FDD modes for the band.

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, one 2×20 MHz band for the S-band MSS operators at 2 GHz and one 2×34 MHz band for the L-band MSS operators at 1.5 GHz.

The US PCS1900 band will be extended with an additional 2×5 MHz block called the *G block*. In order to accommodate the new block, work is ongoing in 3GPP to create an extended PCS1900 band that will be 2×65 MHz, compared to the original Band 2, which is 2×60 MHz.

17.2 FLEXIBLE SPECTRUM USE

Most of the frequency bands identified above 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 very strict.
- *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.

17.3 FLEXIBLE CHANNEL BANDWIDTH OPERATION

The frequency allocations in Figures 17.1 and 17.2 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 17.3. The variation of

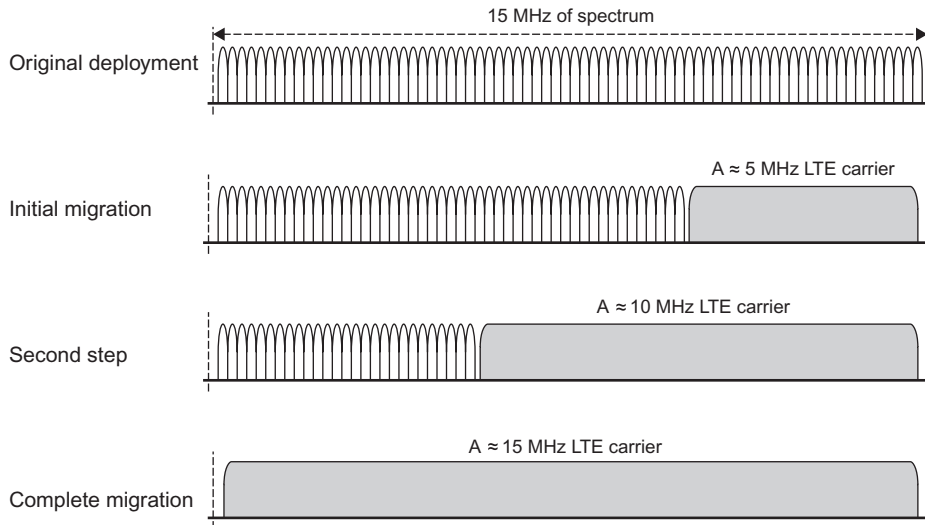


FIGURE 17.3

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

Table 17.3 Channel Bandwidths Specified in LTE	
Channel Bandwidth, BW_{channel} (MHz)	Number of Resource Blocks (N_{RB})
1.4	6
3	15
5	25
10	50
15	75
20	100

possible spectrum scenarios implies a requirement for spectrum flexibility for LTE in terms of the transmission bandwidths supported.

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 UE are defined only for those six channel bandwidths. The channel bandwidths range from 1.4 to 20 MHz, as shown in Table 17.3. 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 in this way is similar to adding new frequency bands.

Figure 17.4 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 17.4 are for a pure OFDM 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.

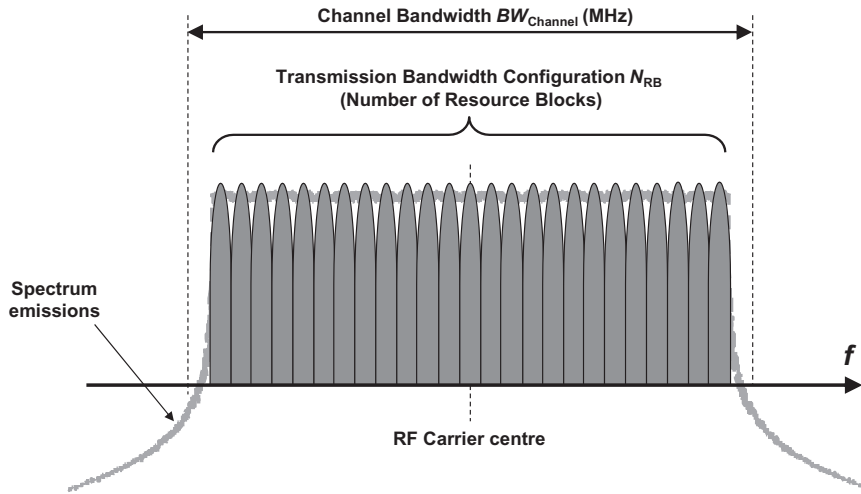


FIGURE 17.4

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

17.4 CARRIER AGGREGATION FOR LTE

The possibility in 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 UE 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.

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. 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 UE. 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 UE.
- *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 UE, due to the possibility for intermodulation and cross-modulation within the UE device when multiple transmitter and 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, which is a configuration not really treated in RF specifications.

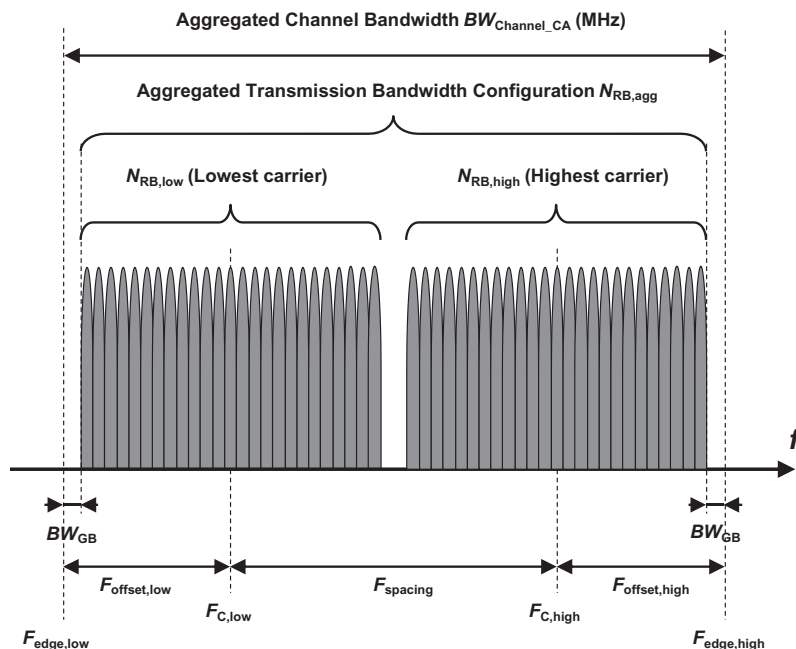


FIGURE 17.5

Definitions for Intra-Band Carrier Aggregation RF parameters, for an example with two aggregated carriers.

Intra-band carrier aggregation is limited to two component carriers and to one paired band (Band 1) and one unpaired (Band 40) band in release 10. Inter-band carrier aggregation is limited to the generic case of aggregating carriers between Bands 1 and 5. The next band pair for which a carrier aggregation capability is specified is a “European” scenario for Bands 3 and 7, which is planned for later inclusion in release 10. The band or set of bands over which carriers are aggregated is defined as a UE capability called *E-UTRA CA Band*. For the base station the band or set of bands defines what is called a *Carrier Aggregation Configuration* for the base station.

For intra-band carrier aggregation, the definitions of $BW_{channel}$ and N_{RB} shown in Figure 17.4 still apply for each component carrier, while new definitions are needed for the Aggregated Channel Bandwidth ($BW_{Channel_CA}$) and the Aggregated Transmission bandwidth Configuration ($N_{RB,agg}$) shown in Figure 17.5. In connection with this, a new capability is defined for the UE called *Carrier Aggregation Bandwidth Class*. There are six classes, where each class corresponds to a range for $N_{RB,agg}$ and a maximum number of component carriers, as shown in Table 17.4. The classes corresponding to aggregation of more than two component carriers or consisting of more than 200 RBs are under study for later releases.

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.

Table 17.4 UE Carrier Aggregation Bandwidth Classes		
Channel Aggregation Bandwidth Classes	Aggregated Transmission BW Configuration	Number of Component Carriers
A	≤ 100	1
B	≤ 100	2
C	101–200	2
D, E, F	Under study (201–500)	Under study

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 9.4, 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

$$F_{\text{Spacing,Nominal}} = \left\lceil \frac{BW_{\text{Channel}(1)} + BW_{\text{Channel}(2)} - 0.1 \lfloor BW_{\text{Channel}(1)} - BW_{\text{Channel}(2)} \rfloor}{2 \cdot 0.3} \right\rceil 0.3. \quad (17.1)$$

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 carriers 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 17.5 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 UE and base station is defined as:

$$BW_{\text{Channel_CA}} = F_{\text{edge,high}} - F_{\text{edge,low}} \quad (17.2)$$

The location of the edges is defined relative to the carriers at the edges through a new parameter F_{offset} (see Figure 17.5) using the following relation to the carrier center positions F_C of the lowest and highest carriers:

$$F_{\text{edge,low}} = F_{C,\text{low}} - F_{\text{offset,low}} \quad (17.3)$$

$$F_{\text{edge,high}} = F_{C,\text{high}} + F_{\text{offset,high}} \quad (17.4)$$

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

² $\lfloor \dots \rfloor$ denotes the “floor” operator, which rounds the number down.

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 UE 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 compatibility reasons, a fundamental parameter such as channel bandwidth and the corresponding reference points (the channel edge) for all RF requirements will have to remain the same. The implication is that the channel edges shown in Figure 17.4 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:

$$F_{\text{offset}} = \frac{BW_{\text{channel}}}{2} \quad (\text{for base station}). \quad (17.5)$$

Unlike the base station, the UE 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 UE, 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} :

$$F_{\text{offset}} = \frac{0.18\text{MHz} \cdot N_{\text{RB}}}{2} + BW_{\text{GB}} \quad (\text{for UE}), \quad (17.6)$$

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 10 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 UE and the base station, but this is in fact not a problem. UE and base-station requirements are defined separately and do not have to cover the same frequency ranges. The aggregated channel BW for both UE 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.

17.5 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 then 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, back-haul 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 17.6 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 17.6, a migration is performed in

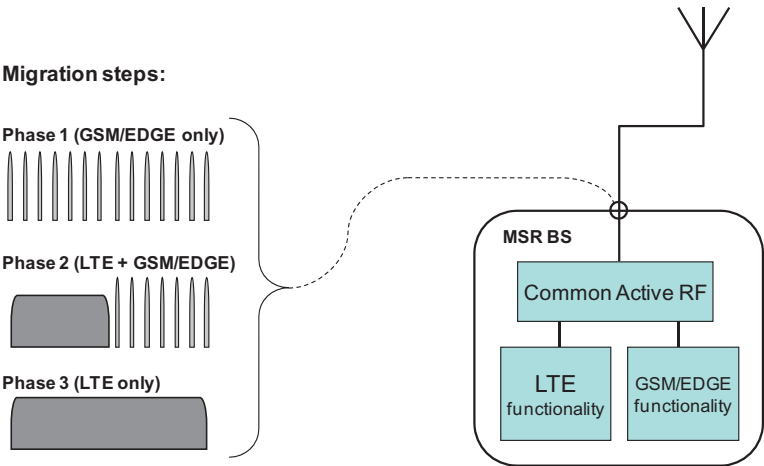


FIGURE 17.6

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

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. 3GPP in release 9 has developed MSR 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 specific 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 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 200kHz, 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.

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

- 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 outside the operating band over the complete frequency range up to 12.75 GHz.

The requirement with the largest impact from MSR is the spectrum mask, or the *operating band unwanted emissions* requirement, as it is called. 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 all supported single RATs and multi-RAT combinations. There are currently six capability sets CS1–CS6 defined in the MSR test specification [81], allowing full flexibility for implementing and deploying base stations compliant to the MSR specification. 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 above in Section 17.4 is also applicable to MSR base stations. Since the MSR specification already has most of the concepts and definitions in place for defining multi-carrier RF requirements, whether aggregated or not, the changes of the MSR requirements compared to non-aggregated carriers are very minor.

17.6 OVERVIEW OF RF REQUIREMENTS FOR LTE

The RF requirements define the receiver and transmitter RF characteristics of a base station or UE. 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 the description below, as it is 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 UE 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 UE 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 UE, while the restrictions on the UE 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, which is more than one *billion* devices per year. In cases where there are differences in how requirements are defined between UE 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 UE in [76]. The RF requirements are divided into transmitter and receiver characteristics. There are also *performance characteristics* for base station and UE 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 UE [74]. These specifications define the test setup, test procedure, test signals, test tolerances, etc. needed to show compliance with the RF and performance requirements.

17.6.1 Transmitter Characteristics

The transmitter characteristics define RF requirements for the wanted signal transmitted from the UE 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 UE and base-station transmitter characteristics arranged according to the three parts defined above is shown in [Table 17.5](#). A more detailed description of the requirements can be found later in this chapter.

Table 17.5 Overview of LTE Transmitter Characteristics

	Base-Station Requirement	UE 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
Unwanted emissions	Operating band unwanted emissions Adjacent Channel Leakage Ratio (ACLR) Spurious emissions Occupied bandwidth Transmitter intermodulation	Spectrum emission mask Adjacent Channel Leakage Ratio (ACLR) Spurious emissions Occupied bandwidth Transmit intermodulation

17.6.2 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 UE and base-station receiver characteristics arranged according to the three parts defined above is shown in [Table 17.6](#). A more detailed description of each requirement can be found later in this chapter.

17.6.3 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 optional requirement and is marked as "regional". For the UE, the same procedure is not possible, since a UE 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*

Table 17.6 Overview of LTE Receiver Characteristics

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

signaling of RF requirements is therefore introduced for LTE, where a UE can be informed at call setup of whether some specific RF requirements apply when the UE is connected to a network.

17.6.4 Band-Specific UE Requirements Through Network Signaling

For the UE, 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 UE when a specific Network Signaling Value (NS_x) is signaled to the UE 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 UE 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 UE RF specification [76], where each value corresponds to a specific condition. The default value for all bands is NS₀₁. 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 UE requirements that have a related Network Signaling Value for some bands:

- NS₀₃, NS₀₄, or NS₀₆ is signaled when specific FCC requirements [88] on UE unwanted emissions apply for operation in a number of US bands.
- NS₀₅ is signaled for protection of the PHS band in Japan when UE 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.

17.6.5 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 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. It is then also clarified from release 9 that the original set of RF parameters is for macro-cell scenarios. The terms macro, 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 BS and UE of 70 dB.
- *Local Area base stations.* This type of base station is intended for pico-cell scenarios, defined with a minimum coupling loss between BS and UE 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 BS and UE of 45 dB is also assumed here. Home BS can be used both for open access and in closed subscriber groups.

The Local Area 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 24 dBm output power for Local Area base stations and to 20 dBm for Home base stations, counting the power over all antennas (up to four). 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 UE 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 UE 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 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 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 Local Area and Home base stations. Receiver dynamic range and in-channel selectivity are also adjusted accordingly.
- All 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-UE) into account.

17.7 OUTPUT POWER LEVEL REQUIREMENTS

17.7.1 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 power limit of 24 dBm output power for Local Area base stations and of 20 dBm for Home base stations, counting the power over all antennas. 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.

17.7.2 UE Output Power and Dynamic Range

The UE 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 UE 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. For each such set of requirement, there is an associated network signaling value NS_x that identifies the allowed A-MPR and the associated conditions, as explained above.

The UE has a definition of the transmitter Off power level, applicable to conditions when the UE 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 UE 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.

17.8 TRANSMITTED SIGNAL QUALITY

The requirements for transmitted signal quality specify how much the transmitted base station or UE 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 measured for base station and UE through *EVM* and *frequency error*. An additional UE requirement is UE in-band emissions.

17.8.1 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 are 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.

17.8.2 UE In-Band Emissions

In-band emissions are emissions within the channel bandwidth. The requirement limits how much a UE 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 UE transmitter affects a real base-station receiver.

17.8.3 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 UE 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.

17.9 UNWANTED EMISSIONS REQUIREMENTS

Unwanted emissions from the transmitter are divided into *out-of-band (OOB) emission* 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 emission 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 UE 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 have a bandwidth of twice the channel bandwidth. For the OOB domain, two overlapping requirements are defined for both base station and UE: *Spectrum Emission Mask* (SEM) and *Adjacent Channel Leakage Ratio* (ACLR). The details of these are further explained below.

17.9.1 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.

