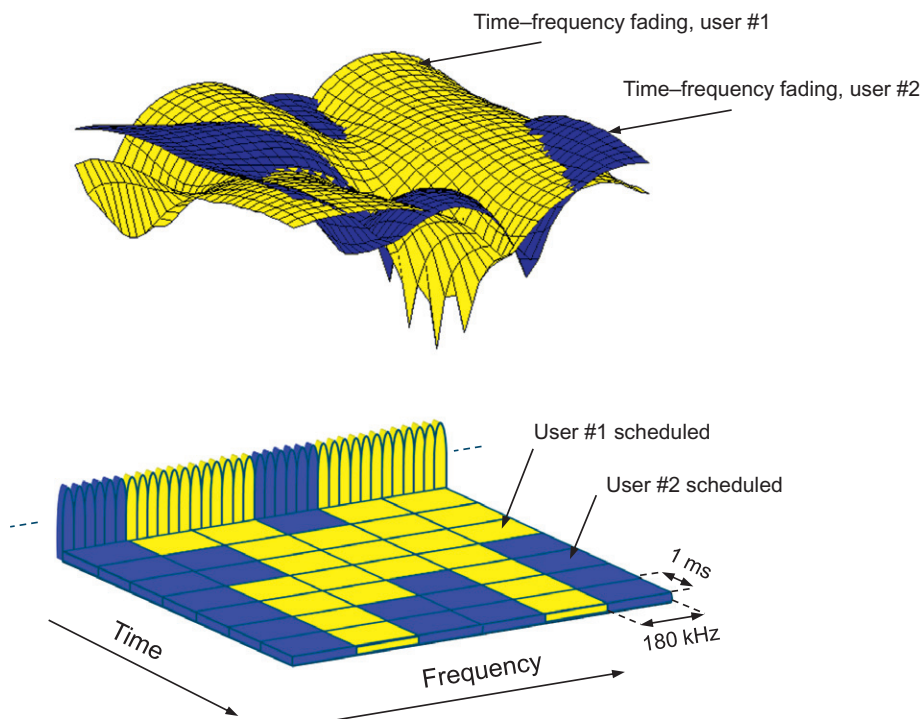


**FIGURE 7.2**

Downlink channel-dependent scheduling in the time and frequency domains.

The possibility of channel-dependent scheduling in the frequency domain is particularly useful at low terminal speeds – in other words, when the channel is varying slowly in time. As discussed in Chapter 6, channel-dependent scheduling relies on channel-quality variations between users to obtain a gain in system capacity. For delay-sensitive services, a time-domain-only scheduler may, due to the delay constraints, be forced to schedule a particular user, despite the channel quality not being at its peak. In such situations, exploiting channel-quality variations also in the frequency domain will help to improve the overall performance of the system. For LTE, scheduling decisions can be taken as often as once every 1 ms and the granularity in the frequency domain is 180 kHz. This allows for relatively rapid channel variations in both the time and frequency domains to be tracked and utilized by the scheduler.

To support downlink scheduling, a terminal may provide the network with *channel-state* reports indicating the instantaneous downlink channel quality in both the time and frequency domains. The channel state is typically obtained by measuring on *reference signals* transmitted in the downlink. Based on the channel-state reports, also referred to as *channel-state information* (CSI), the downlink scheduler can assign resources for downlink transmission to different terminals, taking the channel quality into account in the scheduling decision. In principle, a scheduled terminal can be assigned an arbitrary combination of 180 kHz wide *resource blocks* in each 1 ms scheduling interval.

As already mentioned, the LTE uplink is based on orthogonal separation of different uplink transmissions and it is the task of the uplink scheduler to assign resources in both the time and frequency domains to different terminals. Scheduling decisions, taken once per 1 ms, control what set of terminals are allowed to transmit within a cell during a given time interval and, for each terminal, on what frequency resources the transmission is to take place and what transmission parameters, including the data rate, to use.

Channel conditions can also be taken into account in the uplink scheduling process, similar to the downlink scheduling. However, as will be discussed in more detail in subsequent chapters, obtaining information about the uplink channel conditions may not be feasible or desirable in all situations. Therefore, different means to obtain *uplink diversity* are important as a complement in situations where uplink channel-dependent scheduling is not suitable.

7.1.3 Inter-Cell Interference Coordination

LTE is designed to operate with a one-cell frequency reuse, implying that the same time–frequency resources can be used in neighboring cells. In particular, the basic control channels are designed to operate properly also with the relatively low signal-to-interference ratio that may be experienced in a reuse-one deployment.

From an overall system-efficiency point-of-view, having access to the entire available spectrum in each cell and operating with one-cell reuse is always beneficial. However, it may also lead to relatively large variations in the signal-to-interference ratio, and thus also in the achievable data rates, over the cell area with potentially only relatively low data rates being available at the cell border. Thus, system performance, and especially the cell-edge user quality, can be further enhanced by allowing for some coordination in the scheduling between cells. The basic aim of such *inter-cell interference coordination* (ICIC) is to, if possible, avoid scheduling transmissions to/from terminals at the cell border simultaneously in neighboring cells, thereby avoiding the worst-case interference situations.

To support such interference coordination, the LTE specification includes several messages that can be communicated between eNodeBs using the so-called *X2 interface*, see Chapter 8. These messages provide information about the interference situation and/or scheduling strategies of the eNodeB issuing the message and can be used by an eNodeB receiving the message as input to its scheduling process. LTE interference coordination is discussed in more detail in Chapter 13.

An even more complicated interference situation may occur in so-called heterogeneous network deployments consisting of overlapping cell layers with large differences in the cell output power. This will be briefly discussed in Section 7.3.4 as part of the discussion on LTE release-10 features. The interference handling in heterogeneous network deployments will then be discussed in more detail in Chapter 13.

7.1.4 Hybrid ARQ with Soft Combining

Fast hybrid ARQ with soft combining is used in LTE to allow the terminal to rapidly request retransmissions of erroneously received transport blocks and to provide a tool for implicit rate adaptation. Retransmissions can be rapidly requested after each packet transmission, thereby minimizing the impact on end-user performance from erroneously received packets. Incremental redundancy is used as the soft combining strategy and the receiver buffers the soft bits to be able to perform soft combining between transmission attempts.

7.1.5 Multi-Antenna Transmission

Already from its first release, LTE included support for different multi-antenna transmission techniques as an integral part of the radio-interface specifications. In many respects, the use of multiple antennas is the key technology to reach many of the aggressive LTE performance targets. As discussed in Chapter 5, multiple antennas can be used in different ways for different purposes:

- Multiple receive antennas can be used for receive diversity. For uplink transmissions, this has been used in many cellular systems for several years. However, as dual receive antennas are the baseline for all LTE terminals, the downlink performance is also improved. The simplest way of using multiple receive antennas is classical receive diversity to collect additional energy and suppress fading, but additional gains can be achieved in interference-limited scenarios if the antennas are used not only to provide diversity, but also to suppress interference, as discussed in Chapter 5.
- Multiple transmit antennas at the base station can be used for transmit diversity and different types of beam-forming. The main goal of beam-forming is to improve the received SINR and, eventually, improve system capacity and coverage.
- *Spatial multiplexing*, sometimes referred to as MIMO, using multiple antennas at both the transmitter and receiver is supported by LTE. Spatial multiplexing results in an increased data rate, channel conditions permitting, in bandwidth-limited scenarios by creating several parallel “channels”, as described in Chapter 5. Alternatively, by combining the spatial properties with the appropriate interference-suppressing receiver processing, multiple terminals can transmit on the same time–frequency resource in order to improve the overall cell capacity. This is sometimes referred to as *multi-user MIMO*.

In general, the different multi-antenna techniques are beneficial in different scenarios. As an example, at relatively low SINR, such as at high load or at the cell edge, spatial multiplexing provides relatively limited benefits. Instead, in such scenarios multiple antennas at the transmitter side should be used to raise the SINR by means of beam-forming. On the other hand, in scenarios where there already is a relatively high SINR, for example in small cells, raising the signal quality further provides relatively minor gains as the achievable data rates are then mainly bandwidth limited rather than SINR limited. In such scenarios, spatial multiplexing should instead be used to fully exploit the good channel conditions. The multi-antenna scheme used is under control of the base station, which therefore can select a suitable scheme for each transmission.

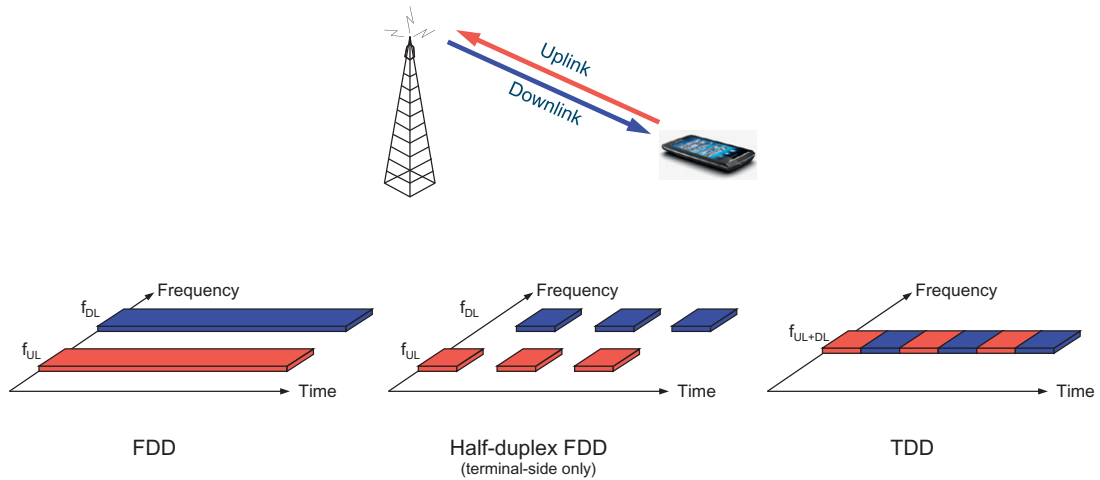
Up to four layers can be spatially multiplexed in release 8. Later releases further enhance the multi-antenna support, as described later.

7.1.6 Spectrum Flexibility

A high degree of spectrum flexibility is one of the main characteristics of the LTE radio-access technology. The aim of this spectrum flexibility is to allow for the deployment of LTE radio access in different frequency bands with different characteristics, including different duplex arrangements and different sizes of the available spectrum. Chapter 17 outlines further details of how spectrum flexibility is achieved in LTE.

7.1.6.1 Flexibility in Duplex Arrangement

One important part of the LTE requirements in terms of spectrum flexibility is the possibility to deploy LTE-based radio access in both paired *and* unpaired spectrum. Therefore, LTE supports both

**FIGURE 7.3**

Frequency- and time-division duplex.

frequency- and time-division-based duplex arrangements. *Frequency-Division Duplex* (FDD), as illustrated on the left in Figure 7.3, implies that downlink and uplink transmission take place in different, sufficiently separated, frequency bands. *Time-Division Duplex* (TDD), as illustrated on the right in Figure 7.3, implies that downlink and uplink transmission take place in different, non-overlapping time slots. Thus, TDD can operate in unpaired spectrum, whereas FDD requires paired spectrum. The required flexibility and resulting requirements to support LTE operation in different paired and unpaired frequency arrangements are further discussed in Chapter 17.

Operation in both paired and unpaired spectrum has been supported by 3GPP radio-access technologies even before the introduction of LTE by means of FDD-based WCDMA/HSPA in combination with TDD-based TD-SCDMA radio. However, this was then achieved by means of, at least in the details, relatively different radio-access technologies leading to additional effort and complexity when developing and implementing dual-mode terminals capable of both FDD and TDD operation. LTE, on the other hand, supports both FDD and TDD *within a single radio-access technology*, leading to a minimum of deviation between FDD and TDD for LTE-based radio access. As a consequence of this, the overview of the LTE radio access provided in the following chapters is, to a large extent, valid for both FDD and TDD. In the case of differences between FDD and TDD, these differences will be explicitly indicated. Furthermore, the TDD mode, also known as TD-LTE, is designed with coexistence between TD-LTE and TD-SCDMA in mind to simplify a gradual migration from TD-SCDMA to TD-LTE.