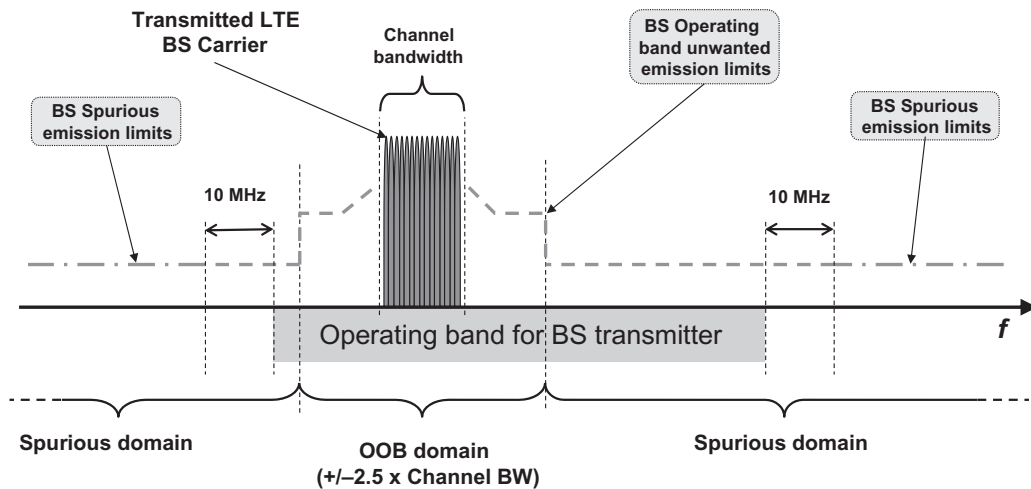
The nonlinear characteristics of the *Power Amplifier* (PA) used to amplify the RF signal must also be taken into account, since it is the source of intermodulation products outside the channel bandwidth. Power back-off to give a more linear operation of the PA can be used, but at the cost of a lower 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.

17.9.2 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 domains is done differently for the LTE base station and UE. Consequently, the spectrum emission masks are also based on different principles.

17.9.2.1 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 bandwidth is handled by not defining an explicit boundary. The solution is a unified concept of *operating band unwanted emissions* 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

**FIGURE 17.7**

Frequency ranges for operating band unwanted emissions and spurious emissions applicable to an LTE base station.

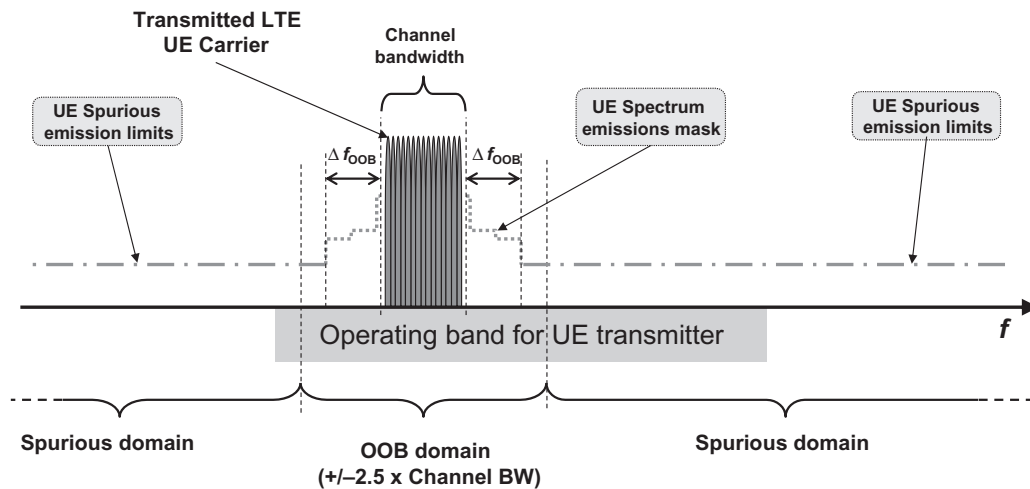
additional 10 MHz on each side, as shown in Figure 17.7. All requirements outside of that range are set by the regulatory spurious emission limits, based on the ITU-R recommendations [90]. As seen in the figure, a large part of the operating band unwanted emissions are defined over a frequency range that for smaller 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 bandwidth 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.

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.

17.9.2.2 UE Spectrum Emission Mask

For implementation reasons, it is not possible to define a generic UE 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 17.8. For 5 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.

**FIGURE 17.8**

Frequency ranges for spectrum emission mask and spurious emissions applicable to an LTE UE.

17.9.3 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 17.9 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 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]:

$$\text{ACIR} = \frac{1}{\frac{1}{\text{ACLR}} + \frac{1}{\text{ACS}}}. \quad (17.7)$$

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

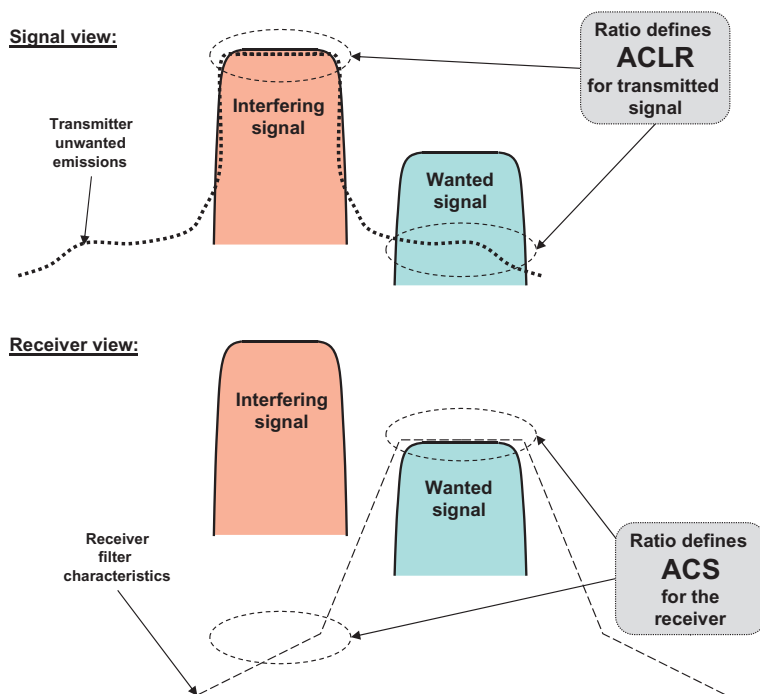


FIGURE 17.9

Illustration of ACLR and ACS, with example characteristics for an “aggressor” interferer and a receiver for a “victim” wanted signal.

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 UE 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 set slightly more relaxed compared to the ACLR, since they are defined in a much narrower measurement bandwidth of 100kHz. 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.

ACLR limits for the UE 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.

17.9.4 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 17.7 – 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.

UE 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 UE 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.

17.9.5 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.

17.9.6 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 UE. 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 UE, there is a similar requirement based on a scenario with another UE transmitted signal appearing at the antenna connector of the UE being specified, but attenuated by 40 dB. The requirement specifies the minimum attenuation of the resulting intermodulation product below the transmitted signal.

17.10 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, which 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 UE, 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 UE is specified as a *maximum signal level* at which the throughput requirement is met.

17.11 RECEIVER SUSCEPTIBILITY TO INTERFERING SIGNALS

There is a set of requirements for base station and UE, 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 possible scenarios with interfering signals of different bandwidths that may be encountered inside and outside the base-station and UE receiver operating band. While the types of requirements are very similar between base station and UE, the signal levels are different, since the interference scenarios for the base station and UE are very different. There is also no UE requirement corresponding to the base-station *In-Channel Selectivity* (ICS) requirement.

The following requirements are defined for LTE base station and UE, starting from interferers with large frequency separation and going close in (see also [Figure 17.10](#)). 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.

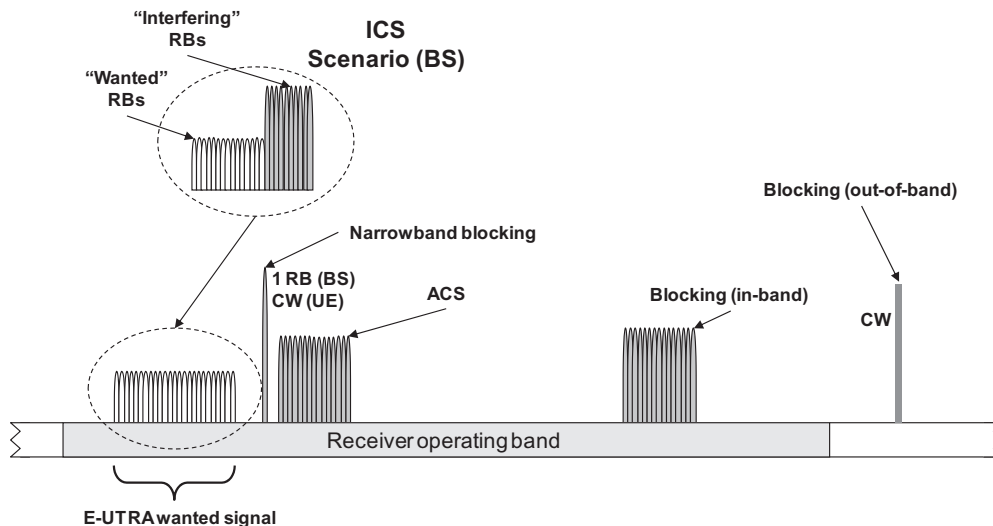


FIGURE 17.10

Base-station and UE requirements for receiver susceptibility to interfering signals in terms of blocking, ACS, narrowband blocking, and in-channel selectivity (BS only).

- **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 band for the base station and the first 15 MHz for the UE. 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-sited with another base station in a different operating band. For the UE, a fixed number of *exceptions* are allowed from the out-of-band blocking requirement, for each assigned frequency channel and at the respective *spurious response frequencies*. At those frequencies, the UE 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 also the discussion in Section 17.9.3). The adjacent interferer is an LTE signal. For the UE, the ACS is specified for two cases with a lower and a higher signal level.
- **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 UE.
- **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 17.10). 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 channel, but at a “desensitized” higher wanted signal level.

18.1 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 obviously 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 yield better data rates and shorter delay. That is, LTE and its evolution can greatly improve both the service experience (end-user viewpoint) and the system capacity (operator viewpoint).

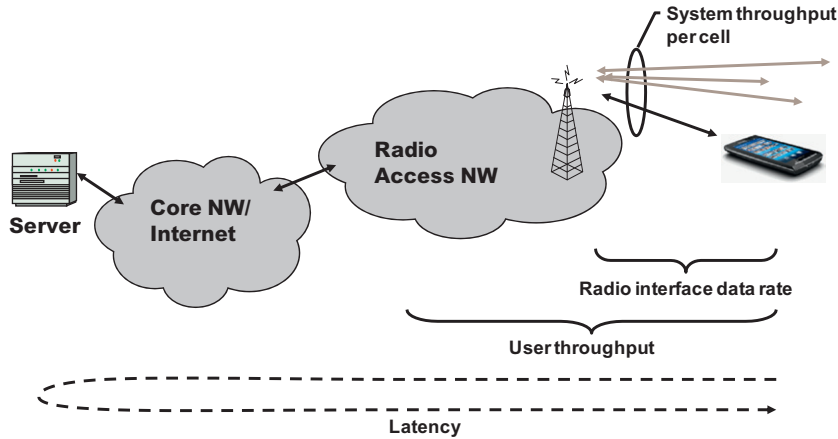
18.1.1 End-User Perspective of Performance

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 18.1](#)). 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 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. Furthermore, 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.

**FIGURE 18.1**

Definitions of data rates for performance.

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.

18.1.2 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.

18.2 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

Table 18.1 LTE Peak Spectral Efficiency [92]					
Peak Spectral Efficiency	ITU 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 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 18.1. The assumptions for deriving the peak spectral efficiency numbers is 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 already 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 18.1. 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.

18.3 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

Table 18.2 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 (GHz)	3.4	2.5	2.0	0.8
Inter-site distance (m)	60	200	500	1732
Terminal speed (km/h)	3	3	30	120
User distribution	100% indoor	50% indoor, 50% outdoor	100% outdoor	100% outdoor
BS antenna height (m)	6	10	25	35
BS antenna gain (dBi)	0	17	17	17
BS output power (dBm/20 MHz)	21	44	49	49
UE output power (dBm)	21	24	24	24

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. Further details are given below. 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].

18.3.1 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 18.2 and the assumptions for the LTE-Advanced system characteristics are given in Table 18.3.

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

Table 18.3 LTE-Advanced System Characteristics for the evaluation in [94]

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	10MHz DL + 10MHz UL for FDD, 10MHz for TDD Double bandwidth for InH case
Antenna configuration at base station (BS)	Vertically co-polarized Antennas with 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 precoded 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; ~1% for 10MHz) has not been deducted

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 18.2.

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 is used 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.

18.3.2 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 18.1 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:

$$\eta = \frac{\sum_{i=1}^N \chi_i}{T \cdot \omega \cdot M}, \quad (18.1)$$

Table 18.4 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

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 18.1), 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.

The normalized user throughput for user i is defined as:

$$\gamma_i = \frac{\chi_i}{T_i \cdot \omega}, \quad (18.2)$$

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 18.4. The requirement levels are shown as dashed lines together with the simulated performance values in Figures 18.2 and 18.5.

18.3.3 Performance Numbers for FDD

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 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 18.2. All results exceed the ITU-R performance requirements in Table 18.4, 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 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].

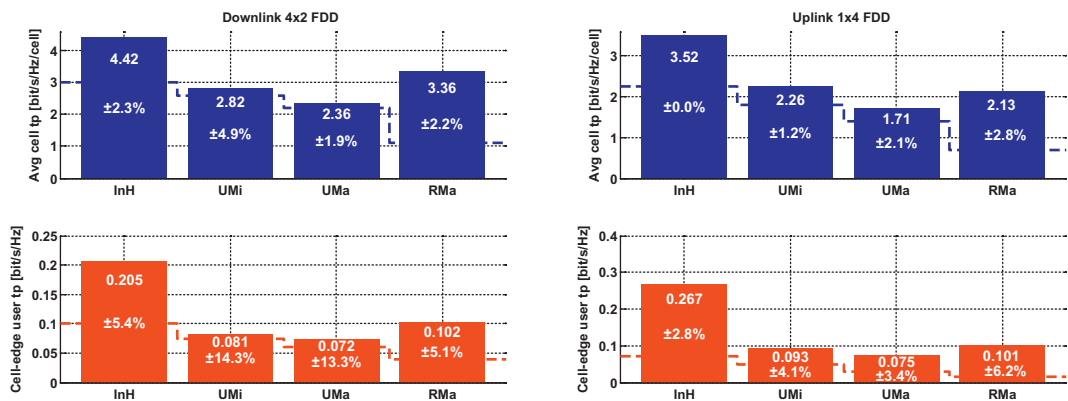


FIGURE 18.2

FDD cell spectral efficiency and cell-edge user spectral efficiency, compared with ITU-R requirements (downlink and uplink).

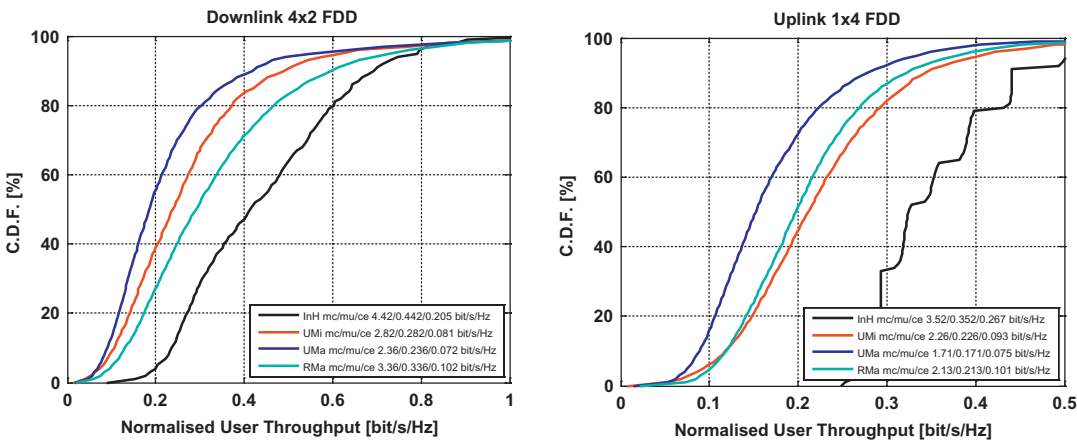


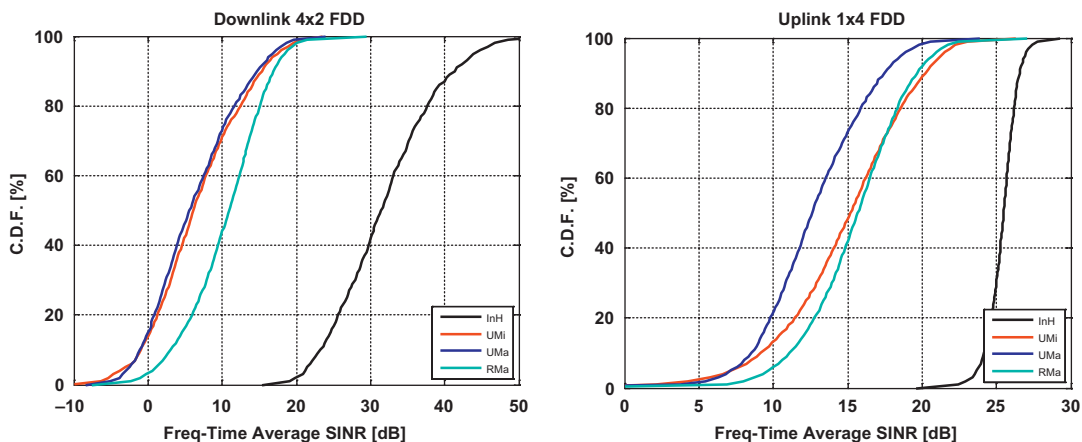
FIGURE 18.3

FDD normalized user throughput distributions (downlink and uplink).