LTE also supports *half-duplex* FDD at the terminal (illustrated in the middle of Figure 7.3). In half-duplex FDD, transmission and reception *at a specific terminal* are separated in both frequency and time. The base station still uses full-duplex FDD as it simultaneously may schedule *different* terminals in uplink and downlink; this is similar to, for example, GSM operation. The main benefit with half-duplex FDD is the reduced terminal complexity as no duplex filter is needed in the terminal. This is especially beneficial in the case of multi-band terminals which otherwise would need multiple sets of duplex filters.

7.1.6.2 Bandwidth Flexibility

An important characteristic of LTE is the possibility for different transmission bandwidths on both downlink and uplink. The main reason for this is that the amount of spectrum available for LTE deployment may vary significantly between different frequency bands and also depending on the exact situation of the operator. Furthermore, the possibility of operating in different spectrum allocations gives the possibility for gradual migration of spectrum from other radio-access technologies to LTE.

LTE supports operation in a wide range of spectrum allocations, achieved by a flexible transmission bandwidth being part of the LTE specifications. To efficiently support very high data rates when spectrum is available, a wide transmission bandwidth is necessary, as discussed in Chapter 2. However, a sufficiently large amount of spectrum may not always be available, either due to the band of operation or due to a gradual migration from another radio-access technology, in which case LTE can be operated with a more narrow transmission bandwidth. Obviously, in such cases, the maximum achievable data rates will be reduced accordingly. As discussed below, the spectrum flexibility is further improved in later releases of LTE.

The LTE physical-layer specifications [61–64] are bandwidth agnostic and do not make any particular assumption on the supported transmission bandwidths beyond a minimum value. As will be seen in the following, the basic radio-access specification, including the physical-layer and protocol specifications, allows for any transmission bandwidth ranging from roughly 1 MHz up to around 20 MHz. At the same time, at an initial stage, radio-frequency requirements are only specified for a limited subset of transmission bandwidths, corresponding to what is predicted to be relevant spectrum-allocation sizes and relevant migration scenarios. Thus, in practice the LTE radio-access technology supports a limited set of transmission bandwidths, but additional transmission bandwidths can easily be introduced by updating only the RF specifications.

7.2 LTE RELEASE 9

After completing the first release of LTE, work continued in 3GPP with introducing additional functionality in the second release of the LTE specifications, release 9. The main enhancements seen in release 9, completed in late 2009, were support for multicast transmission, support for network-assisted positioning services, and enhancements to beam-forming in the downlink.

7.2.1 Multicast and Broadcast Support

Multi-cell broadcast implies transmission of the same information from multiple cells, as described in Chapter 3. By exploiting this at the terminal, effectively using signal power from multiple cell sites at the detection, a substantial improvement in coverage (or higher broadcast data rates) can be achieved. By transmitting not only identical signals from multiple cell sites (with identical coding and modulation), but also synchronizing the transmission timing between the cells, the signal at the terminal will appear exactly as a signal transmitted from a single cell site and subject to multi-path propagation. Due to the OFDM robustness to multi-path propagation, such multi-cell transmission, in 3GPP also referred to as *Multicast/Broadcast Single-Frequency Network* (MBSFN) transmission, will then not only improve the received signal strength, but also eliminate the inter-cell interference as described in

Chapter 3. Thus, with OFDM, multi-cell broadcast/multicast throughput may eventually be limited by noise only and can then, in the case of small cells, reach extremely high values.

It should be noted that the use of MBSFN transmission for multi-cell broadcast/multicast assumes the use of tight synchronization and time alignment of the signals transmitted from different cell sites.

7.2.2 Positioning

Positioning, as the name implies, refers to functionality in the radio-access network to determine the location of individual terminals. Determining the position of a terminal can, in principle, be done by including a GPS receiver in the terminal. Although this is a quite common feature, not all terminals include the necessary GPS receiver and there may also be cases when the GPS service is not available. LTE release 9 therefore introduces positioning support inherent in the radio-access network. By measuring on special reference signals transmitted regularly from different cell sites, the location of the terminal can be determined.

7.2.3 Dual-Layer Beam-Forming

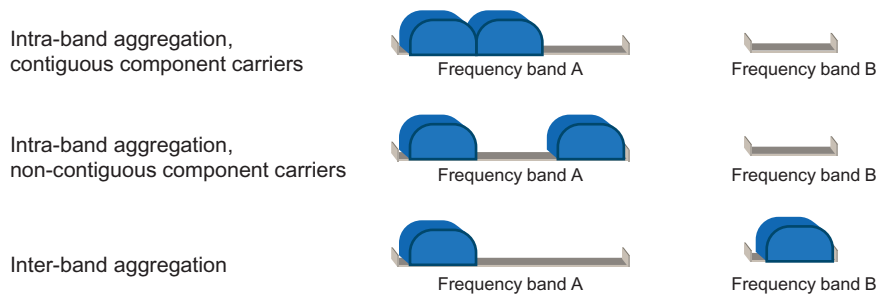
Release 9 enhances the support for combining spatial multiplexing with beam-forming. Although the combination of beam-forming and spatial multiplexing was already possible in release 8, this was then restricted to so-called *codebook-based precoding* (see Chapter 10). In release 9, spatial multiplexing can be combined with so-called *non-codebook-based precoding*, thereby improving the flexibility in deploying various multi-antenna schemes. The enhancements can also be seen as the basis for a further improvement in the area in LTE release 10.

7.3 LTE RELEASE 10 AND IMT-ADVANCED

As described in Chapter 1, *IMT-Advanced* is the term used by the ITU for radio-access technologies beyond IMT-2000. As a first step in defining IMT-Advanced, the ITU defined a set of requirements that any IMT-Advanced compliant technology should fulfill [65]. Examples of these requirements are support for at least 40 MHz bandwidth, peak spectral efficiencies of 15 bit/s/Hz in downlink and 6.75 bit/s/Hz in uplink (corresponding to peak rates of at least 600 and 270 Mbit/s respectively), and control and user plane latency of less than 100 and 10 ms respectively.

One of the main targets of LTE release 10 was to ensure that the LTE radio-access technology would be fully compliant with the IMT-Advanced requirements, thus the name *LTE-Advanced* is often used for LTE release 10. However, in addition to the ITU requirements, 3GPP also defined its own targets and requirements [66] for LTE release 10 (LTE-Advanced). These targets/requirements extended the ITU requirements both in terms of being more aggressive as well as including additional requirements. One important requirement was *backwards compatibility*. Essentially this means that an earlier-release LTE terminal should always be able to access a carrier supporting LTE release-10 functionality, although obviously not being able to utilize all the release-10 features of that carrier.

LTE release 10 was completed in late 2010 and enhances LTE spectrum flexibility through *carrier aggregation*, further extends multi-antenna transmission, introduces support for *relaying*, and provides improvements in the area of inter-cell interference coordination in *heterogeneous network deployments*.

**FIGURE 7.4**

Carrier aggregation.

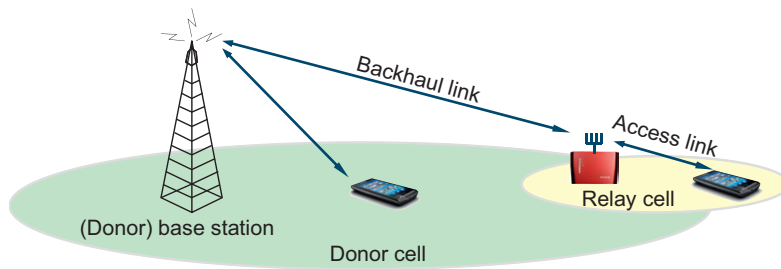
7.3.1 Carrier Aggregation

As mentioned earlier, the first release of LTE already provided extensive support for deployment in spectrum allocations of various characteristics, with bandwidths ranging from roughly 1 MHz up to 20 MHz in both paired and unpaired bands. In LTE release 10 the transmission bandwidth can be further extended by means of so-called *carrier aggregation* (CA), where multiple *component carriers* are aggregated and jointly used for transmission to/from a single terminal. Up to five component carriers, possibly each of different bandwidth, can be aggregated, allowing for transmission bandwidths up to 100 MHz. Backwards compatibility is catered for as each component carrier uses the release-8 structure. Hence, to a release-8/9 terminal each component carrier will appear as an LTE release-8 carrier, while a carrier-aggregation-capable terminal can exploit the total aggregated bandwidth, enabling higher data rates. In the general case, a different number of component carriers can be aggregated for the downlink and uplink.

Component carriers do not have to be contiguous in frequency, which enables exploitation of *fragmented spectrum*; operators with a fragmented spectrum can provide high-data-rate services based on the availability of a wide overall bandwidth even though they do not possess a single wideband spectrum allocation. From a baseband perspective, there is no difference between the cases in Figure 7.4 and they are all supported by LTE release 10. However, the RF-implementation complexity is vastly different with the first case being the least complex. Thus, although spectrum aggregation is supported by the basic specifications, the actual release-10 RF requirements will be strongly constrained, including specification of only a limited number of aggregation scenarios and including support of inter-band aggregation only for the most advanced terminals, but excluding non-contiguous intra-band aggregation.

7.3.2 Extended Multi-Antenna Transmission

In release 10, downlink spatial multiplexing is expanded to support up to eight transmission layers. Along with this, and as discussed in more detail in Chapter 10 an enhanced *reference-signal structure* is introduced to improve the support of various beam-forming solutions. This can be seen as an extension of the release-9 dual-layer beam-forming to support up to eight antenna ports and eight corresponding layers. Together with the support for carrier aggregation, this enables downlink data rates up to 3 Gbit/s or 30 bit/s/Hz.

**FIGURE 7.5**

Example of relaying.

Uplink spatial multiplexing of up to four layers is also part of release 10. It consists of a codebook-based scheme under control of the base station, which means that the structure can also be used for uplink transmit-side beam-forming. Together with the possibility for uplink carrier aggregations, this allows for uplink data rates up to 1.5 Gbit/s or 15 bit/s/Hz.

7.3.3 Relaying

Relaying implies that the terminal communicates with the network via a *relay node* that is *wirelessly connected* to a *donor cell* using the LTE radio-interface technology (Figure 7.5). From a terminal point of view, the relay node will appear as an ordinary cell. This has the important advantage of simplifying the terminal implementation and making the relay node backwards compatible – that is, also accessible to LTE release-8/9 terminals. In essence, the relay is a low-power base station wirelessly connected to the remaining part of the network. One of the attractive features of a relay is the LTE-based wireless backhaul, as this could provide a simple way of improving coverage, for example in indoor environments, by simply placing relays at the problematic locations. At a later stage, if motivated by the traffic situation, the wireless donor-relay link could be replaced by an optical fiber in order to use the precious radio resources in the donor cell for terminal communication instead of serving the relay.

7.3.4 Heterogeneous Deployments

Heterogeneous deployments refer to deployments with a mixture of cells with different downlink transmission power, operating on (partially) the same set of frequencies and with overlapping geographical coverage (Figure 7.6). A typical example is a pico cell placed within the coverage area of a macro cell. Although such deployments were already supported in release 8, release 10 introduced improved inter-cell interference handling focusing on scenarios with large power differences between overlapping cells.

7.4 TERMINAL CAPABILITIES

To support different scenarios, which may call for different terminal capabilities in terms of data rates, as well as to allow for market differentiation in terms of low- and high-end terminals with a

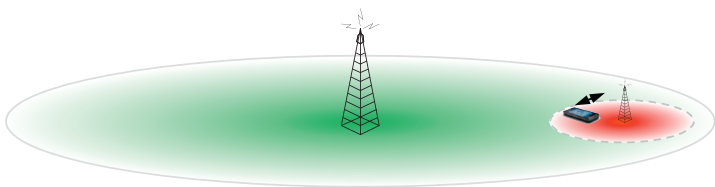


FIGURE 7.6
Example of heterogeneous deployment with a pico cell inside a macro cell.

Table 7.1 UE Categories								
	Category							
	Release 8/9/10					Release 10 only		
	1	2	3	4	5	6	7	8
Downlink peak rate (Mbit/s)	10	50	100	150	300	300	300	3000
Uplink peak rate (Mbit/s)	5	25	50	50	75	50	150	1500
Maximum downlink modulation	64QAM							
Maximum uplink modulation	16QAM				64QAM	16QAM	64QAM	
Max. number of layers for downlink spatial multiplexing	1	2			4	Signaled separately		

corresponding difference in price, not all terminals support all capabilities. Furthermore, terminals from an earlier release of the standard will not support features introduced in later versions of LTE. For example, a release-8 terminal will obviously not support carrier aggregation as this feature was introduced in release 10. Therefore, as part of the connection setup, the terminal indicates not only which release of LTE it supports, but also its capabilities within the release.

In principle, the different parameters could be specified separately, but to limit the number of combinations and avoid a parameter combination that does not make sense, a set of physical-layer capabilities are lumped together to form a UE category (UE, *User Equipment*, is the term used in 3GPP to denote a terminal). In total five different UE categories have been specified for LTE release 8/9, ranging from the low-end category 1 not supporting spatial multiplexing to the high-end category 5 supporting the full set of features in the release-8/9 physical layer specifications. The categories are summarized in Table 7.1 (in simplified form; for the full set of details, see [67]). Note that, regardless

of the category, a terminal is always capable of receiving transmissions from up to four antenna ports. This is necessary as the system information can be transmitted on up to four antenna ports.

In LTE release 10, features such as carrier aggregation and uplink spatial multiplexing are introduced, which calls for additional capability signaling compared to release 8/9, either in the form of additional UE categories or as separate capabilities. Defining new categories for each foreseen combination of the maximum number of component carriers and maximum degree of spatial multiplexing could be done in principle, although the number of categories might become very large and which categories a terminal support may be frequency-band dependent. Therefore, in release 10, three additional UE categories were defined as seen in Table 7.1, and the maximum number of component carriers and degree of spatial multiplexing supported, both in uplink and downlink, are signaled separately from the category number. A release-10 terminal may therefore declare itself as, for example, category 4 but capable of uplink spatial multiplexing. Hence, categories 1–5 may have a slightly different meaning for a release-8/9 and a release-10 terminal, depending on the value of the separately declared capabilities. Furthermore, in order to be able to operate in release-8/9 networks, a release-10 UE has to be able to declare both release-8/9 and release-10 categories.

In addition to the capabilities mentioned in the different UE categories, there are some capabilities specified outside the categories. The duplexing schemes supported is one such example, the support of UE-specific reference signals for FDD in release 8 is another. Whether the terminal supports other radio-access technologies, for example GSM and WCDMA, is also declared separately.

Radio-Interface Architecture

8

This chapter contains a brief overview of the overall architecture of an LTE radio-access network and the associated core network, followed by descriptions of the radio-access network user-plane and control-plane protocols.

8.1 OVERALL SYSTEM ARCHITECTURE

In parallel to the work on the LTE radio-access technology in 3GPP, the overall system architecture of both the *Radio-Access Network* (RAN) and the *Core Network* (CN) was revisited, including the split of functionality between the two network parts. This work was known as the *System Architecture Evolution* (SAE) and resulted in a flat RAN architecture, as well as a new core network architecture referred to as the *Evolved Packet Core* (EPC). Together, the LTE RAN and the EPC can be referred to as the *Evolved Packet System* (EPS).¹

The RAN is responsible for all radio-related functionality of the overall network including, for example, scheduling, radio-resource handling, retransmission protocols, coding and various multi-antenna schemes. These functions will be discussed in detail in the subsequent chapters.

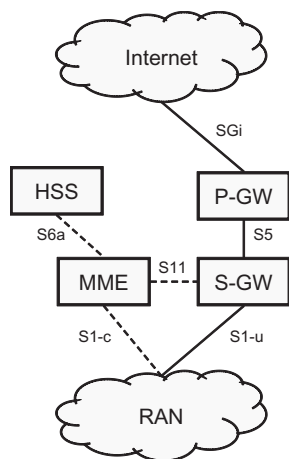
The EPC is responsible for functions not related to the radio interface but needed for providing a complete mobile-broadband network. This includes, for example, authentication, charging functionality, and setup of end-to-end connections. Handling these functions separately, instead of integrating them into the RAN, is beneficial as it allows for several radio-access technologies to be served by the same core network.

Although this book focuses on the LTE RAN, a brief overview of the EPC, as well as how it connects to the RAN, is useful. For an excellent in-depth discussion of EPC, the reader is referred to [9].

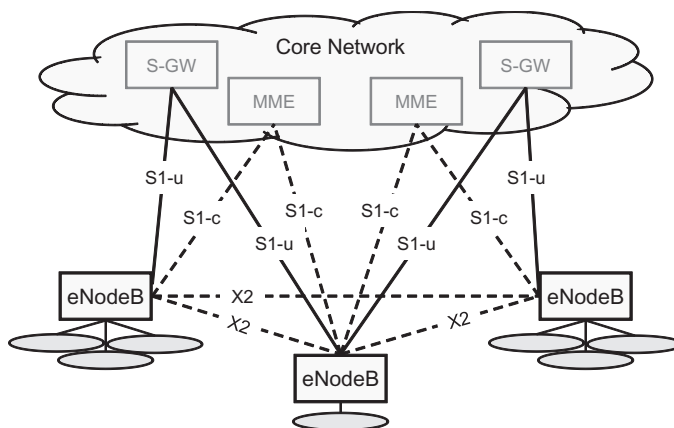
8.1.1 Core Network

The EPC is a radical evolution from the GSM/GPRS core network used for GSM and WCDMA/HSPA. EPC supports access to the *packet-switched domain* only, with no access to the *circuit-switched domain*. It consists of several different types of nodes, some of which are briefly described below and illustrated in [Figure 8.1](#).

¹UTRAN, the WCDMA/HSPA radio-access network, is also part of the EPS.

**FIGURE 8.1**

Core-network (EPC) architecture.

**FIGURE 8.2**

Radio-access-network interfaces.

The *Mobility Management Entity* (MME) is the control-plane node of the EPC. Its responsibilities include connection/release of bearers to a terminal, handling of IDLE to ACTIVE transitions, and handling of security keys. The functionality operating between the EPC and the terminal is sometimes referred to as the *Non-Access Stratum* (NAS), to separate it from the *Access Stratum* (AS) which handles functionality operating between the terminal and the radio-access network.

The *Serving Gateway* (S-GW) is the user-plane node connecting the EPC to the LTE RAN. The S-GW acts as a mobility anchor when terminals move between eNodeBs (see next section), as well as a mobility anchor for other 3GPP technologies (GSM/GPRS and HSPA). Collection of information and statistics necessary for charging is also handled by the S-GW.

The *Packet Data Network Gateway* (PDN Gateway, P-GW) connects the EPC to the internet. Allocation of the IP address for a specific terminal is handled by the P-GW, as well as quality-of-service enforcement according to the policy controlled by the PCRF (see below). The P-GW is also the mobility anchor for non-3GPP radio-access technologies, such as CDMA2000, connected to the EPC.

In addition, the EPC also contains other types of nodes such as *Policy and Charging Rules Function* (PCRF) responsible for quality-of-service (QoS) handling and charging, and the *Home Subscriber Service* (HSS) node, a database containing subscriber information. There are also some additional nodes present as regards network support of *Multimedia Broadcast Multicast Services* (MBMS) (see Chapter 15 for a more detailed description of MBMS, including the related architecture aspects).

It should be noted that the nodes discussed above are *logical* nodes. In an actual physical implementation, several of them may very well be combined. For example, the MME, P-GW, and S-GW could very well be combined into a single physical node.

8.1.2 Radio-Access Network

The LTE radio-access network uses a flat architecture with a single type of node² – the *eNodeB*. The eNodeB is responsible for all radio-related functions in one or several cells. It is important to note that an eNodeB is a *logical* node and not a physical implementation. One common implementation of an eNodeB is a three-sector site, where a base station is handling transmissions in three cells, although other implementations can be found as well, such as one baseband processing unit to which a number of remote radio heads are connected. One example of the latter is a large number of indoor cells, or several cells along a highway, belonging to the same eNodeB. Thus, a base station is a *possible* implementation of, but not *the same* as, an eNodeB.

As can be seen in Figure 8.2, the eNodeB is connected to the EPC by means of the *S1 interface*, more specifically to the S-GW by means of the *S1 user-plane part*, S1-u, and to the MME by means of the *S1 control-plane part*, S1-c. One eNodeB can be connected to multiple MMEs/S-GWs for the purpose of load sharing and redundancy.

The *X2 interface*, connecting eNodeBs to each other, is mainly used to support active-mode mobility. This interface may also be used for multi-cell *Radio Resource Management* (RRM) functions such as *Inter-Cell Interference Coordination* (ICIC) discussed in Chapter 13. The X2 interface is also used to support lossless mobility between neighboring cells by means of packet forwarding.

8.2 RADIO PROTOCOL ARCHITECTURE

With the overall network architecture in mind, the RAN protocol architecture for the user as well as the control planes can be discussed. Figure 8.3 illustrates the RAN protocol architecture (the MME is, as discussed in the previous section, not part of the RAN but is included in the figure for completeness). As seen in the figure, many of the protocol entities are common to the user and control planes. Therefore, although this section mainly describes the protocol architecture from a user-plane perspective, the description is in many respects also applicable to the control plane. Control-plane-specific aspects are discussed in Section 8.3.

The LTE radio-access network provides one or more *Radio Bearers* to which IP packets are mapped according to their Quality-of-Service requirements. A general overview of the LTE (user-plane) protocol architecture for the downlink is illustrated in Figure 8.4. As will become clear in the subsequent discussion, not all the entities illustrated in Figure 8.4 are applicable in all situations. For example, neither MAC scheduling nor hybrid ARQ with soft combining is used for broadcast of the basic system information. The LTE protocol structure related to uplink transmissions is similar to the downlink structure in Figure 8.4, although there are some differences with respect to, for example, transport-format selection.

The different protocol entities of the radio-access network are summarized below and described in more detail in the following sections.

- *Packet Data Convergence Protocol* (PDCP) performs IP header compression to reduce the number of bits to transmit over the radio interface. The header-compression mechanism is based

²The introduction of MBMS (see Chapter 15) in release 9 and relaying (see Chapter 16) in release 10 brings additional node types to the RAN.

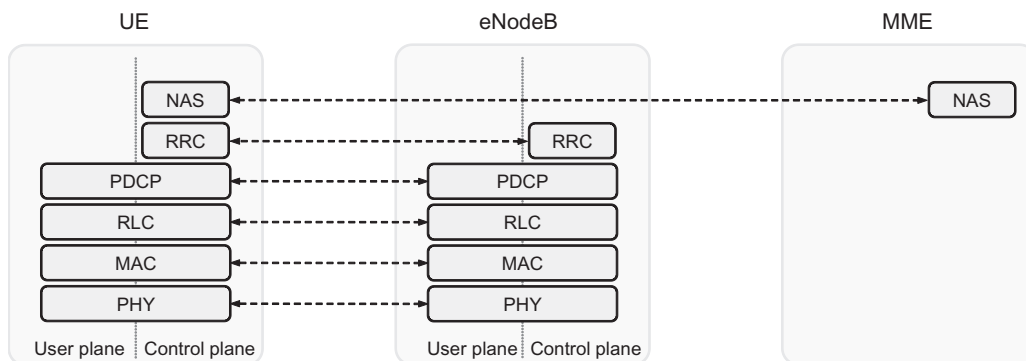


FIGURE 8.3

Overall RAN protocol architecture.