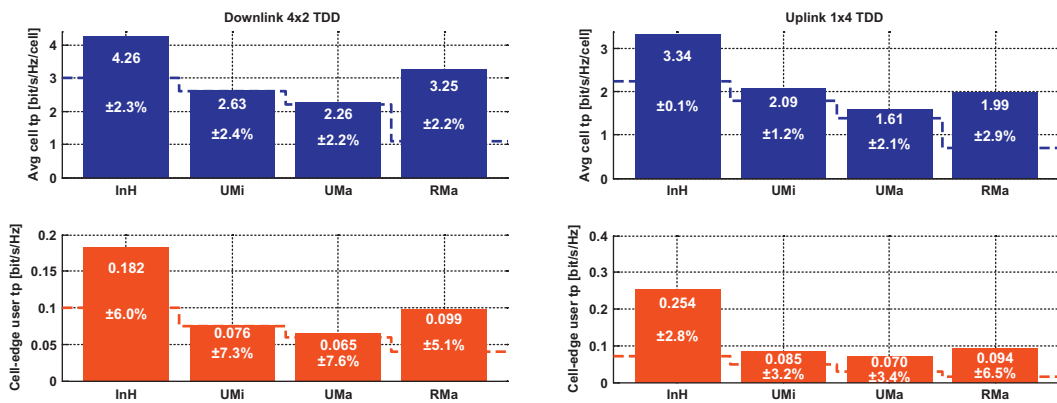
The normalized user throughput distributions are given in Figure 18.3 and Signal-to-Noise-and-Interference Ratio (SINR) distributions are shown in Figure 18.4. The figures demonstrate the impact of the very advantageous SINR distributions achieved in the indoor hotspot scenario and to some extent also for rural macro, resulting in higher spectral efficiency numbers in those environments.

18.3.4 Performance Numbers for TDD

For TDD, the cell spectral efficiency and cell-edge spectral efficiency are shown in Figure 18.5, while normalized user throughput distributions are illustrated in Figure 18.6. There is a small drop in the

**FIGURE 18.4**

FDD SINR distributions (downlink and uplink).

**FIGURE 18.5**

TDD cell spectral efficiency and cell-edge user spectral efficiency, compared with ITU-R requirements (downlink and uplink).

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.

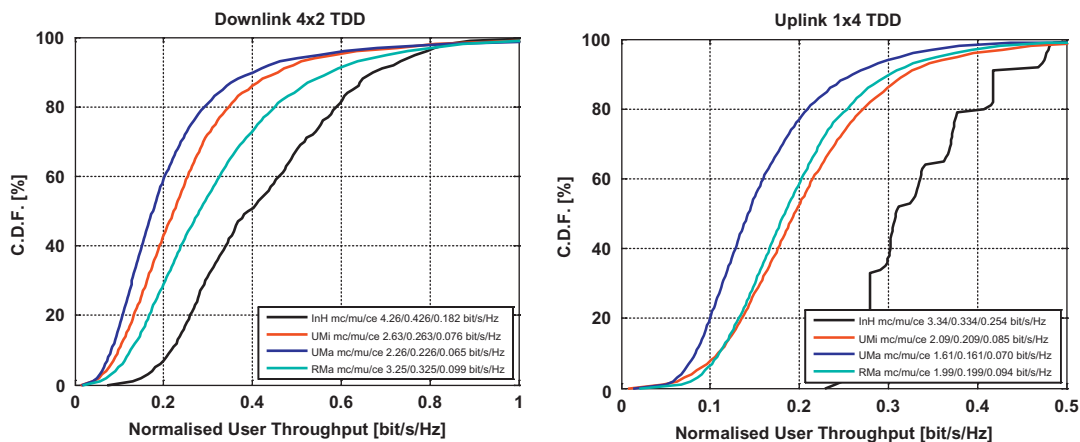


FIGURE 18.6
TDD normalized user throughput distributions (downlink and uplink).

18.4 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.

Other Wireless Communications Systems

19

Most mobile and cellular systems go through a continuous development of the underlying technology and of the features, services, and performance supported. The direction of development is from the baseline “narrowband” systems of the 1990s supporting mainly voice services to today’s “wideband” systems that target a much wider set of services, including broadband wireless data.

The fundamental technologies introduced in the systems to support new and better performing services all fall within the scope of Chapters 2–6 of this book. When the problems to solve are fundamentally the same, the solutions engineered tend to be quite similar. Wideband transmission, higher-order modulation, fast scheduling, advanced receivers, multi-carrier, OFDM, MIMO, etc. have been added as continuous developments and, at other times, as more revolutionary steps for the different technologies.

The application of these different technical solutions to LTE has been discussed in the previous chapters of this book. A similar development is occurring for the other technologies in the IMT-2000 family. This chapter gives a brief overview of HSPA and GSM, developed in 3GPP, 1xEV-DO, developed in 3GPP2, and 806.16m/WiMAX, developed in IEEE in conjunction with WiMAX forum.

19.1 HSPA

Wideband Code-Division Multiple Access (WCDMA), also known as *Universal Terrestrial Radio Access* (UTRA), is the dominant third-generation cellular technology, developed by 3GPP with the first version of the specifications released in 1999. WCDMA is based on code-division multiple access using a chip rate of 3.84Mchip/s, is originally designed for 5MHz of spectrum allocation, and provides wide-area coverage of data rates¹ up to 384kbit/s in both uplink and downlink. Circuit-switched as well as packet-switched services are supported.

High-Speed Packet Access (HSPA) is an evolution of the WCDMA radio interface, adding significantly enhanced support for packet data by applying technologies such as shared-channel transmission, channel-dependent scheduling, link adaptation, and hybrid ARQ with soft combining. The first version of HSPA appeared in release 5 of the WCDMA standard in 2002, but the specifications have

¹The specifications support peak data rates of up to 2Mbit/s in both uplink and downlink, but practical implementations typically support up to 384kbit/s only.

been updated following a similar release scheme as the LTE specifications. Later versions of HSPA include features such as spatial multiplexing, carrier aggregation, and broadcast support. Figure 19.1 outlines the evolutionary steps of HSPA, which are also summarized below:

- Release 5 was the first release of HSPA, introducing the basic set of features (shared-channel transmission, channel-dependent scheduling, link adaptation, and hybrid ARQ with soft combining) in the downlink. Peak data rates up to 14 Mbit/s in the downlink were supported.
- Release 6 provided the basic set of features for uplink transmissions, resulting in an uplink peak data rate of 5.7 Mbit/s.
- Release 7 brought spatial multiplexing of two layers to the downlink, as well as downlink 64QAM and uplink 16QAM, to HSPA. Other enhancements in release 7 included continuous packet connectivity to provide better DRX/DTX capabilities and enhanced CELL_FACH to reduce the latency associated with state switching.
- Release 8 supports simultaneous usage of spatial multiplexing and 64QAM modulation in the downlink.
- Release 9 introduced carrier aggregation in a similar way as later done for LTE, thereby increasing the HSPA bandwidth to 10 MHz.
- Release 10 further increased the bandwidth supported by HSPA to 20 MHz by using four carriers in the downlink. The downlink peak data rate is 168 Mbit/s using roughly 20 MHz of spectrum.

As seen from the list above, the performance of HSPA has increased considerably in less than 10 years since its first incarnation in release 5. The fact that HSPA is an *evolution* and builds upon the basic WCDMA structure sets some constraints on what it is possible to introduce compared to a clean-slate design such as LTE, but also provides the possibility to gradually improve the performance of an already deployed network. This is an important aspect from an operator's perspective. The evolution is still ongoing and HSPA can, in many aspects, provide performance similar to that of LTE.

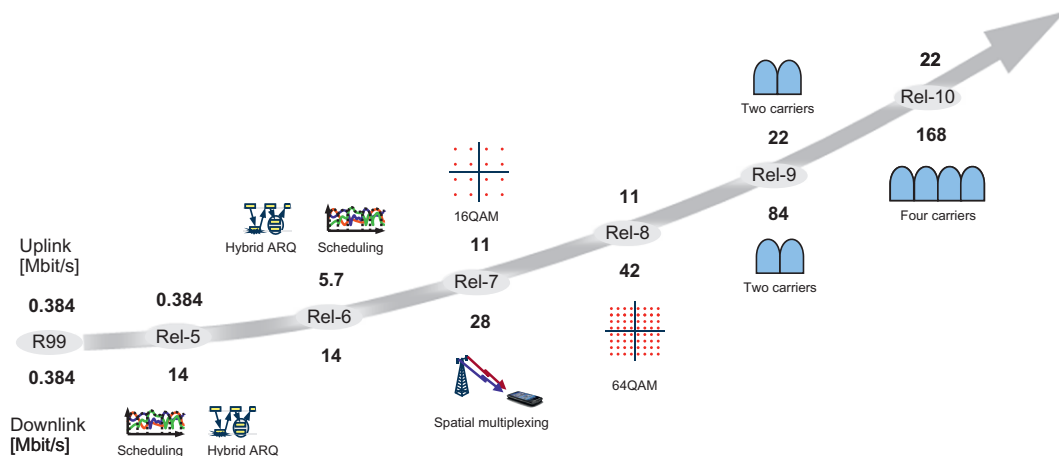


FIGURE 19.1

Evolution of HSPA.

HSPA exists in both FDD and TDD versions but, unlike LTE, the two duplex schemes are designed more or less independently of each other, resulting in significant differences in the physical-layer design. In the following, a brief overview of the most important technology components in the FDD version of HSPA is provided, as well as a short summary of the TDD version. For a more detailed description of HSPA, see [96].

19.1.1 Architecture

The WCDMA radio-access network architecture, upon which HSPA builds, is different from LTE, with the main difference being the existence of a *Radio Network Controller* (RNC), motivated by the support of macro diversity in WCDMA. The functional split between the RNC and the NodeB² is such that the NodeB handles most of the physical-layer processing, while the RNC handles the higher-layer protocols and all the radio-resource management. The WCDMA/HSPA radio-access network architecture is illustrated in Figure 19.2.

HSPA to a large extent relies on the same basic features as LTE, such as channel-dependent scheduling, hybrid ARQ with soft combining, rate adaptation, etc. These features to a large extent rely on rapid adaptation to the instantaneous radio conditions and hence need to be located in the NodeB, while the higher-layer protocols reside in the RNC. At the same time, an important design objective of HSPA was, as far as possible, to retain the original WCDMA functional split between layers and nodes. Minimization of the architectural changes is desirable as it simplifies introduction of HSPA in already deployed networks and also secures operation in environments where not all cells have been upgraded with HSPA functionality. Therefore, HSPA introduced new MAC sublayers in the NodeB, to handle scheduling, rate adaptation, and hybrid-ARQ retransmissions.

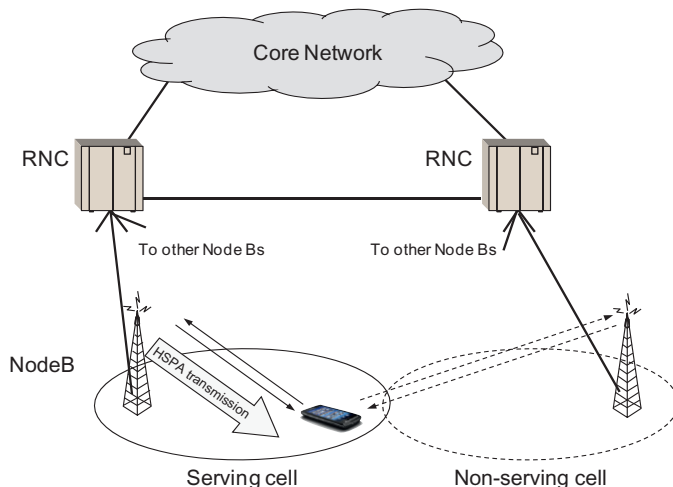


FIGURE 19.2

WCDMA/HSPA architecture.

²In WCDMA/HSPA, the term NodeB is used instead of eNodeB.

Each terminal using HSPA will receive data from one cell, the *serving cell*, which is also the cell responsible for scheduling. Unlike LTE, uplink soft handover is explicitly supported, in which case the uplink data transmission will be received in multiple cells and the terminal will receive power control commands from multiple cells.

19.1.2 Channel-Dependent Scheduling

Channel-dependent scheduling is used in HSPA for similar reasons as in LTE, namely to rapidly adapt to varying traffic and radio-channel conditions. The basic operation is similar to LTE, with scheduling decisions taken by the NodeB once per 2 ms TTI, but the difference in the uplink multiple-access scheme between HSPA and LTE has particularly affected the scheduling design.

The downlink transmission scheme in HSPA is based on reserving a certain fraction of the total downlink radio resources available within a cell, channelization codes and transmission power in the case of WCDMA/HSPA, and dynamically scheduling these resources across the users. The shared resource, illustrated in Figure 19.3, consists of a set of channelization codes of spreading factor 16. The dynamic allocation of the code resource for transmission to a specific user is done on a 2 ms TTI basis (see lower part of Figure 19.3). The reasons for a short TTI are similar as for LTE, namely to follow fast channel variations and to reduce the overall latency. From a scheduling perspective, using a set of orthogonal codes is similar to using a set of orthogonal frequency resources in LTE, and hence the downlink scheduling strategy for HSPA is similar to that of LTE. Channelization codes not used for HSPA data transmission are used for other downlink channels, for example control signaling and non-HSPA services.

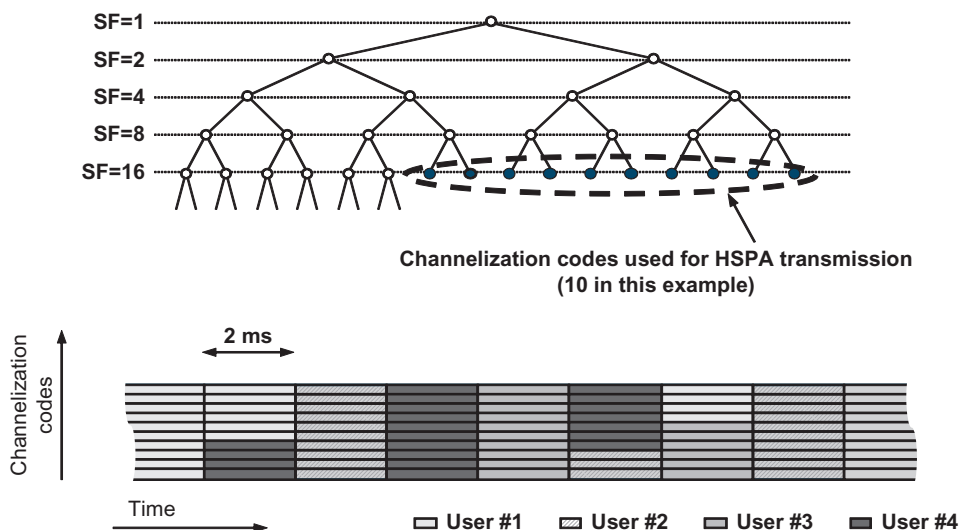


FIGURE 19.3

Shared resource in HSPA downlink – channelization codes.

The uplink transmission scheme, on the other hand, is inherently non-orthogonal and this motivates a different scheduling mechanism than the one used for the orthogonal LTE uplink. The higher the data rate a terminal is using, the higher the received power at the NodeB must be to maintain the E_b/N_0 required for successful demodulation. By increasing the transmission power, the terminal can transmit at a higher data rate. However, due to the non-orthogonal uplink, the received power from one terminal represents interference when receiving from other terminals due to the use of non-orthogonal spreading codes in the uplink. Hence, the shared resource in the uplink is basically the amount of tolerable interference in the cell. If the interference level is too high, some transmissions in the cell, most notably control channels and non-scheduled uplink transmissions, may not be received properly. On the other hand, an interference level that is too low may indicate that terminals are artificially throttled and the full system capacity is not being exploited. Therefore, the task of the uplink scheduler is to give users with data to transmit permission to use as high a data rate as possible without exceeding the maximum tolerable interference level in the cell.

The uplink scheduling framework for HSPA is based on *scheduling grants* sent by the NodeB scheduler to control the terminal transmission activity and *scheduling requests* sent by the terminals to request resources. The scheduling grants essentially control the maximum allowed transmission power the terminal may use; a larger grant implies the terminal may use a higher data rate but also contributes more to the interference level in the cell. Both multi-bit absolute grants, directly setting the grant level in the terminal to the desired level, as well as single-bit relative grants increasing/decreasing the grant level in the terminal are supported. Within the limits set by the scheduling grants, the terminals are allowed to autonomously select the data rate to use. A terminal not utilizing all its granted resources will transmit at a lower power, thereby reducing the intra-cell interference. Hence, shared resources not utilized by one mobile terminal can be exploited by another mobile terminal through statistical multiplexing.

For very high data rates, a non-orthogonal transmission scheme may result in relatively large amounts of interference from high-rate transmissions to low-rate transmissions. HSPA therefore makes it possible to also separate data transmissions in the time domain in the uplink, where terminals are allowed to transmit in certain 2 ms TTIs only. By separating high-rate and low-rate terminals in the time domain, the interference problem can be alleviated.

The fact that HSPA supports uplink macro diversity also has an impact on the scheduling framework. Although the scheduling grants are provided to the terminal by the serving cell only, there is the chance that neighboring cells may affect the grant level in the terminal. In essence, this is an overload indicator, common to multiple terminals, which a cell can set in order to ask terminals in the soft handover region to lower their data rates, thereby reducing the interference level in the non-serving cell.

Different data rates in the uplink as well as in the downlink are supported in a similar way as in LTE by supporting different modulation orders, up to 16QAM in the uplink and up to 64QAM in the downlink, in combination with rate matching to adjust the coding rate. For historical reasons, a 10 ms TTI is supported in the uplink in addition to the 2 ms TTI.

19.1.3 Hybrid ARQ with Soft Combining

Fast hybrid ARQ with soft combining in HSPA uses the same structure with multiple, parallel hybrid-ARQ processes as in LTE. Retransmissions typically take place 12 ms after the initial transmission in the downlink. For the uplink, the corresponding figure is 16 ms for a 2 ms TTI (40 ms for a 10 ms TTI).

Since the RLC is located in the RNC, while the hybrid-ARQ functionality is located in the NodeB, a separate reordering mechanism is located in the NodeB to ensure in-sequence delivery of received transport blocks to the RLC (in LTE, reordering is an integral part of the RLC).

19.1.4 Control-Plane Latency Reductions

Starting in release 7, mechanisms to reduce the overall latency were introduced in HSPA, more specifically *Continuous Packet Connectivity* (CPC) and *Enhanced CELL_FACH*, to match the bursty nature of packet-data traffic.

In WCDMA/HSPA, a terminal can be in one of four different states, URA_PCH, CELL_PCH, CELL_FACH, and CELL_DCH, ranging from low activity, low power consumption, to high activity and higher power consumption. Thus, from a delay perspective it is preferable to keep the terminal in CELL_DCH, while from an interference and power-consumption perspective, one of the paging states is preferred. RRC signaling, originating from the RNC, is used to move the terminal between the different states.

To improve the packet-data support in HSPA, a set of features known as *Continuous Packet Connectivity* (CPC) was introduced in release 7. CPC consists of three building blocks:

1. *Discontinuous transmission* (DTX), to reduce the uplink interference and thereby increase the uplink capacity, as well as to save battery power. Without DTX, the terminal is constantly transmitting on a control channel in the uplink while being in the CELL_DCH state.
2. *Discontinuous reception* (DRX), to allow the terminal to periodically switch off the receiver circuitry and save battery power.
3. *HS-SCCH-less operation*, to reduce the downlink control signaling overhead for small amounts of data, as will be the case for services such as voice-over IP.

The intention with these features is to provide an “always-on” experience for the end-user by keeping the terminal in CELL_DCH for a longer time and avoiding frequent state changes to the low-activity states, as well as improving the capacity for services such as voice-over IP (Figure 19.4).

The purpose of CPC is, as discussed above, to provide an “always-on” user experience by keeping the terminal in the active state CELL_DCH while still providing mechanisms for reduced power consumption. However, eventually the terminal will be switched to CELL_FACH if there has been no transmission activity for a certain period of time. Once the terminal is in CELL_FACH, signaling on the *Forward Access Channel* (FACH), a low-rate common downlink transport channel, is required to

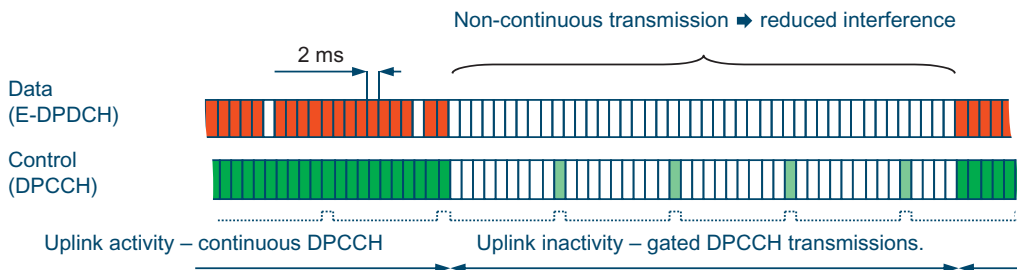


FIGURE 19.4

Simplified illustration of DTX in continuous packet connectivity.

move the terminal to CELL_DCH prior to any data exchange using the HSPA features taking place. The physical resources to which the FACH is mapped is semi-statically configured by the RNC and, to maximize the resources available for HS-DSCH and other downlink channels, the amount of resources (and thus the FACH data rate) is typically kept small, of the order of a few tens of kbit/s.

To reduce the latency associated with state changes, release 7 improved the performance by allowing HSPA to be used also in the CELL_FACH state. This is often referred to as *Enhanced CELL_FACH Operation*. Using the high-rate HSPA transmission scheme also in CELL_FACH allows for a significant reduction in the delays associated with switching to the CELL_DCH state. Instead of using a low-rate semi-statically allocated control channel, the signaling from the network to the terminal can be carried using high-rate HSPA channels. This can result in a significant reduction in call-setup delay and a corresponding improvement in user perception. In release 8, this is further improved by also supporting HSPA functionality in the uplink when being in the CELL_FACH state.

19.1.5 Spatial Multiplexing

Spatial multiplexing in HSPA supports up to two layers in the downlink. Two transport blocks, one per layer, are used in multi-stream transmission and HSPA-MIMO is thus a multi-codeword scheme (see Chapter 5 for a discussion on single- vs. multi-codeword schemes), allowing for a successive-interference-cancellation receiver in the terminal.

Similar to LTE, precoding is used prior to transmitting the two layers. Using precoding is beneficial for several reasons, especially in the case of single-stream transmission. In this case, precoding provides both *diversity gain* and *array gain* as both transmit antennas are used and the weights are selected such that the signals from the two antennas add coherently at the receiver. This results in a higher received carrier-to-interference ratio than in the absence of precoding, thus increasing the coverage for a certain data rate. Furthermore, if a separate power amplifier is used for each physical antenna, precoding ensures that both of them are also used in the case of single-layer transmissions, thereby increasing the total transmit power.

19.1.6 Carrier Aggregation

In later releases of HSPA, carrier aggregation is supported with up to four carriers in the downlink and two carriers in the uplink. A similar structure as for LTE has been adopted, where each carrier in principle is independently processed.

19.1.7 UTRA TDD

In parallel to the development of WCDMA and its evolution to HSPA, 3GPP also works on UTRA TDD. Although the higher layers are virtually identical between FDD and TDD, the physical-layer design is, for historical reasons, quite different. Furthermore, there are three versions of the TDD specifications, with the chip rate being one distinguishing factor.

Initially, TDD supported 3.84Mchip/s only, but later 1.28 and 7.68Mchip/s versions were added. The 3.84 and 7.68 versions share many features, while the 1.28 version, also known as *Time-Division-Synchronous CDMA* (TD-SCDMA), is substantially different from the other two. Of the three versions, TD-SCDMA is the only version deployed on a larger scale, mainly in China, with the other two versions being limited to niche applications.

TD-SCDMA was primarily developed in China as an industry standard within CCSA. In release 4, TD-SCDMA was introduced into 3GPP as an alternative to the 3.84 Mchip/s TDD version. The main differences from WCDMA/HSPA are, apart from the use of TDD instead of FDD, the lower chip rate and resulting (approximate) 1.6 MHz carrier bandwidth, the optional 8PSK higher-order modulation, a different 5 ms time-slot structure, and beam-forming support with eight antennas.

The HSPA enhancements to TD-SCDMA are similar to those applied to FDD, such as the use of 16QAM higher-order modulation and the use of hybrid ARQ. Some main differences are that, due to the TDD frame structure, timing relations are very different and a TTI of 5 ms is used. In general, UTRA TDD development is completed at a later stage in the 3GPP standardization process compared to FDD.

Some features found in TD-SCDMA originate from the work within the CCSA [99], including:

- *Multi-frequency operation.* In this mode, multiple 1.28 Mchip/s carriers are supported in one cell, but BCH is only sent on one carrier called *primary frequency*, in order to decrease inter-cell interference. The carrier on the primary frequency contains all common channels while traffic channels can be on both primary and secondary carriers. Each terminal still operates on a single 1.6 MHz carrier.
- *Beam-forming improvement.* Angle-of-arrival parameters for beam-forming can be signaled in the RACH and FACH data frames.
- *Multi-carrier HSPA.* In a cell using multi-carrier HSPA, data can be transmitted to a terminal on more than one carrier. There is a UE capability defined for receiving up to six carriers.

As discussed in Chapter 9, the LTE TDD mode has been designed to ease coexistence with and migration from TD-SCDMA.

19.2 GSM/EDGE

Worldwide GSM deployment started in 1992 and GSM quickly became the most widely deployed cellular standard in the world, with more than 2.5 billion subscribers today. A major evolutionary step for GSM was initiated in late 1996, at the same time as UTRA development started in ETSI. This step sprung out of the study of higher-order modulation for a 3G TDMA system that was performed within the European FRAMES project (see also Chapter 1).

The enhancement of GSM is called EDGE (*Enhanced Data Rates for GSM Evolution*). The EDGE development was initially focused on higher end-user data rates by introducing higher-order modulation in GSM, both for circuit-switched services and the GPRS packet-switched services. During the continued work, focus moved to the enhancement of GPRS (EGPRS), where other advanced radio-interface components were added, including link adaptation, hybrid ARQ with soft combining, and advanced scheduling techniques, as described in Chapter 6 of this book. In this way GSM became the first cellular standard to add such enhancements,³ later followed by WCDMA/HSPA, CDMA2000, and other technologies. GSM/EDGE is, however, a more narrowband technology than WCDMA/HSPA and CDMA2000, implying that the peak data rates achievable are not as high. HSPA and CDMA2000 have,

³The first cellular technology to deploy higher-order modulation was iDEN, which uses “M-16QAM”. It is, however, a proprietary technology and as such is not standardized.

however, added a time-division structure to make full use of advanced scheduling techniques for high-rate data services. Being a TDMA system, EDGE already has a time-division structure.

19.2.1 Objectives for GSM/EDGE Evolution

An evolved GSM/EDGE standard is developed in 3GPP based on a feasibility study for an “Evolved GSM/EDGE Radio Access Network (GERAN)” [100]. The objective is to improve service performance for GERAN for a range of services, including interactive best-effort services as well as conversational services including *voice-over IP*.

The Evolved GERAN is compatible with the existing GSM/EDGE in terms of frequency planning and coexistence with legacy terminals. It also reuses the existing network architecture (Figure 19.5) and has minimum impact on the BTS, BSC, and core network hardware. For the standardization of Evolved GERAN, a number of performance targets were set [100]:

- *Improved spectrum efficiency.* The target is a 50% increase in an interference-limited scenario (measured in kbit/s/MHz/cell for data or Erlang/MHz/cell for voice).
- *Increased peak data rates.* The target is a 100% increase in both downlink and uplink.
- *Improved coverage for voice and data.* The target is a sensitivity increase of 3 dB in downlink (noise-limited scenario).
- *Improved service availability.* The target is a 50% increase in mean data rate for uplink and downlink at cell edges (when cells are planned for voice).
- *Reduced latency.* The target is a *round-trip time* (RTT) of less than 450 ms for initial access and less than 100 ms after initial access (in non-ideal radio conditions, counting from the terminal to the GGSN and back, as shown in Figure 19.5).

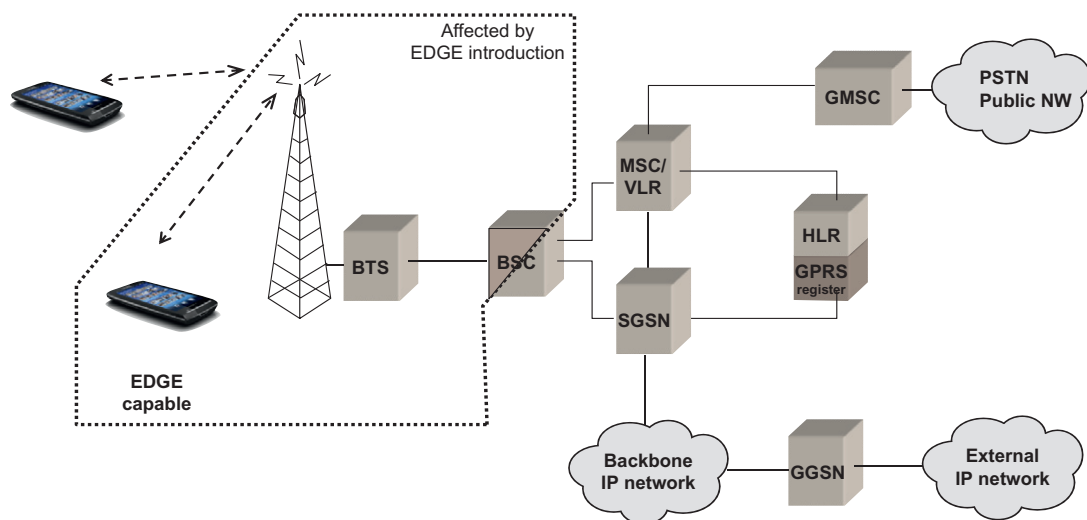


FIGURE 19.5

GSM/EDGE network structure.

A number of alternative solutions were studied in 3GPP to achieve the performance and compatibility targets listed above. These are not fundamentally different from the solutions selected for LTE or other technologies and are all described in a more general context in Chapters 2–6 of this book. The technologies chosen for standardization in 3GPP for GERAN Evolution were:

- Dual-antenna terminals
- Multi-carrier EDGE
- Reduced TTI and fast feedback
- Improved modulation and coding
- Higher symbol rates.

The specific application of these technologies to GERAN is discussed below. The solutions are included in release 7 of the GSM/EDGE specifications.