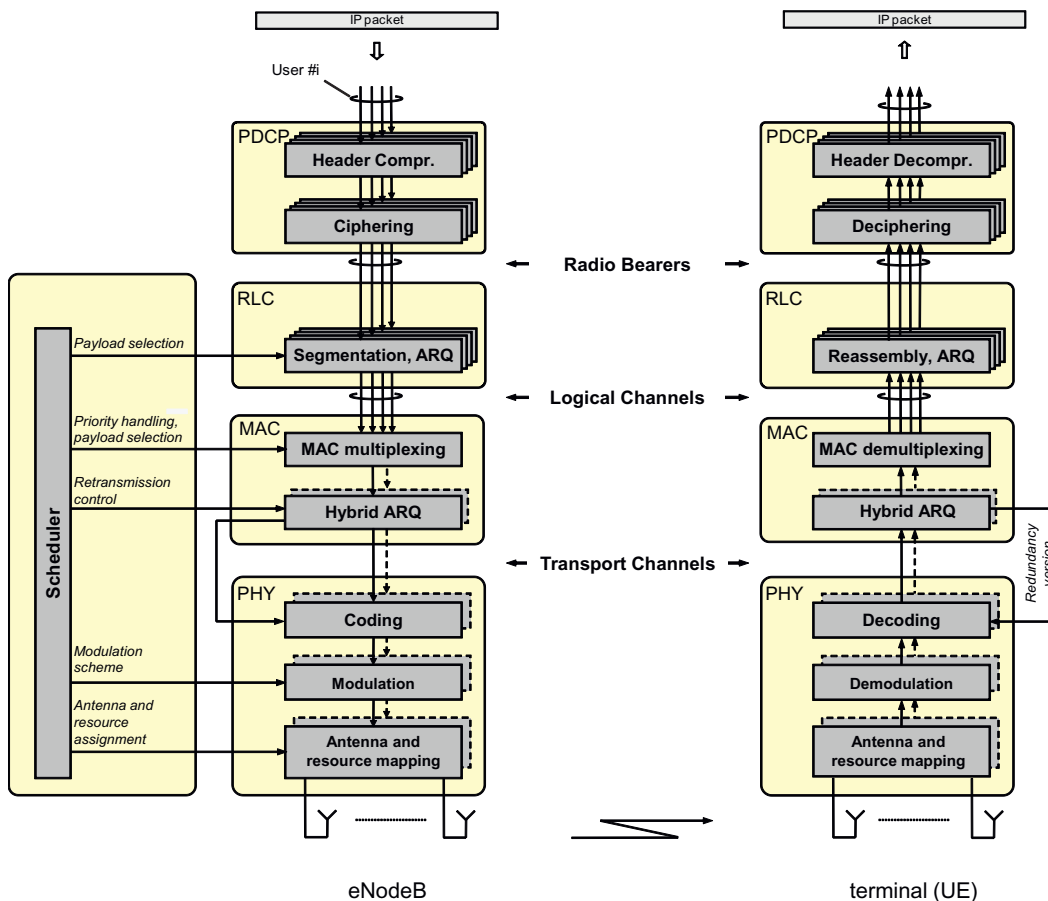


FIGURE 8.4

LTE protocol architecture (downlink).

on Robust Header Compression (ROHC) [68], a standardized header-compression algorithm also used for several mobile-communication technologies. PDCP is also responsible for ciphering and, for the control plane, integrity protection of the transmitted data, as well as in-sequence delivery and duplicate removal for handover. At the receiver side, the PDCP protocol performs the corresponding deciphering and decompression operations. There is one PDCP entity per radio bearer configured for a terminal.

- *Radio-Link Control* (RLC) is responsible for segmentation/concatenation, retransmission handling, duplicate detection, and in-sequence delivery to higher layers. The RLC provides services to the PDCP in the form of *radio bearers*. There is one RLC entity per radio bearer configured for a terminal.
- *Medium-Access Control* (MAC) handles multiplexing of logical channels, hybrid-ARQ retransmissions, and uplink and downlink scheduling. The scheduling functionality is located in the eNodeB for both uplink and downlink. The hybrid-ARQ protocol part is present in both the transmitting and receiving ends of the MAC protocol. The MAC provides services to the RLC in the form of *logical channels*.
- *Physical Layer* (PHY) handles coding/decoding, modulation/demodulation, multi-antenna mapping, and other typical physical-layer functions. The physical layer offers services to the MAC layer in the form of *transport channels*.

To summarize the flow of downlink data through all the protocol layers, an example illustration for a case with three IP packets, two on one radio bearer and one on another radio bearer, is given in [Figure 8.5](#). The data flow in the case of uplink transmission is similar. The PDCP performs (optional) IP-header compression, followed by ciphering. A PDCP header is added, carrying information required for deciphering in the terminal. The output from the PDCP is forwarded to the RLC.

The RLC protocol performs concatenation and/or segmentation of the PDCP SDUs³ and adds an RLC header. The header is used for in-sequence delivery (per logical channel) in the terminal and for identification of RLC PDUs in the case of retransmissions. The RLC PDUs are forwarded to the MAC layer, which multiplexes a number of RLC PDUs and attaches a MAC header to form a transport block. The transport-block size depends on the instantaneous data rate selected by the link-adaptation mechanism. Thus, the link adaptation affects both the MAC and RLC processing. Finally, the physical layer attaches a CRC to the transport block for error-detection purposes, performs coding and modulation, and transmits the resulting signal, possibly using multiple transmit antennas.

The remainder of the chapter contains an overview of the RLC, MAC, and physical layers. A more detailed description of the LTE physical-layer processing is given in Chapters 10 (downlink) and 11 (uplink), followed by descriptions of some specific uplink and downlink radio-interface functions and procedures in the subsequent chapters.

8.2.1 Radio-Link Control

The RLC protocol is responsible for segmentation/concatenation of (header-compressed) IP packets, also known as RLC SDUs, from the PDCP into suitably sized RLC PDUs. It also handles

³In general, the data entity from/to a higher protocol layer is known as a Service Data Unit (SDU) and the corresponding entity to/from a lower protocol layer entity is called a Protocol Data Unit (PDU).

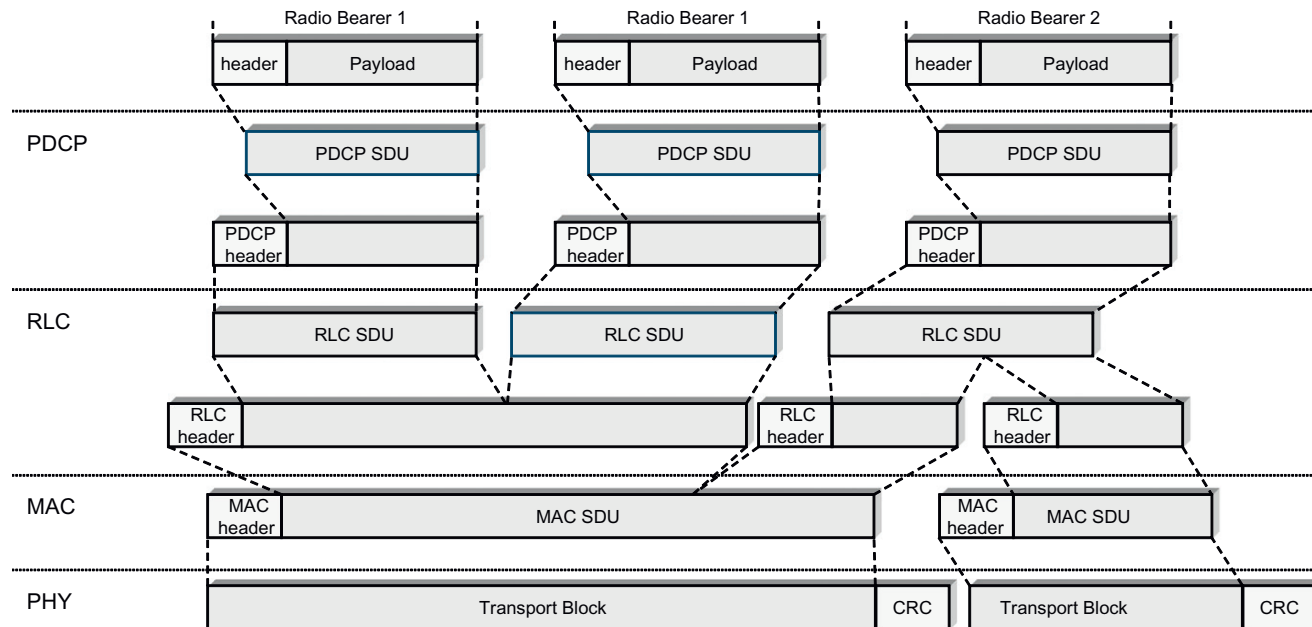
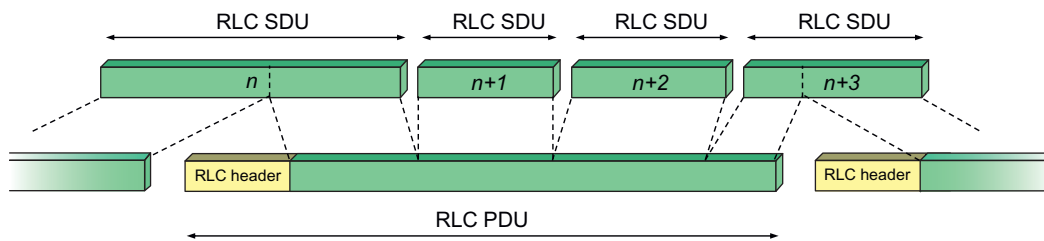


FIGURE 8.5

Example of LTE data flow.

**FIGURE 8.6**

RLC segmentation and concatenation.

retransmission of erroneously received PDUs, as well as removal of duplicated PDUs. Finally, the RLC ensures in-sequence delivery of SDUs to upper layers. Depending on the type of service, the RLC can be configured in different modes to perform some or all of these functions.

Segmentation and concatenation, one of the main RLC functions, is illustrated in Figure 8.6. Depending on the scheduler decision, a certain amount of data is selected for transmission from the RLC SDU buffer and the SDUs are segmented/concatenated to create the RLC PDU. Thus, for LTE the RLC PDU size varies *dynamically*. For high data rates, a large PDU size results in a smaller relative overhead, while for low data rates, a small PDU size is required as the payload would otherwise be too large. Hence, as the LTE data rates may range from a few kbit/s up to 3 Gbit/s, dynamic PDU sizes are motivated for LTE in contrast to earlier mobile-communication technologies, which typically use a fixed PDU size. Since the RLC, scheduler, and rate adaptation mechanisms are all located in the eNodeB, dynamic PDU sizes are easily supported for LTE. In each RLC PDU, a header is included, containing, among other things, a sequence number used for in-sequence delivery and by the retransmission mechanism.

The RLC retransmission mechanism is also responsible for providing error-free delivery of data to higher layers. To accomplish this, a retransmission protocol operates between the RLC entities in the receiver and transmitter. By monitoring the sequence numbers of the incoming PDUs, the receiving RLC can identify missing PDUs. Status reports are then fed back to the transmitting RLC entity, requesting retransmission of missing PDUs. Based on the received status report, the RLC entity at the transmitter can take the appropriate action and retransmit the missing PDUs if needed.

Although the RLC is capable of handling transmission errors due to noise, unpredictable channel variations, etc., error-free delivery is in most cases handled by the MAC-based hybrid-ARQ protocol. The use of a retransmission mechanism in the RLC may therefore seem superfluous at first. However, as will be discussed in Section 8.2.2.3, this is not the case and the use of both RLC- and MAC-based retransmission mechanisms is in fact well motivated by the differences in the feedback signaling.

The details of RLC are further described in Chapter 12.

8.2.2 Medium-Access Control

The MAC layer handles logical-channel multiplexing, hybrid-ARQ retransmissions, and uplink and downlink scheduling. It is also responsible for multiplexing/demultiplexing data across multiple component carriers when carrier aggregation is used.

8.2.2.1 Logical Channels and Transport Channels

The MAC provides services to the RLC in the form of *logical channels*. A logical channel is defined by the *type* of information it carries and is generally classified as a *control channel*, used for transmission of control and configuration information necessary for operating an LTE system, or as a *traffic channel*, used for the user data. The set of logical-channel types specified for LTE includes:

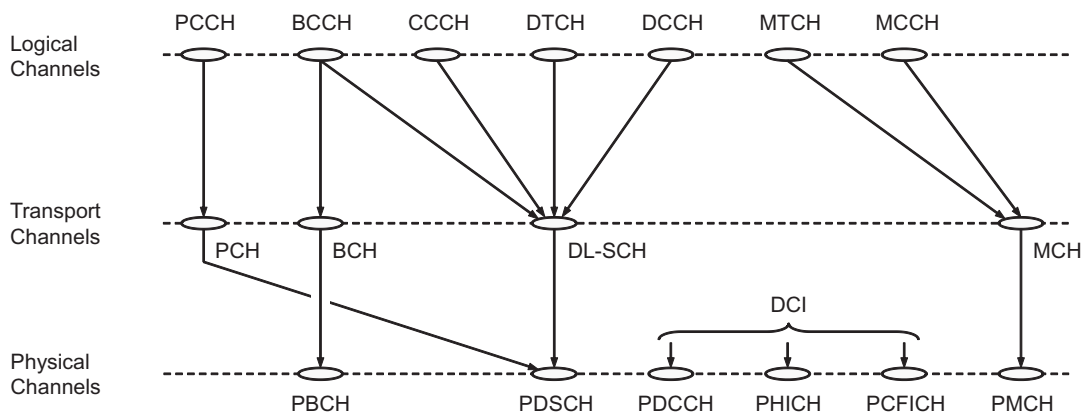
- The *Broadcast Control Channel* (BCCH), used for transmission of *system information* from the network to all terminals in a cell. Prior to accessing the system, a terminal needs to acquire the system information to find out how the system is configured and, in general, how to behave properly within a cell.
- The *Paging Control Channel* (PCCH), used for paging of terminals whose location on a cell level is not known to the network. The paging message therefore needs to be transmitted in multiple cells.
- The *Common Control Channel* (CCCH), used for transmission of control information in conjunction with random access.
- The *Dedicated Control Channel* (DCCH), used for transmission of control information to/from a terminal. This channel is used for individual configuration of terminals such as different handover messages.
- The *Multicast Control Channel* (MCCH), used for transmission of control information required for reception of the MTCH (see below).
- The *Dedicated Traffic Channel* (DTCH), used for transmission of user data to/from a terminal. This is the logical channel type used for transmission of all uplink and non-MBSFN downlink user data.
- The *Multicast Traffic Channel* (MTCH), used for downlink transmission of MBMS services.