19.2.2 Dual-Antenna Terminals

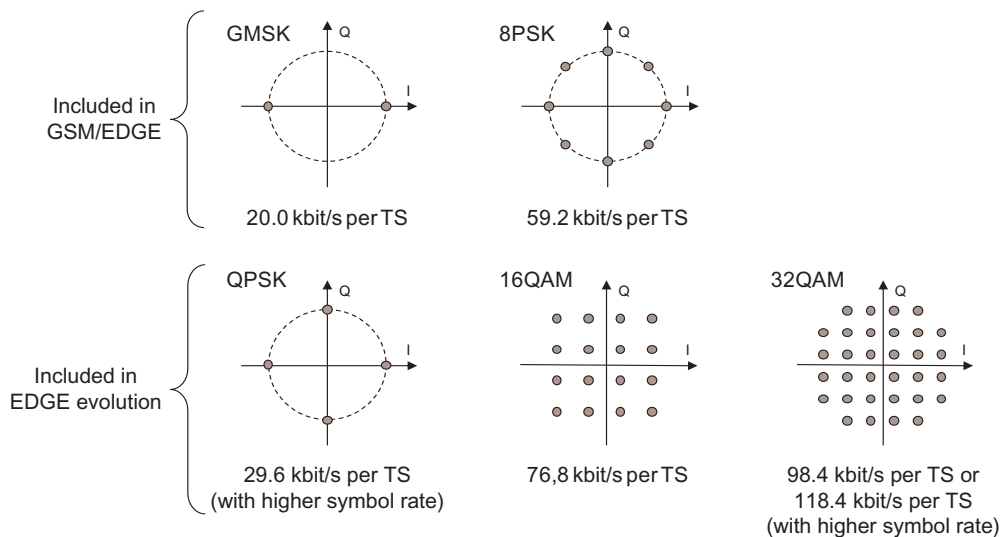
As was shown in Chapter 5, multiple receiver antennas are an effective means against multipath fading and to provide an improved signal-to-noise ratio through “energy gain” when combining the antenna signals. There will thus be improvements for both interference-limited and noise-limited scenarios. There are also possibilities for interference cancellation through multiple antennas. While multiple antennas have an implementation impact for the terminal, there is no impact on the base-station hardware or software.

For GSM/EDGE, a dual-antenna solution called *Mobile Station Receive Diversity* (MSRD) is standardized. Analysis shows that a dual-antenna solution in GSM terminals can give a substantial coverage improvement of up to 6 dB. In addition, the dual-antenna terminals could potentially handle almost 10 dB more interference [79,101].

19.2.3 Multi-Carrier EDGE

The GSM radio interface is based on a TDMA structure with eight time slots and 200 kHz carrier bandwidth. To increase data rates, today’s GPRS and EDGE terminals can use multiple time slots for transmission, reception, or both. Using 8PSK modulation (Figure 19.6) and assigning the maximum of eight time slots, the GSM/EDGE standard gives a theoretical peak data rate of close to 480 kbit/s. However, from a design and complexity point of view, it is best to avoid simultaneous transmission and reception. Today’s terminals typically receive on a maximum of five time slots because they must also transmit (on at least one time slot) as well as measure the signal strength of neighboring cells.

To increase data rates further, multiple carriers for the downlink and the uplink can be introduced, similar to what is discussed in Chapter 2. This straightforward enhancement increases peak and mean data rates in proportion to the number of carriers employed. For example, given two carriers with eight time slots each, the peak data rate will be close to 1 Mbit/s. Using dual carriers can be seen as a straightforward extension of the multi-slot principle, allowing a multi-slot configuration to span more than one carrier. The limiting factor in this case is the complexity and cost of the terminal, which must have either multiple transmitters and receivers or a wideband transmitter and receiver. The use of multiple carriers has only a minor impact on base transceiver stations. A dual-carrier downlink solution is standardized in release 7 of the GSM/EDGE standard.

**FIGURE 19.6**

Existing and new modulation schemes for GSM/EDGE. The highest specified radio-interface data rate using GPRS is shown for each scheme [101]. Note that the view of the nonlinear binary GMSK scheme is simplified in the figure.

19.2.4 Reduced TTI and Fast Feedback

Latency is usually defined as the round-trip time over the radio-access network. It has a major influence on user experience, and conversational services such as voice and video telephony especially require low latency, but services such as web browsing and e-mail download are also improved considerably. A major parameter in the radio interface that has an impact on latency is the *Transmission Time Interval* (TTI), as discussed for LTE in Chapter 8. It is difficult to substantially improve latency without reducing the TTI.

The *round-trip time* (RTT) in GSM/EDGE networks can be 150 ms [101], including network delays but not retransmissions over the radio interface. Radio blocks are transmitted interleaved over four consecutive bursts on one assigned time slot over 20 ms. In GSM/EDGE evolution, a reduced TTI with radio blocks interleaved over four consecutive bursts assigned on *two time slots* in 10 ms is introduced. This reduction of the TTI from 20 to 10 ms can reduce the round-trip time from 150 to less than 100 ms [101]. Multiplexing of users with the two different interleaving schemes is also possible on the same RF carrier.

Faster responses to incorrectly received radio blocks speed up the retransmission of radio blocks and can also help in reducing the latency. In GSM/EDGE evolution, a fast ACK/NACK reporting mechanism is introduced to reduce time for the network to realize that a block is lost. ACK/NACK reports can also “piggy-back” on user data, which reduces the overhead. Combined with the shorter TTI, the total time to retransmit lost radio blocks is thus reduced significantly and the throughput is increased.

19.2.5 Improved Modulation and Coding

The main evolutionary step taken for GSM when EDGE was introduced was the higher-order modulation to enhance the data rates over the radio interface – Enhanced Data Rates for GSM Evolution. 16QAM was considered for EDGE, but 8PSK was finally chosen as a smooth evolutionary step that gives three bits per modulated symbol instead of one. This increases the peak data rate from approximately 20 to 60 kbit/s per time slot.

Figure 19.6 shows the further steps that are standardized for Evolved EDGE, giving 4 bits/symbol for 16QAM and 5 bits/symbol for 32QAM. Since the signal points in the new schemes move more closely together, they are more susceptible to interference. With more bits per symbol, however, the higher data rates allow for more robust channel coding, which can more than compensate for the increased susceptibility to interference. QPSK modulation with 2 bits/symbol is also included in Evolved GSM/EDGE, but used only with the higher symbol rate described below.

Both GSM and EDGE use convolutional codes, while Turbo codes are also standardized for Evolved EDGE. Decoding Turbo codes is more complex than regular convolutional codes. However, since Turbo codes are already used today for WCDMA and HSPA, and many terminals support both GSM/EDGE and WCDMA/HSPA, the decoding circuitry for Turbo codes already exists in the terminals and can be reused for EDGE.

Turbo codes perform well for large code-block sizes, making them better suited for higher data rate EDGE channels using 8PSK or higher-order modulation. It is estimated that, compared to the existing EDGE scheme with 8PSK modulation, the combination of Turbo codes and 16QAM improves the user data rates 30–40% for the median user in a system [100]. Turbo codes are standardized for the downlink in release 7 of GSM/EDGE.

19.2.6 Higher Symbol Rates

The modulation can also be improved by simply increasing the modulation symbol rate. As a part of GERAN Evolution in 3GPP, the combination of higher-order modulation and a 20% increase in symbol rate has been standardized [102]. For both uplink and downlink, two “levels” of terminal capabilities are defined, called Levels A and B. Each level defines a set of modulation schemes, where Level B also applies the higher symbol rate for some modulation schemes. Table 19.1 shows the combinations that are defined.

The higher symbol rate of 325 ksymbol/s will operate with the same nominal carrier bandwidth and carrier raster as legacy GERAN having 271 ksymbol/s, which puts some requirements on the

Table 19.1 Combinations of Modulation Schemes and Symbol Rates in GSM/EDGE Evolution

Terminal Capability	Symbol Rate (ksymbols/s)	Modulation Schemes
Uplink Level A	271	GMSK, 8PSK, 16QAM
Uplink Level B	271	GMSK
	325	QPSK, 16QAM, 32QAM
Downlink Level A	271	GMSK, 8PSK, 16QAM, 32QAM
Downlink Level B	271	GMSK
	325	QPSK, 16QAM, 32QAM

transmitter filter. The amount of transmitter filtering has to be weighed against the receiver complexity and performance, and the amount of interference in adjacent channels. An alternative wider Tx filter has been standardized for the uplink. A wider Tx filter is also studied for the downlink.

19.2.7 Voice Service over Adaptive Multi-User Channels

The focus of GSM/EDGE evolution has been improvements of the data services through higher spectral efficiency and increased peak data rates. At the same time, voice is still a predominant service for many operators and an improvement of voice performance is also desirable. Earlier improvements of voice capacity have mainly been in the voice coding area, where half-rate voice coding was a major step taken in earlier releases. There are eight time slots in the GSM TDMA frame structure, with one full-rate or two half-rate users occupying each slot.

For release 9 of the GERAN specification, multi-user channels on each time slot were investigated and a new mode called *Voice services over Adaptive Multi-user channels* (VAMOS) was introduced [103]. In VAMOS mode, *two* users are multiplexed on the same physical resource – that is, on the same frequency and TDMA time slot. The two users form a *VAMOS pair*, each user being mapped on one *VAMOS subchannel*. In the uplink, each subchannel is modulated with GSMK. In the downlink, a new *Adaptive Quadrature Phase-Shift Keying* (AQPSK) modulation scheme is used jointly for the two subchannels. The two subchannels in a VAMOS pair have different training sequences to facilitate joint demodulation of the two subchannels.

The AQPSK modulation scheme shown in Figure 19.7 is used in the downlink when both subchannels in a VAMOS pair have a burst scheduled for transmission. The bits from one subchannel in the VAMOS pair are mapped on the first bit that forms the QPSK constellation (corresponding to I) and the bits from the second user subchannel are mapped on the second bit (corresponding to Q). Depending on the choice of the parameter α , the power ratio between the two subchannels mapped on I and Q can be controlled, thereby providing partially independent power control of the two subchannels.

Use of the VAMOS mode was studied in detail by GERAN [103], with the conclusion that there is no harmful degradation of voice services for legacy users.

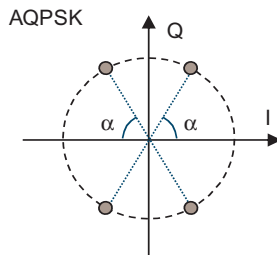


FIGURE 19.7

The AQPSK modulation scheme used in the downlink for two users forming a VAMOS pair.

19.3 CDMA2000 AND HRPD/1x EV-DO

CDMA2000 evolved as a cellular standard under the name IS-95 and later became a part of the IMT-2000 family of technologies described in Chapter 1. When it became a more global IMT-2000 technology, the name was changed to CDMA2000 and the specification work was moved from the *US Telecommunications Industry Association* (TIA) to 3GPP2. Being a sister organization to 3GPP, 3GPP2 is responsible for CDMA2000 specifications.

The CDMA2000 standard is going through an evolution similar to that of WCDMA/HSPA. In the different evolutionary steps, the same gradual shift of focus from voice and circuit-switched data to best-effort data and broadband data can be seen as for WCDMA/HSPA. The basic principles used are also very similar to those found in HSPA.

The evolutionary steps of CDMA2000 are shown in Figure 19.8. After the CDMA2000 1x standard was formed as an input to ITU for IMT-2000, two parallel evolutionary tracks were initiated for better support of data services. The first one was EV-DO (*Evolution-Data Only*),⁴ which has continued to be the main track, as further described below. It is also called HRPD (*High-Rate Packet Data*). A parallel track was EV-DV (*Evolution for integrated Data and Voice*), developed to give parallel support of data and circuit-switched services on the same carrier. It is not used at the moment and has not been developed further within 3GPP2.

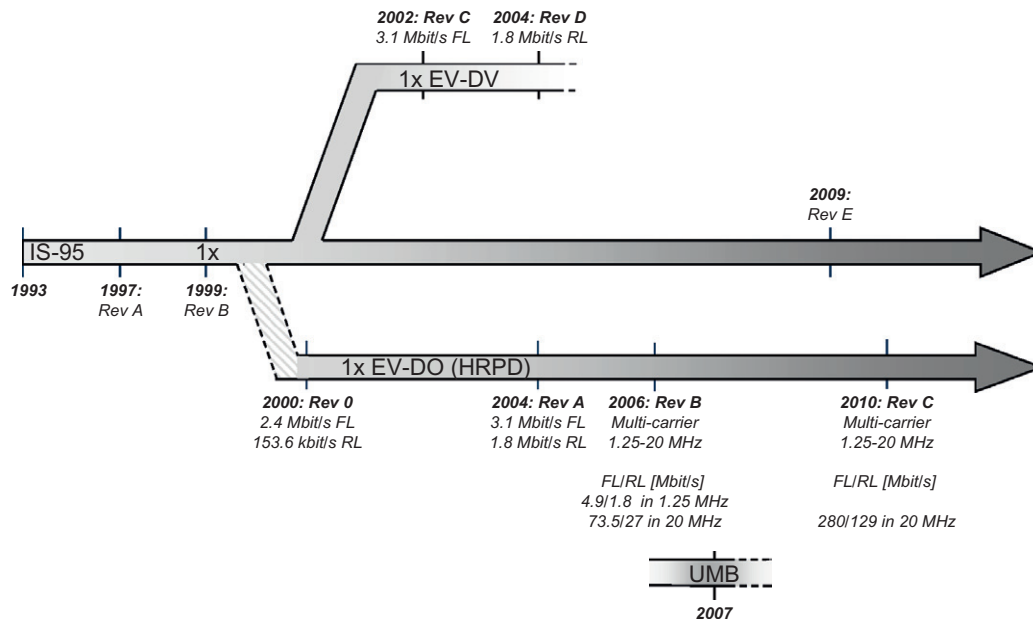


FIGURE 19.8

The evolution from IS-95 to CDMA2000 1x and 1x EV-DO.

⁴The abbreviation “DO” is also interpreted as “Data Optimized”, since the EV-DO carrier targets data services.

The figure also illustrates UMB (*Ultra Mobile Broadband*), an OFDM-based standard including support for multi-antenna transmission and channel bandwidths up to 20MHz. It is similar to LTE in terms of the technologies and features used, with one main difference that UMB uses OFDM in the uplink while LTE uses single-carrier modulation. UMB is not backwards compatible with CDMA2000. At the moment UMB is not used, and has not been developed further within 3GPP2, but some features from UMB, most notably the OFDM-based multi-antenna schemes, have subsequently been the basis for the corresponding features in EV-DO Rev C.

19.3.1 CDMA2000 1x

The CDMA2000 standard, which originally supported both a single-carrier (1x) and a multi-carrier (3x) mode, was adopted by ITU-R under the name IMT-2000 CDMA *Multi-Carrier* (MC) [2]. It offers several improvements over the earlier versions of IS-95 that give better spectral efficiency and higher data rates. The most important aspect from a 3G evolution perspective, however, is that 1x has been a platform for further evolution of packet-data modes, as shown in Figure 19.8. While 1x EV-DV is not undergoing any new development, 1x EV-DO is deployed today and is going through several new evolutionary steps in Rev 0, Rev A, Rev B, and Rev C that are described further below. While all steps up to Rev A are fundamentally based on direct sequence spread spectrum and a 1.25 MHz carrier bandwidth, Rev B and Rev C diverge from this by including wider carrier bandwidths and for Rev C also OFDM operation.

The 3x mode of CDMA2000 was never deployed in its original form, but was still an essential component of the submission of CDMA2000 to the ITU-R. However, multi-carrier CDMA is today once again included in the CDMA2000 evolution through EV-DO Rev B.

The 1x version of CDMA2000 (that is, not the EV-DO/HRPD track in Figure 19.8) has evolved into Rev E (the letters C and D were used for the EV-DV versions). Compared to Rev B, several small enhancements have been introduced in order to increase the voice capacity. Receiver requirements have also been defined to facilitate terminals simultaneously receiving circuit-switched voice call on one CDMA2000 carrier and packet data on EV-DO carriers.

19.3.2 1x EV-DO Rev 0

EV-DO Rev 0 defines a new uplink and downlink structure for CDMA2000 1x, where DO originally implied “*Data Only*”. The reason is that an EV-DO carrier has a structure optimized for data that does not support the voice and circuit-switched data services of a CDMA2000 1x carrier. In this way, the whole carrier works as a shared downlink resource for data transmission. An operator would deploy an additional carrier for EV-DO, thereby separating circuit-switched and packet-switched connections on different carriers. A drawback is reduced flexibility in that there cannot be a simultaneous packet-data and legacy circuit-switched service (such as voice) to the same user on one carrier. 1x EV-DO has later also been named HRPD (*High-Rate Packet Data*).