From the physical layer, the MAC layer uses services in the form of *transport channels*. A transport channel is defined by *how* and *with what characteristics* the information is transmitted over the radio interface. Data on a transport channel is organized into *transport blocks*. In each *Transmission Time Interval* (TTI), at most one transport block of dynamic size is transmitted over the radio interface to/from a terminal in the absence of spatial multiplexing. In the case of spatial multiplexing (MIMO), there can be up to two transport blocks per TTI.

Associated with each transport block is a *Transport Format* (TF), specifying *how* the transport block is to be transmitted over the radio interface. The transport format includes information about the transport-block size, the modulation-and-coding scheme, and the antenna mapping. By varying the transport format, the MAC layer can thus realize different data rates. Rate control is therefore also known as *transport-format selection*.

The following transport-channel types are defined for LTE:

- The *Broadcast Channel* (BCH) has a fixed transport format, provided by the specifications. It is used for transmission of parts of the BCCH system information, more specifically the so-called *Master Information Block* (MIB), as described in Chapter 14.
- The *Paging Channel* (PCH) is used for transmission of paging information from the PCCH logical channel. The PCH supports *discontinuous reception* (DRX) to allow the terminal to save battery power by waking up to receive the PCH only at predefined time instants. The LTE paging mechanism is also described in Chapter 14.

**FIGURE 8.7**

Downlink channel mapping.

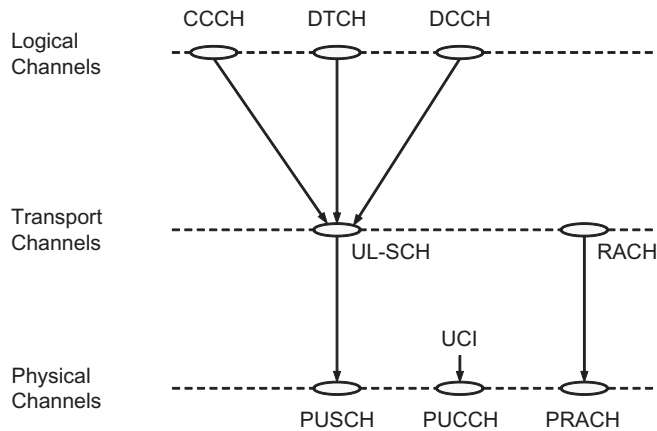
- The *Downlink Shared Channel* (DL-SCH) is the main transport channel used for transmission of downlink data in LTE. It supports key LTE features such as dynamic rate adaptation and channel-dependent scheduling in the time and frequency domains, hybrid ARQ with soft combining, and spatial multiplexing. It also supports DRX to reduce terminal power consumption while still providing an always-on experience. The DL-SCH is also used for transmission of the parts of the BCCH system information not mapped to the BCH. There can be multiple DL-SCHs in a cell, one per terminal⁴ scheduled in this TTI, and, in some subframes, one DL-SCH carrying system information.
- The *Multicast Channel* (MCH) is used to support MBMS. It is characterized by a semi-static transport format and semi-static scheduling. In the case of multi-cell transmission using MBSFN, the scheduling and transport format configuration is coordinated among the transmission points involved in the MBSFN transmission. MBSFN transmission is described in Chapter 15.
- The Uplink Shared Channel (UL-SCH) is the uplink counterpart to the DL-SCH – that is, the uplink transport channel used for transmission of uplink data.

In addition, the *Random-Access Channel* (RACH) is also defined as a transport channel, although it does not carry transport blocks.

Part of the MAC functionality is multiplexing of different logical channels and mapping of the logical channels to the appropriate transport channels. The supported mappings between logical-channel types and transport-channel types are given in Figure 8.7 for the downlink and Figure 8.8 for the uplink. The figures clearly indicate how DL-SCH and UL-SCH are the main downlink and uplink transport channels respectively. In the figures, the corresponding physical channels, described further below, are also included and the mapping between transport channels and physical channels is illustrated.

To support priority handling, multiple logical channels, where each logical channel has its own RLC entity, can be multiplexed into one transport channel by the MAC layer. At the receiver, the

⁴For carrier aggregation, a UE may receive multiple DL-SCHs, one per component carrier.

**FIGURE 8.8**

Uplink channel mapping.

MAC layer handles the corresponding demultiplexing and forwards the RLC PDUs to their respective RLC entity for in-sequence delivery and the other functions handled by the RLC. To support the demultiplexing at the receiver, a MAC header, shown in Figure 8.9, is used. To each RLC PDU, there is an associated sub-header in the MAC header. The sub-header contains the identity of the logical channel (LCID) from which the RLC PDU originated and the length of the PDU in bytes. There is also a flag indicating whether this is the last sub-header or not. One or several RLC PDUs, together with the MAC header and, if necessary, padding to meet the scheduled transport-block size, form one transport block which is forwarded to the physical layer.

In addition to multiplexing of different logical channels, the MAC layer can also insert the so-called *MAC control elements* into the transport blocks to be transmitted over the transport channels. A MAC control element is used for inband control signaling – for example, timing-advance commands and random-access response, as described in Sections 11.5 and 14.3 respectively. Control elements are identified with reserved values in the LCID field, where the LCID value indicates the type of control information. Furthermore, the length field in the sub-header is removed for control elements with a fixed length.

The MAC multiplexing functionality is also responsible for handling of multiple component carriers in the case of carrier aggregation. The basic principle for carrier aggregation is independent processing of the component carriers in the physical layer, including control signaling, scheduling and hybrid-ARQ retransmissions, while carrier aggregation is invisible to RLC and PDCP. Carrier aggregation is therefore mainly seen in the MAC layer, as illustrated in Figure 8.10, where logical channels, including any MAC control elements, are multiplexed to form one (two in the case of spatial multiplexing) transport block(s) per component carrier with each component carrier having its own hybrid-ARQ entity.

8.2.2.2 Scheduling

One of the basic principles of LTE radio access is shared-channel transmission – that is, time–frequency resources are dynamically shared between users. The *scheduler* is part of the MAC layer

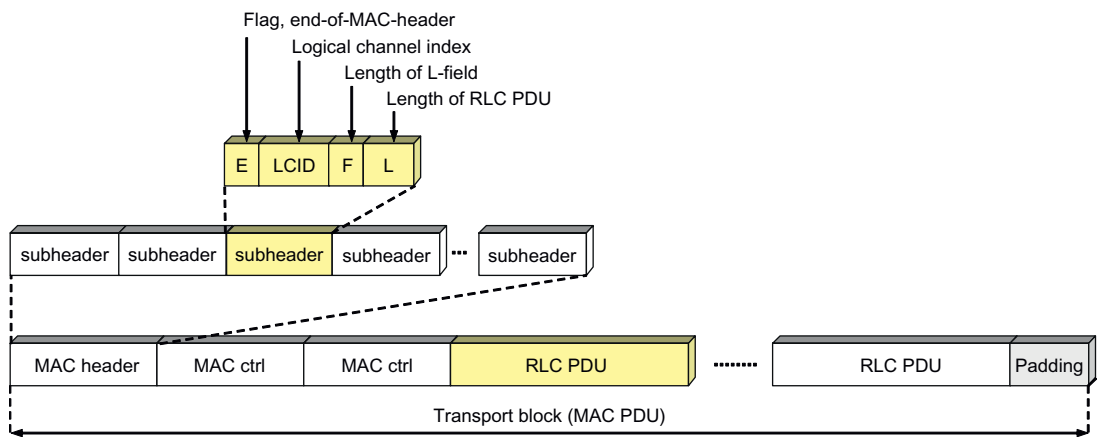


FIGURE 8.9
MAC header and SDU multiplexing.

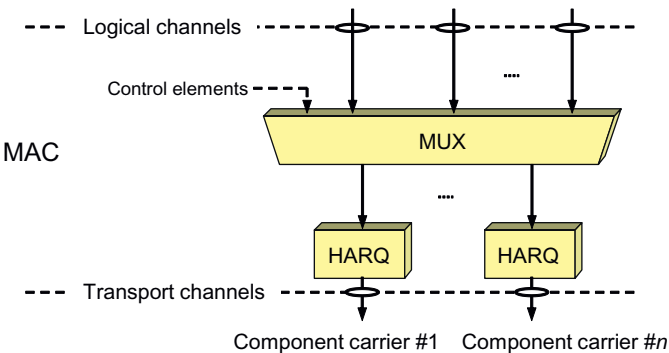


FIGURE 8.10
MAC operation in the case of multiple component carriers.

(although often better viewed as a separate entity as illustrated in Figure 8.4) and controls the assignment of uplink and downlink resources in terms of so-called *resource-block pairs*. Resource blocks correspond to a time–frequency unit of 1 ms times 180 kHz, as described in more detail in Chapter 9.

The basic operation of the scheduler is so-called *dynamic* scheduling, where the eNodeB in each 1 ms interval takes a scheduling decision and sends scheduling information to the selected set of terminals. However, there is also a possibility for semi-persistent scheduling where a semi-static scheduling pattern is signaled in advance to reduce the control-signaling overhead. Coordination of scheduling decisions across multiple cells residing in different eNodeBs is supported using signaling over the X2 interface.

Uplink and downlink scheduling are separated in LTE, and uplink and downlink scheduling decisions can be taken independently of each other (within the limits set by the uplink/downlink split in the case of half-duplex FDD operation).

The downlink scheduler is responsible for (dynamically) controlling which 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. Transport-format selection (selection of transport-block size, modulation scheme, and antenna mapping) and logical-channel multiplexing for downlink transmissions are controlled by the eNodeB, as illustrated in the left part of Figure 8.11. As a consequence of the scheduler controlling the data rate, the RLC segmentation and MAC multiplexing will also be affected by the scheduling decision. The outputs from the downlink scheduler can be seen in Figure 8.4.

The uplink scheduler serves a similar purpose, namely to (dynamically) control which terminals are to transmit on their respective UL-SCH and on which uplink time–frequency resources (including component carrier). Despite the fact that the eNodeB scheduler determines the transport format for the terminal, it is important to point out that the uplink scheduling decision is taken *per terminal* and not per radio bearer. Thus, although the eNodeB scheduler controls the payload of a scheduled terminal, the terminal is still responsible for selecting *from which radio bearer(s)* the data is taken. The terminal autonomously handles logical-channel multiplexing according to rules, the parameters of which can be configured by the eNodeB. This is illustrated in the right part of Figure 8.11, where the eNodeB scheduler controls the transport format and the terminal controls the logical-channel multiplexing.

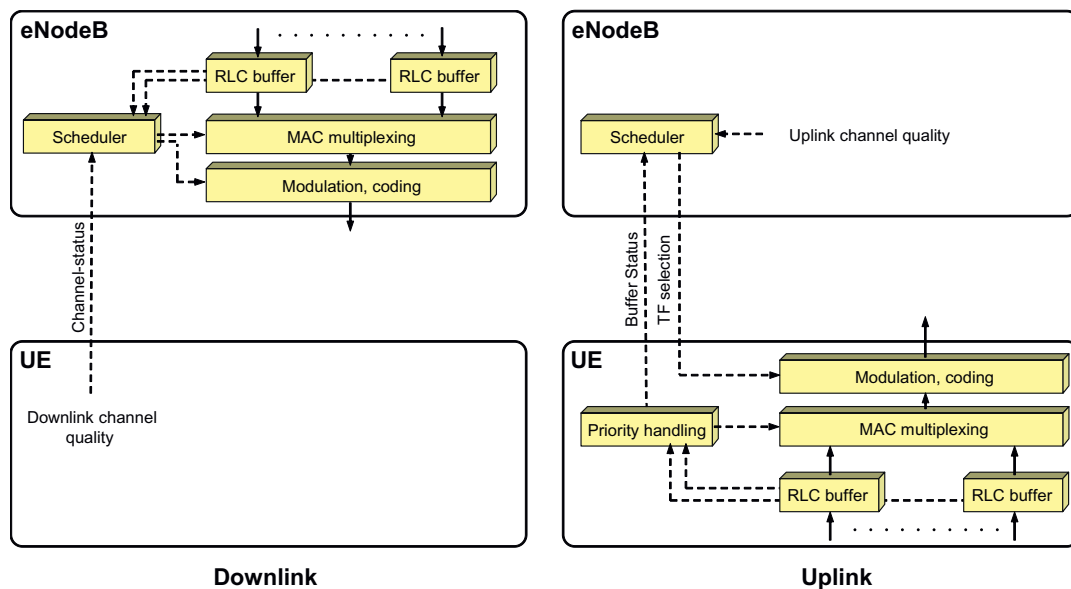


FIGURE 8.11

Transport-format selection in downlink and uplink.