With EV-DO Rev 0, a peak data rate of 2.4Mbit/s is supported in the downlink on a 1.25 MHz carrier. There are several components in EV-DO Rev 0, several of which have similarities with the HSPA evolution described in Section 19.1:

- *Shared-channel transmission.* EV-DO Rev 0 has a *Time-Division Multiplexing* (TDM) downlink, with transmission to only one user at the time with the full power of the Base Station (BS). This makes the downlink a resource that is shared between the users in the time domain only. This is

similar to the “shared-channel transmission” for HSPA, with the difference that HSPA can also share the downlink in the code domain, primarily with non-HSPA users.

- *Channel-dependent scheduling.* The adaptive data scheduler of EV-DO Rev 0 takes into account fairness, queue sizes, and measures of the channel state. It thereby exploits multi-user diversity in fading channels in a way similar to the channel-dependent scheduling used in HSPA.
- *Short TTI.* The TTI is reduced from 20 ms in CDMA2000 to 1.67 ms in EV-DO Rev 0. This is important to enable fast channel-dependent scheduling and rapid retransmissions, also resulting in lower latency. The TTI for EV-DO Rev 0 is thereby of the same order as the 2 ms TTI used for HSPA.
- *Rate control.* EV-DO Rev 0 employs rate control through adaptive modulation and coding, thereby maximizing the throughput for a given channel condition. This is similar to HSPA, but an EV-DO base station follows the rate request from the mobile, while in HSPA the feedback is a recommendation and the NodeB takes the final decision.
- *Higher-order modulation.* EV-DO Rev 0 supports 16QAM modulation in the downlink, which is similar to the first HSPA release.
- *Hybrid ARQ.* The EV-DO Hybrid-ARQ scheme is similar to the scheme used for HSPA.
- *Virtual SOHO.* EV-DO Rev 0 does not use soft handover in the downlink like CDMA2000. Instead, the terminal supports “virtual soft handover” via adaptive server selection initiated by the terminal, which can be seen as fast cell selection within the “active set” of base stations. These server changes may result in some packet transmission delays [104].
- *Receive diversity in the mobile.* EV-DO Rev 0 has terminal performance numbers specified assuming receive diversity, similar to performance requirements specified for HSPA advanced receivers.

19.3.3 1x EV-DO Rev A

The next step in the evolution of CDMA2000 is 1x EV-DO Rev A. The focus is on an uplink improvement similar to the enhanced uplink of HSPA, but it also includes an updated downlink, a more advanced quality-of-service handling, and an add-on multicast mode [104].

The downlink of EV-DO Rev A is based on the EV-DO Rev 0 downlink, with the following differences:

- *Higher peak rates.* EV-DO Rev A downlink supports 3.1 Mbit/s as compared to the 2.4 Mbit/s of EV-DO Rev 0. Rev A also offers a finer quantization of data rates.
- *Shorter packets.* New transmission formats for EV-DO Rev A enable 128-, 256-, and 512-bit packets. This, together with new multi-user packets, where data to multiple terminals share the same packet in the downlink, improves support for lower-rate, delay-sensitive services.

The major enhancements in EV-DO Rev A compared to Rev 0 are in the uplink. This results in a more packet-oriented uplink with higher capacity and data rates. Peak uplink data rates of up to 1.8 Mbit/s are supported:

- *Higher-order modulation.* In addition to BPSK modulation in EV-DO Rev 0, the uplink physical layer of Rev A supports QPSK and optionally 8PSK modulation.
- *Hybrid ARQ.* Improved performance is achieved through an uplink hybrid ARQ scheme, similar to HSPA.
- *Reduced latency.* The use of smaller packet sizes and a shorter TTI enables a reduced latency of up to 50% compared to EV-DO Rev 0 [105].

- *Capacity/latency trade-off*. Each packet can be transmitted in one of two possible transmission modes: *LoLat* gives low latency through a higher power level, ensuring that the packet is received within the latency target, while *HiCap* gives higher total capacity by allowing for more retransmissions and lower transmit power levels. A similar effect can be obtained in HSPA through different HARQ profiles [96].

19.3.4 1x EV-DO Rev B

The next step for the 1x EV-DO series of standards is Rev B, which enables higher data rates by aggregation of multiple carriers. This fundamental way of increasing the data rate is also discussed in Chapter 2. Rev B permits up to sixteen 1.25 MHz carriers to be aggregated, forming a 20 MHz wide system [106], and giving a theoretical peak data rate of up to 46.5 Mbit/s in the downlink. For reasons of cost, size, and battery life, Rev B devices will most likely support up to three carriers [105], giving a peak downlink data rate of 9.3 Mbit/s.

The lower layers of the radio interface are similar to and compatible with those in Rev A, making it possible for both single-carrier Rev 0 and Rev A terminals to function on a Rev B network that supports multi-carrier operation.

Carriers do not have to be symmetrically allocated in the uplink and the downlink. For asymmetric applications such as file downloads, a larger number of carriers can be set up for downlink than for uplink. This reduces the amount of overhead for the uplink transmission. One uplink channel can carry feedback information for multiple forward link channels operating on multiple downlink carriers to a single terminal in asymmetric mode.

19.3.5 1x EV-DO Rev C

The latest step in the evolution of 1x EV-DO is Rev C, which introduces both smaller software-based enhancements as well as larger enhancements requiring hardware modifications.

The major new features in Rev C are inclusion of OFDM and support for spatial multiplexing. OFDM is supported in the forward link only. A subcarrier spacing of 6.8 kHz is used, resulting in an OFDM symbol duration of 162.76 μ s including a cyclic prefix of 16.28 μ s. OFDM- and CDMA-based transmissions are time multiplexed to support Rev C and legacy terminals on the same carrier.

Spatial multiplexing supports up to four transmission layers in the forward link. Forward-link spatial multiplexing is only supported in conjunction with OFDM. In the reverse link, which only supports CDMA-based transmission, up to two layers are supported together with codebook-based precoding under control of the base station. Higher-order modulation in the reverse link has been increased to 64QAM.

19.4 IEEE 802.16E, MOBILE WiMAX AND 802.16m

The IEEE 802.16 family of radio-access technologies is developed by IEEE, initially targeting fixed wireless access applications but later being extended to also support mobility, thus providing a radio-access technology that can be used to offer mobile-data/broadband services.

The initial 802.16 specification was designed for line-of-sight operation in the 10–60 GHz frequency range. The next revision, 802.16a, introduced support for non-line-of-sight operation with focus on the 2–11 GHz frequency range, but still targeting the fixed-wireless-access application. Support for mobility was introduced as part of IEEE 802.16e, finalized in December 2005. 802.16e

is the basis for all currently commercial mobile-data networks based on the 802.16 technology. Part of 802.16e has also, at a relatively late stage (2007), been approved by the ITU as an additional IMT-2000 technology in parallel with WCDMA/HSPA, TD-SCDMA, and CDMA2000/1xEV-DO. The 802.16e-based IMT-2000 technology is only supporting TDD operation and is, in the ITU, referred to as *IMT-2000 OFDMA TDD WAN*.

The latest step taken within the IEEE 802.16 community is the development of the 16m version of the specification (IEEE 802.16m), a key goal of which was to ensure compliance with the ITU requirements on IMT-Advanced [65]. In October 2010, 802.16m was approved by the ITU as a second IMT-Advanced-compliant technology under the name *WirelessMAN-Advanced*, in parallel with LTE/LTE-Advanced.

The IEEE 802.16 specifications cover only the physical and MAC layers of the overall protocol stack, with no definition of higher layers. Furthermore, the 802.16 specifications contain an extremely large set of different features, often serving essentially the same purpose but doing it in different ways, including even multiple alternatives for the basic physical-layer transmission scheme. Including all these options and alternative features in a mobile terminal is not feasible. Thus, it is generally accepted that the 802.16 specifications in themselves do not provide a complete and implementable specification for a mobile-communication system.

The *WiMAX forum* is an industry-led, non-profit cooperation formed to promote and certify compatibility and interoperability of 802.16-based products. As part of this, the WiMAX forum selects the set of features, from the full set of features specified by the IEEE, that should be included in any equipment that conforms to the different *WiMAX System Profiles*. The first such profile, System Profile Release 1.0, was published in 2007, with the second profile, System Profile Release 1.5, finalized in 2009. System Profile Release 2.0 corresponds to the 802.16m specifications currently under development.

The system profiles for 802.16e developed by the WiMAX forum are also, jointly, often referred to as Mobile WiMAX or just WiMAX.

19.4.1 IEEE 802.16e and Mobile WiMAX

Although IEEE 802.16 includes several alternatives for the basic physical-layer transmission scheme, including both OFDM and single-carrier transmission, Mobile WiMAX is based on OFDM transmission supporting different bandwidths for both uplink and downlink. In contrast to LTE, which is based on a single numerology for all transmission bandwidths, different bandwidths of 802.16e may correspond to partly different numerology, including different subcarrier spacing. However, the main focus of Mobile WiMAX has been on the 5 and 10 MHz bandwidths,⁵ both corresponding to a subcarrier spacing $\Delta f \approx 10.94 \text{ kHz}$ and only differing in the number of subcarriers within the transmission bandwidth. Note that, although wider bandwidths are supported by the 802.16e specifications, 10 MHz is the maximum transmission bandwidth supported by the current Mobile WiMAX profiles.

802.16e includes support for both TDD and FDD operation, including the possibility for half-duplex FDD (see Chapter 9 for a discussion on half-duplex FDD in the context of LTE). However, WiMAX has mainly focused on the TDD mode of operation. In particular, TDD was the only duplex operation supported by the first WiMAX release (1.0). Also, as already mentioned, the WiMAX/802.16e-based

⁵There are additional bandwidths supported by Mobile WiMAX, more specifically 3.5, 7.5, and 8.75 MHz.

IMT-2000 standard only supports TDD operation. This overview will therefore focus on the TDD mode of operation for Mobile WiMAX.

As illustrated in Figure 19.9, in the time domain, a 5 ms frame⁶ is divided into two parts, a downlink part and an uplink part, together consisting of 48 OFDM symbols.⁷ With one symbol used as preamble, there is a maximum of 47 symbols available for downlink and uplink transmission. These symbols can be assigned for downlink and uplink transmission respectively, by means of a number of defined downlink/uplink configurations, ranging from 32:12 (35 downlink symbols and 12 uplink symbols) to 26:21 (26 downlink symbols and 21 uplink symbols). This only leaves a very small time available for downlink/uplink switching. If additional switching time is needed, for example to handle large cells, one can simply discard a number of symbols at the end of the downlink part of the frame.

Downlink control signaling for scheduling and hybrid-ARQ operation is transmitted at the beginning of the downlink part. Hence, unlike LTE, where such control signaling can be provided in each downlink subframe, in 802.16e such control signaling can be provided once per 5 ms frame only. Hybrid ARQ with soft combining is also supported both in 802.16 and by Mobile WiMAX. However, in contrast to HSPA and LTE, the first Mobile WiMAX releases only support Chase combining and not incremental redundancy.

Similar to LTE, 802.16e and Mobile WiMAX supports QPSK, 16QAM, and 64QAM⁸ data modulation, providing spectral efficiency up to 6 bit/s/symbol. Also similar to LTE, adaptive modulation and coding (“link adaptation”) can be used to adjust the modulation scheme and coding rate, and thus the data rate, to match the instantaneous channel conditions. The 802.16e specifications support a range of channel coding schemes, including both Turbo codes, similar to HSPA and LTE, and so-called LDPC codes [53]. However, LDPC codes are not supported in current Mobile WiMAX releases.

802.16e allows for mapping of data to the basic OFDM time–frequency grid in different ways to achieve diversity either by means of resource allocations that spread the coded information in the frequency domain, or optimization for channel-dependent scheduling and link adaptation by concentrating the transmission in the frequency domain. The different resource assignment schemes can be used

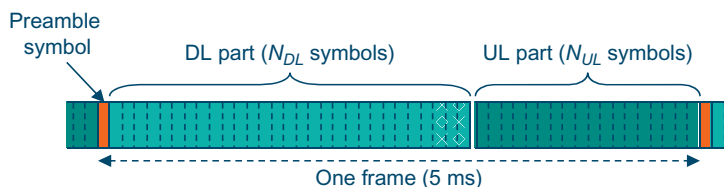


FIGURE 19.9

Basic frame structure for IEEE802.16e/Mobile WiMAX ($N_{DL}:N_{UL} = 29:18$).

⁶Additional frame lengths are specified in 802.16e. However, 5 ms is the only frame length included in the Mobile WiMAX profiles.

⁷This assumes a cyclic-prefix length of approximately 11.4 μ s. The 802.16e specifications include both larger and smaller cyclic-prefix lengths, corresponding to somewhat less/more OFDM symbols within a 5 ms frame. However, these cyclic-prefix lengths are not part of the Mobile WiMAX system profiles.

⁸64QAM for the uplink is not supported/mandatory in the first Mobile WiMAX releases.

within the same subframe by dividing the subframe into different time-domain regions or “zones”. Uplink channel-dependent scheduling was not part of the first release of Mobile WiMAX but was added in later releases.

A wide range of multi-antenna transmission schemes are included in 802.16e. Already since the first release (1.0), Mobile WiMAX has included a subset of these, more specifically open-loop spatial multiplexing. Later releases have introduced additional multi-antenna transmission techniques for both downlink and uplink.

19.4.2 IEEE 802.16m – WiMAX for IMT-Advanced

The work on 802.16m was initiated with the aim to extend the family of 802.16 radio-access technologies and ensure fulfillment of the IMT-Advanced requirements defined by the ITU.

In contrast to LTE-Advanced being an evolution of LTE release 8/9, 802.16m is not a direct evolution of 802.16e, adding new complementary features to extend the performance and capabilities of the radio-access technology. Rather, 802.16m is, in many respects, a new radio-access technology although retaining several of the basic characteristics of 802.16e, including the basic OFDM numerology. This also implies that 802.16e and 802.16m can coexist on the same carrier by means of time multiplexing the two radio-access technologies within the 802.16e 5 ms frame structure. In contrast, earlier releases of LTE are included as integrated parts of LTE release 10, providing a more straightforward backwards compatibility and support of legacy terminals, for example.

Technology wise, 802.16m introduces many features similar to LTE release 10, including the use of multi-carrier transmission (carrier aggregation) for bandwidths beyond 20 MHz and support for relaying functionality.

802.16m also introduces substantially shorter subframes of length roughly 0.6 ms to reduce hybrid-ARQ round-trip time and, in general, allow for reduced latency over the radio interface.

Regarding the resource assignment, 802.16m does not support the resource mapping schemes specified for 802.16e.⁹ Rather, 802.16m introduces *physical resource units* consisting of a number of frequency-contiguous subcarriers during one subframe, very similar to the LTE resource blocks. Actually, the number of subcarriers in a resource unit equals 18 which, together with the subcarrier spacing of 10.94 kHz, results in a resource unit with a bandwidth that is very similar to the LTE resource-block bandwidth of 180 kHz. Above the physical resource units are *logical resource units* that can be distributed or localized, where distributed resource units are used for cases when frequency diversity is used. Once again, this is very similar to LTE localized and distributed resource mapping, as described in Chapter 10.

Taking into account the many similarities between LTE and 802.16m, it is not surprising that performance evaluations indicate similar performance of the two radio-interface technologies. Thus, similar to LTE, 802.16m also fulfils all the requirements for IMT-Advanced as defined by the ITU.

As already mentioned, 802.16m was approved by the ITU in October 2010, as a second IMT-Advanced-compliant technology under the name WirelessMAN-Advanced. Thus, similar to LTE, IEEE 802.16m can also be seen as a fully compliant 4G technology.

⁹They can, of course, be supported on a 802.16m carrier as part of 802.16e/802.16m time multiplexing.

19.5 SUMMARY

The IMT-2000 technologies and the other technologies introduced above are developed in different standardization bodies, but all show many commonalities. The reason is that they target the same type of application and operate under similar conditions and scenarios. The fundamental constraints for achieving high data rates, good quality of service, and system performance will require that a set of efficient tools are applied to reach those targets.

The technologies described in Chapters 2–6 of this book form a set of such tools and it turns out that many of these are applied across most of the technologies discussed here. To cater for high data rates, different ways to transmit over wider bandwidth is employed, such as single- and multi-carrier transmission including OFDM, often with the addition of higher-order modulation. Multi-antenna techniques are employed to exploit the “spatial domain”, using techniques such as receive and transmit diversity, beam-forming, and multi-layer transmission. Most of the schemes also employ dynamic link adaptation and scheduling to exploit variations in channel conditions. In addition, coding schemes such as Turbo codes are combined with advanced retransmission schemes such as hybrid ARQ.