Although the scheduling strategy is implementation specific and not specified by 3GPP, the overall goal of most schedulers is to take advantage of the channel variations between terminals and preferably schedule transmissions to a terminal on resources with advantageous channel conditions. A benefit of the use of OFDM in LTE is the possibility to exploit channel variations in both time and frequency domains through channel-dependent scheduling. This was mentioned earlier in Chapter 7 and illustrated in Figure 7.2. For the larger bandwidths supported by LTE, where a significant amount of frequency-selective fading will often be experienced, the possibility for the scheduler to also exploit frequency-domain channel variations becomes increasingly important compared to exploiting time-domain variations only. The possibility to also exploit frequency-domain variations is beneficial especially at low speeds, where the variations in the time domain are relatively slow compared to the delay requirements set by many services.

Downlink channel-dependent scheduling is supported through *channel-state reports*, transmitted by the UE and reflecting the instantaneous channel quality in the time and frequency domains, as well as information necessary to determine the appropriate antenna processing in the case of spatial multiplexing. In the uplink, the channel-state information necessary for uplink channel-dependent scheduling can be based on a *sounding reference signal* transmitted from each terminal for which the eNodeB wants to estimate the uplink channel quality. To aid the uplink scheduler in its decisions, the terminal can transmit buffer-status information to the eNodeB using a MAC message. Obviously, this information can only be transmitted if the terminal has been given a valid scheduling grant. For situations when this is not the case, an indicator that the terminal needs uplink resources is provided as part of the uplink L1/L2 control-signaling structure (see Chapter 11).

Interference coordination, which tries to control the inter-cell interference on a slow basis, as mentioned in Chapter 7, is also part of the scheduler. As the scheduling strategy is not mandated by the specifications, the interference-coordination scheme (if used) is vendor specific and may range from simple higher-order reuse deployments to more advanced schemes. The mechanisms supporting inter-cell interference coordination are discussed in Chapter 13.

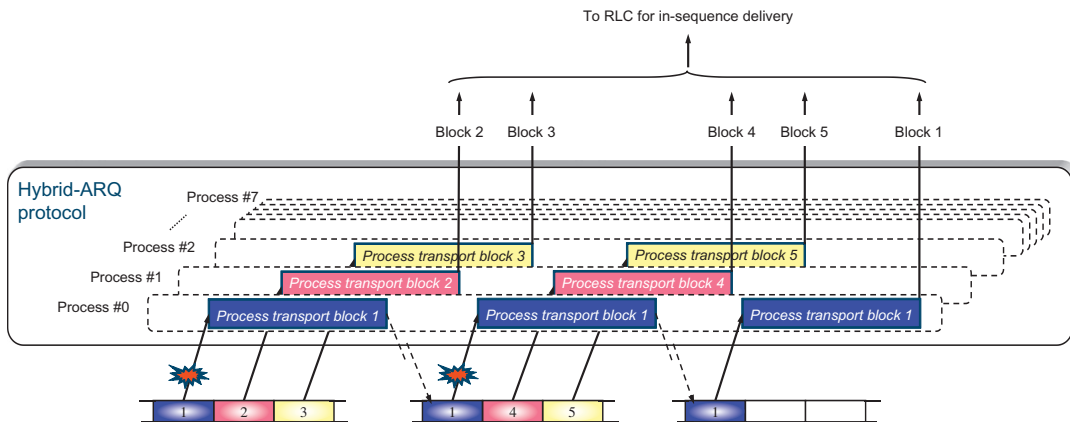
8.2.2.3 Hybrid ARQ with Soft Combining

Hybrid ARQ with soft combining provides robustness against transmission errors. As hybrid-ARQ retransmissions are fast, many services allow for one or multiple retransmissions, thereby forming an implicit (closed loop) rate-control mechanism. The hybrid-ARQ protocol is part of the MAC layer, while the actual soft combining is handled by the physical layer.⁵

Obviously, hybrid ARQ is not applicable for all types of traffic. For example, broadcast transmissions, where the same information is intended for multiple terminals, typically do not rely on hybrid ARQ. Hence, hybrid ARQ is only supported for the DL-SCH and the UL-SCH, although its usage is optional.

The LTE hybrid-ARQ protocol uses multiple parallel stop-and-wait processes. Upon reception of a transport block, the receiver makes an attempt to decode the transport block and informs the transmitter about the outcome of the decoding operation through a single acknowledgement bit indicating whether the decoding was successful or if a retransmission of the transport block is required. Further details on transmission of hybrid-ARQ acknowledgements are found in Chapters 10–12. Clearly, the receiver must know to which hybrid-ARQ process a received acknowledgement is associated. This is

⁵The soft combining is done before or as part of the channel decoding, which clearly is a physical-layer functionality.

**FIGURE 8.12**

Multiple parallel hybrid-ARQ processes.

solved by using the timing of the acknowledgement for association with a certain hybrid-ARQ process. Note that, in the case of TDD operation, the time relation between the reception of data in a certain hybrid-ARQ process and the transmission of the acknowledgement is also affected by the uplink/downlink allocation.

The use of multiple parallel hybrid-ARQ processes, illustrated in Figure 8.12, for each user can result in data being delivered from the hybrid-ARQ mechanism out of sequence. For example, transport block 5 in the figure was successfully decoded before transport block 1 which required retransmissions. In-sequence delivery of data is therefore ensured by the RLC layer. The same reordering mechanism in the RLC is also used to handle in-sequence delivery across multiple component carriers in the case of carrier aggregation. As mentioned above, the hybrid-ARQ retransmissions are handled independently per component carrier, which may result in out-of-sequence delivery in a similar way as within a component carrier.

Downlink retransmissions may occur at any time after the initial transmission – that is, the protocol is asynchronous – and an explicit hybrid-ARQ process number is used to indicate which process is being addressed. In an asynchronous hybrid-ARQ protocol, the retransmissions are in principle scheduled similarly to the initial transmissions. Uplink retransmissions, on the other hand, are based on a synchronous protocol, the retransmission occurs at a predefined time after the initial transmission and the process number can be implicitly derived. In a synchronous protocol the time instant for the retransmissions is fixed once the initial transmission has been scheduled, which must be accounted for in the scheduling operation. However, note that the scheduler knows from the hybrid-ARQ entity in the eNodeB whether a terminal will perform a retransmission or not.

The hybrid-ARQ mechanism will rapidly correct transmission errors due to noise or unpredictable channel variations. As discussed above, the RLC is also capable of requesting retransmissions, which at first sight may seem unnecessary. However, the reason for having two retransmission mechanisms on top of each other can be seen in the feedback signaling – hybrid ARQ provides fast retransmissions but due to errors in the feedback the residual error rate is typically too high for, for example,

good TCP performance, while RLC ensures (almost) error-free data delivery but slower retransmissions than the hybrid-ARQ protocol. Hence, the combination of hybrid ARQ and RLC provides an attractive combination of small round-trip time and reliable data delivery. Furthermore, as the RLC and hybrid ARQ are located in the same node, tight interaction between the two is possible, as discussed in Chapter 12.

8.2.3 Physical Layer

The physical layer is responsible for coding, physical-layer hybrid-ARQ processing, modulation, multi-antenna processing, and mapping of the signal to the appropriate physical time–frequency resources. It also handles mapping of transport channels to physical channels, as shown in Figures 8.7 and 8.8.

As mentioned in the introduction, the physical layer provides services to the MAC layer in the form of transport channels. Data transmission in downlink and uplink use the DL-SCH and UL-SCH transport-channel types respectively. There is at most one or, in the case of spatial multiplexing, two transport blocks per TTI on a DL-SCH or UL-SCH. In the case of carrier aggregation, there is one DL-SCH (or UL-SCH) per component carrier.

A *physical channel* corresponds to the set of time–frequency resources used for transmission of a particular transport channel and each transport channel is mapped to a corresponding physical channel, as shown in Figures 8.7 and 8.8. In addition to the physical channels with a corresponding transport channel, there are also physical channels without a corresponding transport channel. These channels, known as L1/L2 control channels, are used for *downlink control information* (DCI), providing the terminal with the necessary information for proper reception and decoding of the downlink data transmission, and *uplink control information* (UCI) used for providing the scheduler and the hybrid-ARQ protocol with information about the situation at the terminal.

The physical-channel types defined in LTE include the following:

- The *Physical Downlink Shared Channel* (PDSCH) is the main physical channel used for unicast data transmission, but also for transmission of paging information.
- The *Physical Broadcast Channel* (PBCH) carries part of the system information, required by the terminal in order to access the network.
- The *Physical Multicast Channel* (PMCH) is used for MBSFN operation.
- The *Physical Downlink Control Channel* (PDCCH) is used for downlink control information, mainly scheduling decisions, required for reception of PDSCH, and for scheduling grants enabling transmission on the PUSCH.
- The *Physical Hybrid-ARQ Indicator Channel* (PHICH) carries the hybrid-ARQ acknowledgement to indicate to the terminal whether a transport block should be retransmitted or not.
- The *Physical Control Format Indicator Channel* (PCFICH) is a channel providing the terminals with information necessary to decode the set of PDCCHs. There is only one PCFICH per component carrier.
- The *Physical Uplink Shared Channel* (PUSCH) is the uplink counterpart to the PDSCH. There is at most one PUSCH per uplink component carrier per terminal.
- The *Physical Uplink Control Channel* (PUCCH) is used by the terminal to send hybrid-ARQ acknowledgements, indicating to the eNodeB whether the downlink transport block(s) was successfully received or not, to send channel-state reports aiding downlink channel-dependent scheduling, and for requesting resources to transmit uplink data upon. There is at most one PUCCH per terminal.

- The *Physical Random-Access Channel* (PRACH) is used for random access, as described in Chapter 14.

Note that some of the physical channels, more specifically the channels used for downlink control information (PCFICH, PDCCH, PHICH) and uplink control information (PUCCH), do not have a corresponding transport channel.

The remaining downlink transport channels are based on the same general physical-layer processing as the DL-SCH, although with some restrictions in the set of features used. This is especially true for PCH and MCH transport channels. For the broadcast of system information on the BCH, a terminal must be able to receive this information channel as one of the first steps prior to accessing the system. Consequently, the transmission format must be known to the terminals a priori and there is no dynamic control of any of the transmission parameters from the MAC layer in this case. The BCH is also mapped to the physical resource (the OFDM time–frequency grid) in a different way, as described in more detail in Chapter 14.

For transmission of paging messages on the PCH, dynamic adaptation of the transmission parameters can, to some extent, be used. In general, the processing in this case is similar to the generic DL-SCH processing. The MAC can control modulation, the amount of resources, and the antenna mapping. However, as an uplink has not yet been established when a terminal is paged, hybrid ARQ cannot be used as there is no possibility for the terminal to transmit a hybrid-ARQ acknowledgement.

The MCH is used for MBMS transmissions, typically with single-frequency network operation, as described in Chapter 3, by transmitting from multiple cells on the same resources with the same format at the same time. Hence, the scheduling of MCH transmissions must be coordinated between the cells involved and dynamic selection of transmission parameters by the MAC is not possible.

8.3 CONTROL-PLANE PROTOCOLS

The control-plane protocols are, among other things, responsible for connection setup, mobility, and security. Control messages transmitted from the network to the terminals can originate either from the MME, located in the core network, or from the *Radio Resource Control* (RRC), located in the eNodeB.