As mentioned above, one reason that solutions become similar between systems is that they target similar problems for the systems. It is also true to some extent that some technologies and their corresponding abbreviations go in and out of “fashion”. Most 2G systems were developed using TDMA, while many 3G systems are based on CDMA and the evolution towards 4G is based on OFDM. Another reason for this stepwise shift of technologies is of course that as technology develops, more complex implementations are made possible. A closer look at many of the evolved wireless communication systems of today also shows that they often combine multiple techniques from previous steps, and are built on a mix of TDMA, OFDM, and spread spectrum components.

Final Thoughts

20

The focus of this book has been on the LTE radio-access technology, from the first version in release 8, via release 9 to release 10. LTE release 10 is the latest version of the LTE radio-access specifications, sometimes also referred to as *LTE-Advanced* and approved by the ITU as the first *IMT-Advanced* technology.

The basic technologies described in Chapters 2–6 served as a foundation for the detailed discussion and explanation of the LTE radio-access technology in the subsequent chapters, providing information about the LTE radio-network architecture, the basic transmission structure, and all the protocols and procedures needed for a complete radio-access network.

The capabilities and system performance of LTE were illustrated in Chapter 18 and it was shown that all the requirements defined by the ITU for IMT-Advanced were fulfilled, in many cases by a significant margin.

At the same time as ITU-R is completing the work on the formal IMT-Advanced recommendations, activities are initiated to look at what will come beyond IMT-Advanced in the 2012–2020 time frame. Several regional workshops on “*IMT for the 2010–2020 Decade*” are planned for 2011 and a new report looking at global broadband wireless services and marketplace for IMT has been initiated within ITU-R WP5D. The basis for the continued work is that future mobile broadband consumers will expect similar user experience and access to similar services and applications in the mobile environment as for fixed broadband, with the added convenience of mobility. Studies within ITU-R will include long-term market forecasts and take into account new user behavior, specifically considering personal computers and smart phones.

20.1 WHERE TO GO IN THE FUTURE?

The evolution of LTE does not end with LTE release 10. Rather LTE will continue to evolve into release 11, release 12, and so on, with each new release bringing additional capabilities and further enhanced system performance into the LTE radio-access technology. Not only will the additional capabilities provide better performance in existing applications, they may also open up for, or even be motivated by, new application areas. Examples hereof are home automation, smart transportation, security, and e-books, but the list is continuously growing as additional applications benefiting from mobile connectivity are emerging. Similarly, new ways of deploying cellular networks, for example more extensive use of massive beam-forming or ubiquitous access to optical fibers in the backhaul, may call for enhancements in the radio interface.

Initial 3GPP discussions on release 11 have already started, with actual work expected to be initiated in early 2011, aiming for completion of this release in mid-2012. In the following, some possible areas for the future evolution of LTE will be briefly discussed. Some of these areas may happen in release 11 or 12, while others may be more futuristic.

20.1.1 Advanced Multi-Cell Coordination

Coordinated Multi-Point transmission and reception (CoMP) refers to a wide range of different techniques with the common denominator being the dynamic coordination of transmission and/or reception at multiple geographically separated sites with the aim to enhance system performance and end-user service quality. Many different coordination schemes of very different characteristics fall under the joint umbrella of CoMP, ranging from dynamic inter-cell scheduling coordination to joint transmission/reception at multiple sites. In the former case, CoMP can, to a large extent, be seen as an extension of the inter-cell interference coordination that is already part of LTE, discussed in Chapter 13.

Joint reception means that the signals received at multiple sites are jointly processed for enhanced reception performance. Maximum-ratio combining and interference-rejection combining are examples of schemes that can be used to combine the uplink transmission received at multiple points. This is, in many respects, similar to softer handover used within a site, for example in WCDMA/HSPA-based systems, but extended to multiple sites.

Joint transmission implies that data is transmitted to a mobile terminal jointly from several sites, thereby not only reducing the interference but also increasing the received power. The transmission from the sites can also take the instantaneous channel conditions at the different terminals into account to enhance the received signal strength, while at the same time reducing the interference between different transmissions.

In general, both joint reception and transmission pose high requirements on low latency in the communication between the network node involved in the joint processing and the different antennas involved in the reception/transmission. Hence, in practice, it can be expected that the different sites may be connected in the form of a *centralized RAN* (C-RAN) deployment for example, as briefly discussed in Chapter 13.

20.1.2 Network Energy Efficiency

Low energy consumption for mobile terminals has been an important requirement ever since the emergence of hand-held terminals roughly 20 years ago. The driving force has been the reduction in battery size and improved battery time. Today, reduced energy consumption *also in the radio-access network* is receiving increased attention for several reasons:

- The cost of energy is a far from negligible part of the overall operational cost for the operator. Thus, reduced energy consumption is one component in the everlasting quest for reduced operating costs for network operators.
- In some rural areas, it may not even be possible to connect the base station to the electrical grid. With sufficiently low energy consumption, reasonably sized solar panels could be used as power source, instead of the diesel generators commonly used today.

- In today's world, where energy consumption and the related climate impact is seen as one of the great challenges for the future, the mobile industry should lead the way for reduced energy consumption.

Regarding the third point, it is important to bear in mind though that the entire ICT (*Information and Communication Technologies*) sector today (2010) contributes roughly 2% of the overall world energy consumption and the contribution of the mobile-communication sector is just a fraction of that. At the same time, the emergence of mobile broadband communication opens up tremendous opportunities for reducing the global energy consumption in general by, for example, virtual meetings replacing traveling to physical meetings.

Still, low energy consumption in mobile networks is important. In particular, taking into account large future traffic increases, where traffic growth of several hundred times up to perhaps 1000 times can be expected in a longer-term perspective, low energy consumption of the mobile networks will most likely be as important a performance metric as capacity, data rates and latency, and must be treated as such.

Reducing the energy consumption of mobile networks is, to a large extent, an implementation issue. However, it is important to ensure that the basic principles of the radio-access technology allows for low energy consumption. One key characteristic is that the energy consumption should scale with the use of the radio resources, not only on average but also on a short-term basis. That is, during points in time where there is no traffic, the energy consumption should be reduced to an absolute minimum, for example by avoiding all but absolutely necessary transmissions. LTE already provides several tools that can be used for this. Nevertheless, the future evolution of LTE should further strive for minimizing transmission of signals strictly not needed.

This will be even more important as the network becomes more and more dense with more and more network nodes. In such network deployments, which may typically be heterogeneous networks as discussed in Chapter 13, pico cells may often be deployed to provide high data rates, rather than being needed for capacity reasons. Thus, the load per cell may be relatively low and each pico cell may often be more or less "empty", further stressing the importance of very low energy consumption when idle.

20.1.3 Machine-Type Communication

Similar to earlier mobile-communication systems, LTE has been designed with data services in mind and much effort has been put into developing techniques for providing high data rates and low latencies for services such as file downloading and web browsing. However, with the increased availability of mobile broadband, connectivity has also become a realistic option for machine-type communication. Machine-type communication spans a wide range of applications, from massive deployment of low-cost battery-powered sensors to remote-controlled utility meters, to surveillance cameras. Many of these applications can be handled by LTE already; communicating with a surveillance camera, for example, is not significantly different from uploading a file, in which case high data rates are paramount. However, other applications may not require transmission of large amounts of data or low latency, but rather pose challenges in terms of a vast amount of devices connecting to the network.

According to some sources [107], 50 billion connections, most of them machine-type communications, can be expected in the year 2020. Handling such a large number of devices is likely to be a challenge mainly for the core network, but improvements in the area of connection setup and power-efficient handling of control signaling in the radio-access network may be of interest.

20.1.4 New ways of Using Spectrum

Spectrum has been, is, and will continue to be a scarce resource for the mobile-communication industry. In particular, in light of the continuous increase in the data rates requested and the corresponding need for wider bandwidths, it is expected that spectrum will remain a scarce resource.

Historically, up until now, the mobile industry has relied on spectrum dedicated for mobile communication and licensed to a certain operator. This will also clearly be the main track for the future. However, in situations where licensed spectrum is not available, other possibilities for increasing the spectrum availability are of interest. This could include the use of unlicensed spectrum, or secondary spectrum primarily used for other communication services, as a complement to operation in the licensed spectrum. Broadcast spectrum not used (in some areas) is often referred to as “white space” [108]. Related to this is the concept of *cognitive radio* [109,110] – that is, radio-access-related functionality related to “smart” selection of spectrum usage. However, the applicability of cognitive radio to cellular communication is a relatively new area and further studies are required to assess the feasibility and impact of such usage.

20.1.5 Direct Device-to-Device Communication

One possible longer-term evolution path would be to extend the LTE radio-access technology with support for direct device-to-device communication – that is, make it possible for two mobile terminals to communicate directly with each other without going via the network. This has been studied in academia for some years, for example as part of the European research project WINNER [111]. Scenarios where the possibility for direct device-to-device communication could be of interest include situations where no network infrastructure is available, for example national security and public safety (NSPS), and situations where network infrastructure is available but when communication directly between the terminals could be more efficient, either in terms of requiring a smaller amount of radio resources for the same communication quality or allowing for improved communication quality. In the latter case, one could envision different degrees of network interaction in the device-to-device communication. However, this area is still in its infancy and scenarios where device-to-device communication offers benefits and how to interact with infrastructure-based communication need to be better understood.

Related to device-to-device communication is the possibility to also use LTE as a radio-access technology in the home, for example a digital camera connecting to the TV. Although such connectivity is possible to some extent today by using (semi-proprietary) technology, basing this type of communication on LTE could be attractive. When outside the home, the devices can communicate with the cellular LTE network, but while at home they can seamlessly connect to the local LTE network covering the home. Furthermore, for devices that are to be used at home only, operation without a SIM card and interaction with an external core network should be possible. Adding this type of functionality to LTE is most likely a relatively simple task.

20.2 CONCLUDING REMARKS

The list of technology areas above should be seen as examples. Some of these technologies will most likely be part of future releases of LTE, while others may not happen at all. There may also be other technologies, not listed above and maybe yet to be discovered, that could be of interest for LTE evolution. What is clear though is that LTE is a very flexible platform that can evolve in different directions to meet the future needs of wireless communication. Given the size of the LTE ecosystem, such evolution is a very attractive path for future wireless communication.

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.1457-9, May 2010.
- [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, Working document towards a preliminary draft new recommendation ITU-R M.[IMT-RSPEC] ITU-R WP5D, Contribution 870, Attachment 5.11.
- [8] ITU-R, Frequency arrangements for implementation of the terrestrial component of international mobile telecommunications-2000 (IMT-2000) in the bands 806–960MHz, 1710–2025MHz, 2110–2200MHz and 2500–2690MHz, Recommendation ITU-R M.1036-3, July 2007.
- [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 multi-user 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 and 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 and 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 2, 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 MobiComp, 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 beamforming 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.
- [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, CRAN – Road 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. Björkén, 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>.

Index

1.28 Mcps TDD, 395
16QAM modulation, 19–21, 399–400
3G, 2–11
3GPP, 9–11
3GPP2, 9, 402–403
32QAM modulation, 399–400
64QAM modulation, 19–21
7.68 Mcps TDD, 395
8PSK modulation, 399–400

A

Absolute grant, 393
Access Stratum (AS), 110
ACIR, 370
Acknowledged mode (for RLC), 261
Acknowledged Resource Indicator (ARI), 184, 238
ACLR, 370–371
ACS, 370–371, 374
Adaptive Modulation and Coding (AMC), 81
Antenna port, 148
Automatic Repeat Request (ARQ), 90–93
 in LTE, 247–259
 for LTE relaying, 335–337

B

Bandwidth utilization, 16
Base-station classes, 365
BCCH *see* Broadcast Control Channel
BCH *see* Broadcast Channel
 transmission of, 305–307
Beam forming:
 classical, 68–70
 in LTE *see* Pre-coding, LTE

Blind decoding of PDCCH, 199–202
Blocking, receiver:
 general, 362, 374
 narrowband, 374
BM-SC *see* Broadcast Multicast Service Center
Broadcast, 43–44
 see also MBMS
Broadcast Control Channel (BCCH), 116
Broadcast Channel (BCH), 116
Broadcast Multicast Service Center (BM-SC), 325
Buffer status report, 277–280

C

Carrier aggregation:
 control signalling for, 193–199
 overview, 104
 physical resources, 132–134
 RF characteristics, 355–358
Carrier Indicator, 183, 190
Carrier raster, 136
CCCH *see* Common Control Channel
CCE *see* Control Channel Element
CDD *see* Cyclic Delay Diversity
CDMA2000, 10, 402–405
Cell identity, 301
Cell-identity group, 302
Cell search, 301–304
Centralized RAN (C-RAN), 292
Channel capacity, 15
Channel coding, LTE:
 downlink, 146–148
 uplink, 204
Channel rank, 166
Channel State Information (CSI), 152

Channel-state report, 282–287
 Channel-Quality Indicator (CQI), 242, 283
 Charging, 109–110
 Chase combining, 90–92
 Closed Subscriber Group (CSG), 294
 CN *see* Core Network
 Code block, 146
 Coexistence
 requirements, 352
 with TD-SCDMA, 140–141
 Cognitive radio, 414
 Co-location requirements, 352
 Common Control Channel (CCCH), 116
 Common Subframe Allocation (CSA), 328
 CoMP *see* Coordinated Multipoint transmission
 Component carrier, 104, 132
 Contention resolution, 312, 319
 Continuous Packet Connectivity (CPC), 390
 Control-channel element, 195
 Controlling RNC *see* RNC
 Control region, 131, 173–174
 Coordinated multipoint transmission, 412
 Core Network (CN), 109–110
 CQI *see* Channel-Quality Indicator
 CRC *see* Cyclic Redundancy Check
 C-RNTI, 125, 319
 Cross-carrier scheduling, 193–194, 254
 CSG *see* Closed Subscriber Group
 CSI *see* Channel-State Information
 Cubic metric, 54
 Cyclic-Delay Diversity (CDD), 66
 Cyclic prefix:
 extended, 127–128
 MBSFN, 131–132
 normal, 127–128
 OFDM, 32–34, 40
 Random-access preamble, 315–317
 single-carrier, 40

Cyclic Redundancy Check (CRC), 90
 for DL-SCH, 144–145
 for PDCCH, 195–196
 for UL-SCH, 203–204

D

Data region, 131, 173
 DC-subcarrier, 130–131
 DCCCH *see* Dedicated Control Channel
 DCI *see* Downlink Control Information
 DCI format, 180–185
 Dedicated Control Channel (DCCCH), 116
 Delay diversity, 65–66
 Device-to-device communication, 414
 DFT-spread OFDM (DFTS-OFDM), 52–57
 localized vs. distributed, 55–57
 receiver, 54–55
 user multiplexing, 55
 in LTE, 96–97, 205
 Discontinuous Reception (DRX), 287–290
 Discontinuous Transmission (DTX), 394
 DL-SCH *see* Downlink Shared Channel
 Donor cell, 332
 Downlink Control Information (DCI), 123, 179
 Downlink Shared Channel (DL-SCH), 117
 processing, 143
 Downlink Traffic Channel (DTCH), 116
 DRX *see* Discontinuous Reception
 DTCH *see* Downlink Traffic Channel
 DTX *see* Discontinuous Transmission
 DwPTS, 138–140

E

EDGE, 396–401
 eNB *see* eNodeB
 eNodeB, 111
 EPC *see* Evolved Packet Core

EPS *see* Evolved Packet System

Equalization, 23, 45–49

Decision-feedback, 49–50

Frequency-domain, 47–49

Time-domain, 45–47

EV-DO, 402–405

EV-DV, 402

EVM requirement, 367

Evolved Packet Core, 109–110

Evolved Packet System, 109

F

FDD *see* Frequency Division Duplex

Femto base station *see* Home-eNodeB

Frame, LTE, 127, 131

Frequency bands, 347–352

numbering of, 349

Frequency Division Duplex, 3, 6, 101, 136–137,
281–282, 384

Frequency hopping, uplink, 207–209

Frequency-Switched Transmit Diversity, 164

FSTD *see* Frequency-Switched Transmit
Diversity

G

GERAN, 10, 397–401

GGSN, 397

GMSK modulation, 399

GSM, 2–3, 7–8, 396–401

migration from, 352–354

multi-standard radio for, 359–361

Guard period, 139

H

Half-duplex FDD, 137–138

in relation to scheduling, 281–282

HARQ *see* Hybrid ARQ

Heterogeneous network deployments,
293–299

Higher-order modulation, 19–21

High Interference Indicator (HII), 292

High Speed Packet Access (HSPA), 389–395
in TDD, 395–396

HII *see* High Interference Indicator

Home-eNodeB, 293, 299

Home Subscriber Service (HSS), 110

HRPD, 402

HSPA *see* High Speed Packet Access

HSS *see* Home Subscriber Service

HS-SCCH-less operation, 394

Hybrid ARQ, 89–93

adaptive, 250–251

control signalling for, 177–179, 229–234

in LTE, 99, 121–123, 247–259

process, 248

relaying, 335–337

synchronous, 250–251

with soft combining, 120–123

I

ICIC *see* Inter-cell interference coordination

ICS, 373–374

IEEE 802.16, 405–408

IMT-2000, 4–6

IMT-Advanced, 4–6, 11–13, 103, 408

Incremental redundancy (IR), 90–93, 249,
253

In-sequence delivery, 262

Inter-cell interference coordination (ICIC), 99,
290–299

in heterogeneous deployments, 294–299

in home-eNodeB, 299

Interference cancellation, 75

Interference coordination *see Inter-cell*

Interference Coordination

Interference Rejection Combining (IRC), 62

Intermodulation:

receiver, 374–375

transmitter, 372

Internet Protocol (IP), 7

IR *see* Incremental Redundancy

ITU, 3–6

see also IMT, IMT-Advanced

L

Latency, 378–379

Layer mapping, LTE, 166, 222

Link adaptation, 79–81, 113, 389

see also Rate adaptation

Logical channels, 113, 116

LTE-Advanced, 103

see also Release 10, IMT-Advanced

M

MAC, 115

control element, 118

Machine-type Communication, 413–414

Master Information Block (MIB), 305–307

MBMS, 323–330

MBMS Control Channel (MCCH), 116,
325–329

MBMS Traffic Channel (MTCH);, 116, 325, 329

MBSFN, 102, 128, 131–132, 324–328

MBSFN area, 324

MBSFN subframe, 131–132, 326–327, 335

MCCH *see* MBMS Control Channel

MCH *see* Multicast Channel

MCH Scheduling Information (MSI), 329

MCH Scheduling Period (MSP), 328

MCH subframe allocation (MSA), 329

MIB *see* Master Information Block

MIMO *see* spatial multiplexing

MME *see* Mobility Management Entity

Mobility Management Entity (MME), 110

MSR *see* Multi-Standard Radio

MTCH *see* MBMS Traffic Channel

Multi-carrier transmission, 21–25

for CDMA2000, 405

for EDGE, 398

for HSPA, 395

for LTE *see carrier aggregation*

for TD-SCDMA, 396

Multicast, 43–44

see also MBMS

Multicast/Broadcast over Single-Frequency

Network *see* MBSFN

Multicast Channel (MCH), 117, 325–328

Multi-cell/multicast Coordination Entity
(MCE), 324

Multi-Standard Radio (MSR) base stations,
359–361

N

NAS *see* Non-Access Stratum

New-data indicator, 184, 191, 249–253

NodeB, 391

Non-Access Stratum (NAS), 110

O

OCC *see* Orthogonal Cover Codes

OFDM *see* Orthogonal Frequency Division
Multiplex

OFDMA, 41

OI *see* Overload Indicator

Operating bands, 348–352

Operating band unwanted emissions, 363,
368–369

Orthogonal Cover Codes (OCC)
 for downlink demodulation reference signals,
 156
 for CSI reference signals, 159
 for uplink reference signals, 216
 Orthogonal Frequency Division Multiplex
 (OFDM), 27–44
 demodulation, 29
 FFT/IFFT implementation, 30–32
 in LTE, 127–131
 Out-of-band
 blocking, 374
 emissions, 367–368
 Overload Indicator (OI), 292

P

Packet Data Network Gateway (P-GW), 110
 Packet Data Convergence Protocol (PDCP),
 111
 Paging, 319–321
 Paging Channel (PCH), 116
 Paging Control Channel (PCCH), 116
 Partial path-loss compensation *see*
 path-loss compensation
 Path-loss compensation, 270
 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), 123
 Physical channel, 123
 Physical Control Format Indicator Channel
 (PCFICH), 123, 174–177
 Physical Downlink Control Channel (PDCCH),
 123, 179–181, 195–202
 Physical Downlink Shared Channel (PDSCH),
 123, 143–152
 Physical Hybrid-ARQ Indicator Channel, 123,
 177–179
 Physical-layer cell identity, 301
 Physical Multicast Channel (PMCH), 123,
 326–327
 Physical Random Access Channel (PRACH),
 124, 312
 Physical Resource Block (PRB):
 downlink, 129–130, 149–152
 uplink, 129–130, 205–207
 Physical Uplink Control Channel (PUCCH),
 123, 227–242
 transmit diversity for
 Physical Uplink Shared Channel (PUSCH),
 123, 203–210
 PMCH *see* Physical Multicast Channel
 PMI *see* Precoding-Matrix Indicator
 Policy and Charging Rules Function (PCRF),
 110
 Positioning, 103
 Power control, 79–81
 for PRACH, 315
 for PUCCH, 267–269
 for PUSCH, 269–271
 for SRS, 271–272
 in LTE uplink, 265–272
 Power Headroom, 271–272, 280
 PRACH *see* Physical Random Access Channel

PRB *see* Physical Resource Block

Pre-coding, LTE, 161–173

codebook-based, 165

closed-loop, 167

non-codebook-based, 169

open-loop, 168

Precoding-Matrix Indicator (PMI), 167, 242, 283

Protocol Data Unit (PDU), 113

PUCCH *see* Physical Uplink Control Channel

PUSCH *see* Physical Uplink Shared Channel

Q

QoS, 110

QPP, 146

QPSK, 19–20

R

RACH *see* Random Access Channel

Radio access network, 95, 111

Radio bearer, 111

Radio frame *see* Frame

Radio Link Control (RLC), 113–115, 247, 257–264

Radio Network Controller (RNC), 391

Radio Resource Control (RRC), 124–125

Radio Resource Management (RRM), 111, 265

RAKE, 45

RAN *see* Radio access network

Random access, 310–319

preamble, 312–317

Random Access Channel (RACH), 117

Rank indicator, 242, 283

Rate adaptation, 97–99, 115, 272

see also link adaptation

Rate control, 106–7

in LTE, 304

Rate matching, 147, 204

Receive diversity, 60, 100

Redundancy version, 92–93, 147, 184, 191, 204, 251, 253–254

Reference sensitivity:

for LTE BS, 373

for LTE terminal, 373

Reference signal:

Cell-specific, 153–156

CRS *see reference signal, cell specific*

CSI *see reference signal, channel-state information*

Demodulation (DM-RS, in downlink)

see reference signal, UE-specific

for channel-state information, 158–161

for MBSFN, 153, 327–328

for positioning, 153

for sounding *see* sounding reference signal

for uplink demodulation, 211–217

power-boosted, 154UE-specific, 153

Reference-signal sequence:

phase-rotated/“cyclically shifted”, 213–214

frequency-domain, 210–211

Reference symbol, 35–36, 153

frequency shifted, 154

Relay, 331–345

access link, 332

amplify-and-forward *see repeater*

architecture, 333–334

backhaul link, 332

control signalling *see* Relay Physical

Downlink Control Channel

decode-and-forward, 331

hybrid-ARQ operation, 336–337

inband, 332

outband, 332

Relay Physical Downlink Control Channel

(R-PDCCH), 337–342

Relative Narrowband Transmit Power (RNTP), 292

Release 8 (LTE), 95–102

Release 9 (LTE), 95, 102–103

Release 10 (LTE), 103–105
see also LTE-Advanced

Reordering, 262

Repeater, 331

Resource-allocation type, 185–186, 192–193

Resource block:
 in downlink, 129–130, 149–153
 in uplink, 129–130, 205–210
 signaling of, 185–187

Resource-block pair, 130

Resource element, 129

Resource element group, 175

RI *see* Rank Indicator

RLC *see* Radio Link Control

Roaming, 2, 7, 11

RNC *see* Radio Network Controller

RNTP *see* Relative Narrowband Transmit Power

R-PDCCH *see* Relay Physical Downlink Control Channel

RRC *see* Radio Resource Control

RRC_IDLE, 125–126

RRC_CONNECTED, 125–126

RRM *see* Radio Resource Management

S

S1 interface, 111

SAE, 8, 95

Scheduling:
 assignment, 179, 181–189
 channel-dependent, 79, 81–89, 97–99
 downlink, 81–85, 273–275

frequency-domain, 87, 97–98

grant, 180, 189–93

greedy filling, 85, 86

half-duplex FDD, 281–82

max-C/I, 82–85

proportional fair, 83–84

request, 229, 234, 238–239, 277–280

round robin, 83–84

semi-persistent, 119, 280–281

uplink, 85–87, 275–280

SDU *see* Service Data Unit

Search space, 199

Selectivity:
 adjacent channel *see* ACS
 in-channel *see* ICS

Self-backhauling *see* Relay

Served traffic, 383

Service Data Unit (SDU), 113

Serving cell, 279

Serving Gateway (S-GW), 110

SFBC *see* Space-Frequency (Block) Coding

SFN *see* Single-frequency network

SFN *see* System Frame Number

S-GW *see* Serving Gateway

Shared-channel transmission, 97, 119

SI *see* System Information message

SIB *see* System Information Block

SIC *see* Successive Interference cancellation

Single-frequency network (SFN), 44
see also MBSFN

SI-RNTI *see* System Information RNTI

Slot, 128–30

Soft handover, 391

Sounding reference signal, 217–221

Space-Frequency Block Coding (SFBC), 68, 163–165, 306

Space Time (Block) Coding (ST(B)C), 66–68

Spatial multiplexing, 71–79, 100, 105
 multi-codeword based, 75–77
 pre-coder based, 74–75
 single-codeword based, 75–77
 in LTE *see* Pre-coding (LTE)

Special subframe, 138–40

Spectral efficiency, 337
 performance, 380–387
 requirement, 380, 384

Spectrum flexibility, 100–102, 354

Spurious emissions, 363, 367–369, 372

Standardization, 8–13

Subcarrier spacing:
 in general, 38
 LTE, 127
 reduced, 128

Subframe, 128

Successive Interference Cancellation (SIC)
 see Interference cancellation

Symbol quadruplet, 175

Synchronization signal, 302–304

System Frame Number (SFN), 128, 305

System Information, 305

System Information Block (SIB), 305, 308–310

System Information message (SI), 308–310

System Information RNTI (SI-RNTI), 308

System performance, 377–379

System throughput, 379, 383

T

TDD *see* Time Division Duplex

TD-SCDMA, 140, 395–396

Time Division Duplex, 3, 6, 101, 138–141, 385

Timing advance, 245–246
 for relaying, 343–344

Transmission mode, 162

Transmission rank, 166

Transmission Time Interval (TTI), 116

Transmit Diversity, 67–69, 163–165
 for PUCCH, 226

Transparent mode (for RLC), 261

Transport block, 116
 signaling of size, 187–189

Transport channels, 116–118

Transport format, 116

Transport format selection, 116, 120, 273

TTI *see* Transmission Time Interval

U

UCI *see* Uplink Control Information

UE categories, 105–106

UL-SCH *see* Uplink Shared Channel

UMB, 403

Unacknowledged mode (for RLC), 261

Unwanted emissions, 362, 367–372

Uplink Control Information (UCI), 123

Uplink Shared Channel (UL-SCH), 117

UpPTS, 138–140

User Equipment (UE), 106
 categories, 106–107

User throughput, 378

UTRA FDD *see* WCDMA

UTRA TDD, 11, 395–396

V

VAMOS, 401

Virtual Resource Block (VRB), 150
 distributed vs. localized, 150–151
 downlink, 150
 uplink, 208

VRB *see* Virtual Resource Block

W

WCDMA, 3, 389–395

White space, 414

WiMAX, 405–408

X

X2 Interface, 111

Z

Zadoff-Chu sequences, 212

APPENDIX 4

Agenda item: 9.6
Source: Nokia, Alcatel-Lucent Shanghai Bell
Title: Beam Forming Impacts
SID: FS_NR_newRAT - Release 14
Document for: Discussion and Decision

1 Introduction

A new study item was agreed at RAN#71 that aims to develop a New Radio Access Technology (NR) [1]. In the study item document, key requirements were listed and it was proposed that NR system should be able to use frequency bands up to 100 GHz. To compensate the increased path loss of higher frequency bands, the need for beam forming was identified. In this contribution we propose to include the study of potential impacts of common channel beam forming on Layer 2 protocols and procedures design.

2 Discussion

The effect of higher frequencies on coverage and the compensation of path loss by using multiple narrow beams for downlink common channels are illustrated on Figure 1. In lower frequency bands (e.g. current LTE bands < 6GHz) the required cell coverage may be provided by forming a wide sector beam for transmitting downlink common channels. However, utilizing wide sector beam on higher frequencies (>> 6GHz) the cell coverage is reduced with same antenna gain. Thus, in order provide required cell coverage on higher frequency bands, higher antenna gain is needed to compensate the increased path loss. To increase the antenna gain over a wide sector beam, larger antenna arrays (number of antenna elements ranging from tens to hundreds) are used to form high gain beams.

As a consequence the high gain beams are narrow compared to a wide sector beam so multiple beams for transmitting downlink common channels are needed to cover the required cell area. The number of concurrent high gain beams that access point is able to form may be limited by the cost and complexity of the utilized transceiver architecture. In practice, on higher frequencies, the number of concurrent high gain beams is much less than the total number of beams required to cover the cell area. In other words, the access point is able to cover only part of the cell area by using a subset of beams at any given time.

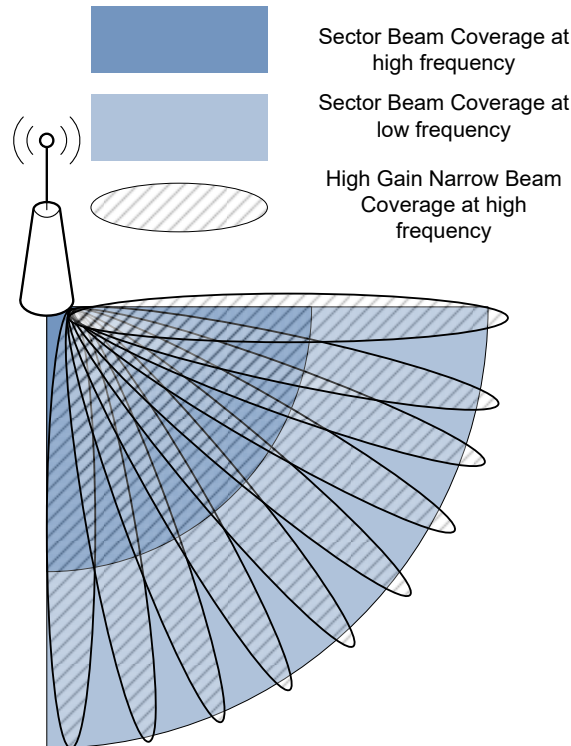


Figure 1. Illustration of providing cell coverage with sector beams and multiple high gain narrow beams

In DL, common channels including synchronization signals, system information and paging will be sweeping across beams to cover the entire cell area. In UL, reception of random access channel (RACH) will also be impacted by the presence of beams. Thus, RAN2 should study the impact of beamforming when designing and defining NR L2 protocols and procedures.

Proposal 1: RAN2 should study the potential impact of beam formed common control plane and beam forming in general on NR Layer 2 procedures and protocol design

Beam formed common control plane may also have an impact to UE mobility. Due to the need for potentially multiple narrow beams to cover a cell area UE may move from coverage area of one beam to another beam inside the cell *the impact of so-called beam level mobility* and its visibility to Layer 2 and Layer 3 should be studied. Furthermore it should be also studied how the beam forming affects inter-cell mobility procedures

Proposal 2: RAN2 should study the potential impact of beam formed common control plane and beam forming in general on the NR mobility and its visibility to Layer 2

3 Conclusion

Proposal 1: RAN2 should study the potential impact of beam formed common control plane and beam forming in general on NR Layer 2 procedures and protocol design

Proposal 2: RAN2 should study the potential impact of beam formed common control plane and beam forming in general on the NR mobility and its visibility to Layer 2

References

- [1] [RP-160671](#), New SID Proposal: Study on New Radio Access Technology, NTT DOCOMO
- [2] [3GPP TR 38.913](#), Study on Scenarios and Requirements for Next Generation Access Technologies