NAS control-plane functionality, handled by the MME, includes EPS bearer management, authentication, security, and different idle-mode procedures such as paging. It is also responsible for assigning an IP address to a terminal. For a detailed discussion about the NAS control-plane functionality, see [9].

The RRC is located in the eNodeB and is responsible for handling the RAN-related procedures, including:

- Broadcast of system information necessary for the terminal to be able to communicate with a cell. Acquisition of system information is described in Chapter 14.
- Transmission of paging messages originating from the MME to notify the terminal about incoming connection requests. Paging, discussed further in Chapter 14, is used in the RRC_IDLE state (described further below) when the terminal is not connected to a particular cell. Indication of system-information update is another use of the paging mechanism, as is public warning systems.

- Connection management, including setting up bearers and mobility within LTE. This includes establishing an RRC context – that is, configuring the parameters necessary for communication between the terminal and the radio-access network.
- Mobility functions such as cell (re)selection.
- Measurement configuration and reporting.
- Handling of UE capabilities; when connection is established the terminal will announce its capabilities as all terminals are not capable of supporting all the functionality described in the LTE specifications, as briefly discussed in Chapter 7.

RRC messages are transmitted to the terminal using *signaling radio bearers* (SRBs), using the same set of protocol layers (PDCP, RLC, MAC and PHY) as described in Section 8.2. The SRB is mapped to the common control channel (CCCH) during establishment of connection and, once a connection is established, to the dedicated control channel (DCCH). Control-plane and user-plane data can be multiplexed in the MAC layer and transmitted to the terminal in the same TTI. The aforementioned MAC control elements can also be used for control of radio resources in some specific cases where low latency is more important than ciphering, integrity protection, and reliable transfer.

8.3.1 State Machine

In LTE, a terminal can be in two different states, as illustrated in Figure 8.13, RRC_CONNECTED and RRC_IDLE.

In RRC_CONNECTED, there is an RRC context established – that is, the parameters necessary for communication between the terminal and the radio-access network are known to both entities. The cell to which the terminal belongs is known and an identity of the terminal, the *Cell Radio-Network Temporary Identifier* (C-RNTI), used for signaling purposes between the terminal and the network, has been configured. RRC_CONNECTED is intended for data transfer to/from the terminal, but *discontinuous reception* (DRX) can be configured in order to reduce terminal power consumption (DRX is described in further detail in Chapter 13). Since there is an RRC context established in the eNodeB in RRC_CONNECTED, leaving DRX and starting to receive/transmit data is relatively fast as no connection with its associated signaling is needed.

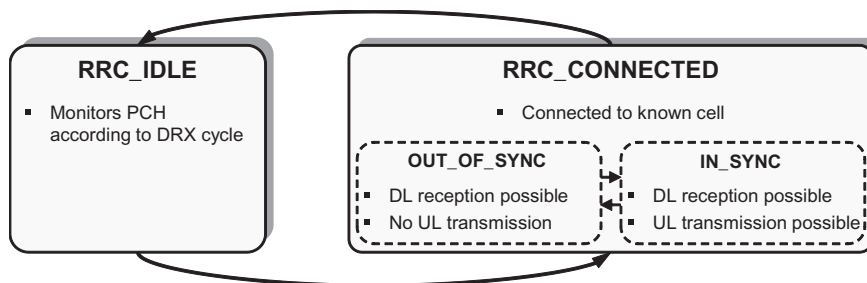


FIGURE 8.13

LTE states.

Although expressed differently in the specifications, RRC_CONNECTED can be thought of as having two substates, IN_SYNC and OUT_OF_SYNC, depending on whether the uplink is synchronized to the network or not. Since LTE uses an orthogonal FDMA/TDMA-based uplink, it is necessary to synchronize the uplink transmission from different terminals such that they arrive at the receiver at (approximately) the same time. The procedure for obtaining and maintaining uplink synchronization is described in Chapter 11, but in short the receiver measures the arrival time of the transmissions from each actively transmitting terminal and sends timing-correction commands in the downlink. As long as the uplink is synchronized, uplink transmission of user data and L1/L2 control signaling is possible. If no uplink transmission has taken place within a configurable time window, timing alignment is obviously not possible and the uplink is declared to be non-synchronized. In this case, the terminal needs to perform a random-access procedure to restore uplink synchronization prior to transmission of uplink data or control information.

In RRC_IDLE, there is no RRC context in the radio-access network and the terminal does not belong to a specific cell. No data transfer may take place as the terminal sleeps most of the time in order to reduce battery consumption. Uplink synchronization is not maintained and hence the only uplink transmission activity that may take place is random access, discussed in Chapter 14, to move to RRC_CONNECTED. When moving to RRC_CONNECTED the RRC context needs to be established in both the radio-access network and the terminal. Compared to leaving DRX this takes a somewhat longer time. In the downlink, terminals in RRC_IDLE periodically wake up in order to receive paging messages, if any, from the network, as described in Chapter 14.

Physical Transmission Resources

In Chapter 8, the overall LTE architecture was discussed, including an overview of the different protocol layers. Prior to discussing the detailed LTE downlink and uplink transmission schemes, a description of the basic time- and frequency-domain structure of LTE will be provided in this chapter.

9.1 OVERALL TIME–FREQUENCY STRUCTURE

OFDM is the basic transmission scheme for both the downlink and uplink transmission directions in LTE although, for the uplink, specific means are taken to reduce the cubic metric of the transmitted signal, thereby allowing for improved efficiency for the terminal transmitter power amplifier. Thus, for data transmission, DFT precoding is applied before OFDM modulation, leading to *DFT-spread OFDM* or *DFTS-OFDM*, as described in Chapter 4. It should be noted though that DFTS-OFDM is only applied to uplink *data* transmission – that is, for the transmission of the UL-SCH transport channel. As will be seen in Chapter 11, for other uplink transmissions, such as the transmission of *L1/L2 control signaling* and different types of *reference-signal transmissions*, other means are taken to limit the cubic metric of the transmitted signal.

The LTE OFDM subcarrier spacing equals 15 kHz for both downlink and uplink. As discussed in Chapter 3, the selection of the subcarrier spacing in an OFDM-based system needs to carefully balance overhead from the cyclic prefix against sensitivity to Doppler spread/shift and other types of frequency errors and inaccuracies. The choice of 15 kHz for the LTE subcarrier spacing was found to offer a good balance between these two constraints.

Assuming an FFT-based transmitter/receiver implementation, 15 kHz subcarrier spacing corresponds to a sampling rate $f_s = 15000 \cdot N_{\text{FFT}}$, where N_{FFT} is the FFT size. It is important to understand though that the LTE specifications do not in any way mandate the use of FFT-based transmitter/receiver implementations and even less so a particular FFT size or sampling rate. Nevertheless, FFT-based implementations of OFDM are common practice and an FFT size of 2048, with a corresponding sampling rate of 30.72 MHz, is suitable for the wider LTE carrier bandwidths, such as bandwidths of the order of 15 MHz and above. However, for smaller carrier bandwidths, a smaller FFT size and a correspondingly lower sampling rate can very well be used. The sampling rate above illustrates another factor influencing the choice of the LTE subcarrier spacing, namely a desire to simplify implementation of dual-mode LTE/HSPA terminals. Assuming a power-of-two FFT size and a subcarrier spacing of 15 kHz, the sampling rate $\Delta f \cdot N_{\text{FFT}}$ will be a multiple or sub-multiple of the HSPA chip rate of 3.84 Mchip/s.

In addition to the 15 kHz subcarrier spacing, a *reduced subcarrier spacing* of 7.5 kHz, with a corresponding OFDM symbol time that is twice as long, is also defined for LTE. The introduction of the reduced subcarrier spacing specifically targeted MBSFN-based multicast/broadcast transmissions (see Chapter 15). However, currently the 7.5 kHz subcarrier numerology is only partly implemented in the LTE specifications. Thus, at least for LTE up to and including release 10, only the 15 kHz subcarrier spacing is fully supported. The remaining discussions within this and the following chapters will assume the 15 kHz subcarrier spacing unless explicitly stated otherwise.

In the time domain, LTE transmissions are organized into (*radio*) *frames* of length 10 ms, each of which is divided into ten equally sized *subframes* of length 1 ms, as illustrated in Figure 9.1. Each subframe consists of two equally sized *slots* of length $T_{\text{slot}} = 0.5$ ms, with each slot consisting of a number of OFDM symbols including cyclic prefix.¹

To provide consistent and exact timing definitions, different time intervals within the LTE specifications are defined as multiples of a basic time unit $T_s = 1/(15000 \cdot 2048)$. The basic time unit T_s can thus be seen as the sampling time of an FFT-based transmitter/receiver implementation with an FFT size equal to 2048. The time intervals outlined in Figure 9.1 can thus also be expressed as $T_{\text{frame}} = 307200 \cdot T_s$, $T_{\text{subframe}} = 30720 \cdot T_s$, and $T_{\text{slot}} = 15360 \cdot T_s$ for the frame, subframe, and slot durations respectively.

On a higher level, each frame is identified by a *System Frame Number* (SFN). The SFN is used to control different transmission cycles that may have a period longer than one frame, such as paging sleep-mode cycles and periods for channel-status reporting. The SFN period equals 1024, thus the SFN repeats itself after 1024 frames or roughly 10 seconds.

The 15 kHz LTE subcarrier spacing corresponds to a useful symbol time $T_u = 2048 \cdot T_s$ or approximately 66.7 μ s. The overall OFDM symbol time is then the sum of the useful symbol time and the cyclic-prefix length T_{CP} . As illustrated in Figure 9.1, LTE defines two cyclic-prefix lengths, the *normal* cyclic prefix and an *extended* cyclic prefix, corresponding to seven and six OFDM symbols per slot respectively. The exact cyclic-prefix lengths, expressed in the basic time unit T_s , are given in Figure 9.1. It can be noted that, in the case of the normal cyclic prefix, the cyclic-prefix length for the first OFDM symbol of a slot is somewhat larger compared to the remaining OFDM symbols. The reason for this is simply to fill the entire 0.5 ms slot, as the number of basic time units T_s per slot (15360) is not divisible by seven.

The reasons for defining two cyclic-prefix lengths for LTE are twofold:

- A longer cyclic prefix, although less efficient from a cyclic-prefix-overhead point of view, may be beneficial in specific environments with extensive delay spread, for example in very large cells. It is important to have in mind, though, that a longer cyclic prefix is not necessarily beneficial in the case of large cells, even if the delay spread is very extensive in such cases. If, in large cells, link performance is limited by noise rather than by signal corruption due to residual time dispersion not covered by the cyclic prefix, the additional robustness to radio-channel time dispersion, due to the use of a longer cyclic prefix, may not justify the corresponding additional energy overhead of a longer cyclic prefix.

¹This is valid for the “normal” downlink subframes. As described in Section 9.5.2, the *special subframe* present in the case of TDD operation is not divided into two slots but rather into three fields. However, the number of OFDM symbols in the special subframe is the same as for the normal subframes illustrated in Figure 9.1.

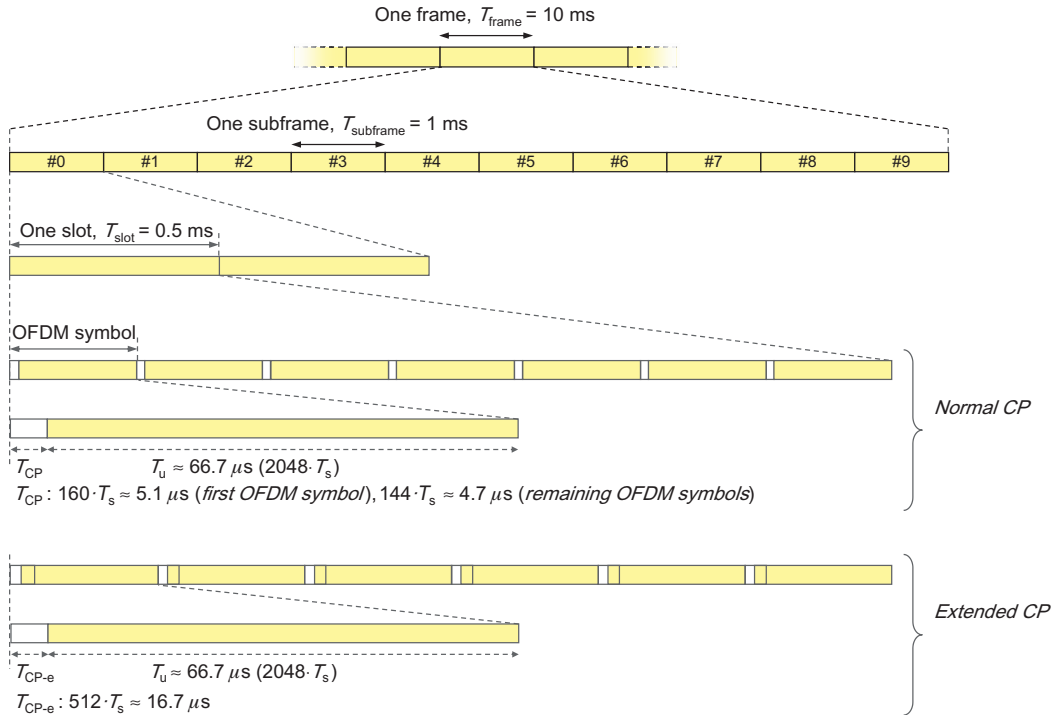


FIGURE 9.1

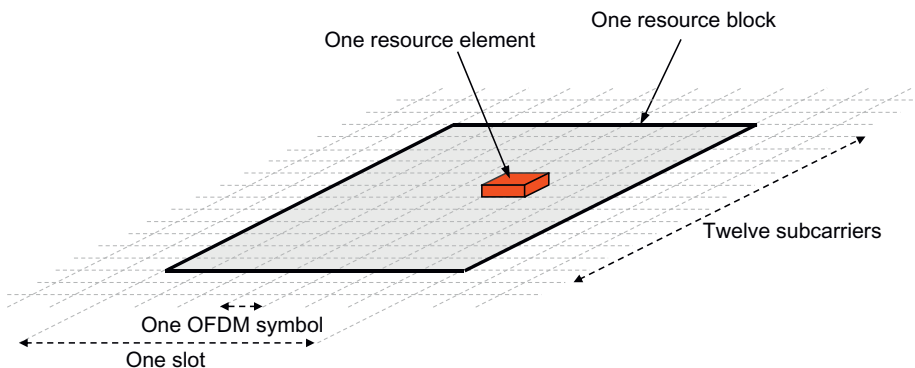
LTE time-domain structure.

- 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. In the case of MBSFN operation, the extended cyclic prefix is therefore used.

It should be noted that different cyclic-prefix lengths may be used for different subframes within a frame. As an example, to be discussed further in Chapter 15, MBSFN-based multicast/broadcast transmission is always confined to a limited set of subframes, in which case the use of the extended cyclic prefix, with its associated additional cyclic-prefix overhead, may only be applied to these subframes.²

A *resource element*, consisting of one subcarrier during one OFDM symbol, is the smallest physical resource in LTE. Furthermore, as illustrated in Figure 9.2, resource elements are grouped into *resource blocks*, where each resource block consists of 12 consecutive subcarriers in the frequency domain and one 0.5 ms slot in the time domain. Each resource block thus consists of $7 \cdot 12 = 84$

²The extended cyclic prefix is then actually applied only to the so-called MBSFN part of the MBSFN subframes (see Section 9.2).

**FIGURE 9.2**

The LTE physical time–frequency resource.

resource elements in the case of a normal cyclic prefix and $6 \cdot 12 = 72$ resource elements in the case of an extended cyclic prefix.

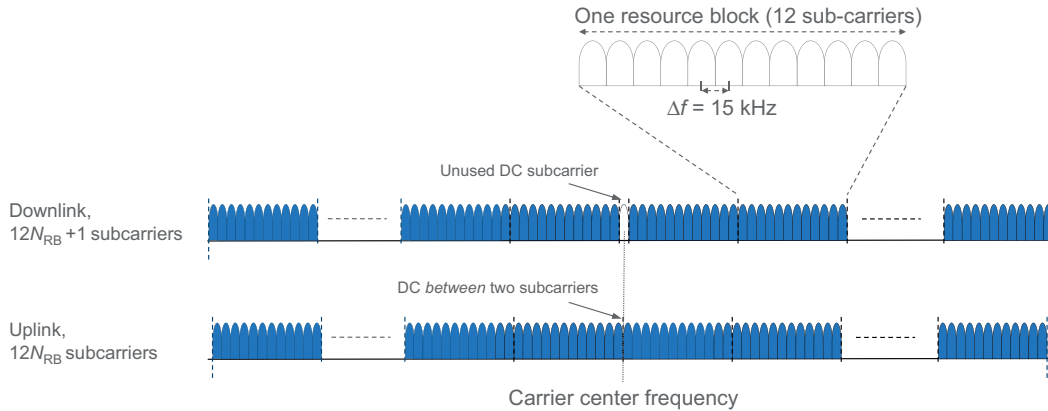
Although resource blocks are defined over one slot, the basic time-domain unit for dynamic scheduling in LTE is one subframe, consisting of two consecutive slots. The reason for defining the resource blocks over one slot is that *distributed downlink transmission*, as described in Chapter 10 and *uplink frequency hopping* (described in Chapter 11) are defined on a slot or resource-block basis. The minimum scheduling unit, consisting of two time-consecutive resource blocks within one subframe (one resource block per slot), can be referred to as a *resource-block pair*.

The LTE physical-layer specification allows for a carrier to consist of any number of resource blocks in the frequency domain, ranging from a minimum of six resource blocks up to a maximum of 110 resource blocks. This corresponds to an overall transmission bandwidth ranging from roughly 1 MHz up to in the order of 20 MHz with very fine granularity and thus allows for a very high degree of LTE bandwidth flexibility, at least from a physical-layer-specification point of view. However, as mentioned in Chapter 7, LTE radio-frequency requirements are, at least initially, only specified for a limited set of transmission bandwidths, corresponding to a limited set of possible values for the number of resource blocks within a carrier. Also note that, in LTE release 10, the total bandwidth of the transmitted signal can be as large as 100 MHz by aggregating multiple carriers, as mentioned in Chapter 7 and further described in Section 9.3.

The resource-block definition above applies to both the downlink and uplink transmission directions. However, there is a minor difference between the downlink and uplink in terms of where the carrier center frequency is located in relation to the subcarriers.

In the downlink (upper part of Figure 9.3), there is an unused *DC-subcarrier* that coincides with the carrier center frequency. The reason why the DC-subcarrier is not used for downlink transmission is that it may be subject to disproportionately high interference due, for example, to local-oscillator leakage.

On the other hand, in the uplink (lower part of Figure 9.3), no unused DC-subcarrier is defined and the center frequency of an uplink carrier is located *between* two uplink subcarriers. The presence of an unused DC-carrier in the center of the spectrum would have prevented the assignment of the entire cell bandwidth to a single terminal and still retain the assumption of mapping to consecutive

**FIGURE 9.3**

Frequency-domain structure for LTE.

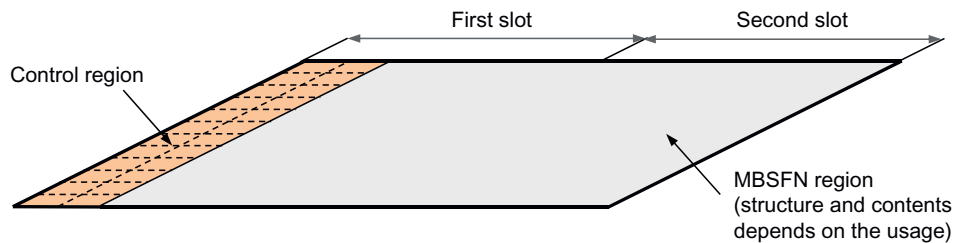
inputs of the OFDM modulator, something that is needed to retain the low-cubic-metric property of the DFTS-OFDM modulation used for uplink data transmission.

9.2 NORMAL SUBFRAMES AND MBSFN SUBFRAMES

In LTE, each downlink subframe (and the DwPTS in the case of TDD; see Section 9.5.2 for a discussion of the TDD frame structure) is normally divided into a *control region*, consisting of the first few OFDM symbols, and a *data region*, consisting of the remaining part of the subframe. The detailed usage of the resource elements in the two regions will be discussed in detail in Chapter 10; at this stage all we need to know is that the control region carries L1/L2 signaling necessary to control uplink and downlink data transmissions.

Additionally, already from the first release of LTE, so-called *MBSFN subframes* have also been defined. The original intention with MBSFN subframes was, as indicated by the name, to support MBSFN transmission, as described in Chapter 15. However, MBSFN subframes have also been found to be useful in other contexts, for example as part of relaying functionality, as discussed in Chapter 16. Hence, MBSFN subframes are therefore better seen as a generic tool and not related only to MBSFN transmission.

An MBSFN subframe, illustrated in Figure 9.4, consists of a control region of length one or two OFDM symbols, which is in essence identical to its counterpart in a normal subframe, followed by an *MBSFN region* whose contents depend on the usage of the MBSFN subframe. The reason for keeping the control region also in MBSFN subframes is, for example, to transmit control signaling necessary for uplink transmissions. All terminals, from LTE release 8 and onwards, are capable of receiving the control region of an MBSFN subframe. This is the reason why MBSFN subframes have been found useful as a generic tool to introduce, in a backwards-compatible way, new types of signaling and transmission not part of an earlier release of the LTE radio-access specification. Terminals not capable of receiving transmissions in the MBSFN region will simply ignore those transmissions.

**FIGURE 9.4**

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

Information about the set of subframes that are configured as MBSFN subframes in a cell is provided as part of the system information. In principle, an arbitrary pattern of MBSFN subframes can be configured with the pattern repeating after 40 ms.³ However, as information necessary to operate the system (to be more specific, synchronization signals, system information and paging, all of which will be discussed in detail in later chapters) needs to be transmitted in order for terminals to find and connect to a cell, subframes where such information is provided cannot be configured as MBSFN subframes. Therefore, subframes 0, 4, 5, and 9 for FDD and subframes 0, 1, 5, and 6 for TDD cannot be configured as MBSFN subframes, leaving the remaining six subframes as candidates for MBSFN subframes.

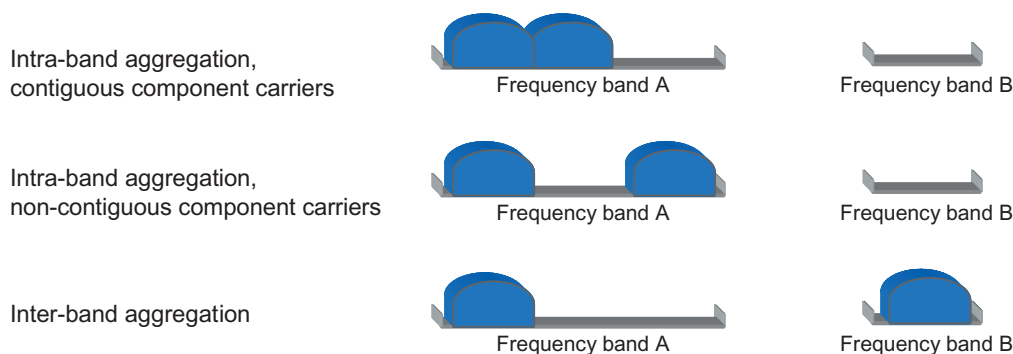
9.3 CARRIER AGGREGATION

The possibility for *carrier aggregation* was introduced in LTE release 10. In the case of carrier aggregation, multiple LTE carriers, each with a bandwidth up to 20 MHz, can be transmitted in parallel to/from the same terminal, thereby allowing for an overall wider bandwidth and correspondingly higher per-link data rates. In the context of carrier aggregation, each carrier is referred to as a *component carrier*⁴ as, from an RF point-of-view, the entire set of aggregated carriers can be seen as a single (RF) carrier.

Up to five component carriers, possibly of different bandwidths up to 20 MHz, can be aggregated allowing for overall transmission bandwidths up to 100 MHz. A terminal capable of carrier aggregation may receive or transmit simultaneously on multiple component carriers. Each component carrier can also be accessed by an LTE terminal from earlier releases – that is, component carriers are *backwards compatible*. Thus, in most respects and unless otherwise mentioned, the physical-layer description in the following chapters applies to each component carrier separately in the case of carrier aggregation.

³The reason for this value is that both the 10 ms frame length and the 8 ms hybrid-ARQ round-trip time are factors in the 40 ms, which is important for relaying operation (see also Chapter 16).

⁴In the specifications, the term *cell* is used instead of component carrier, but as the term cell is something of a misnomer in the uplink case, the term component carrier is used herein.

**FIGURE 9.5**

Different types of carrier aggregation.

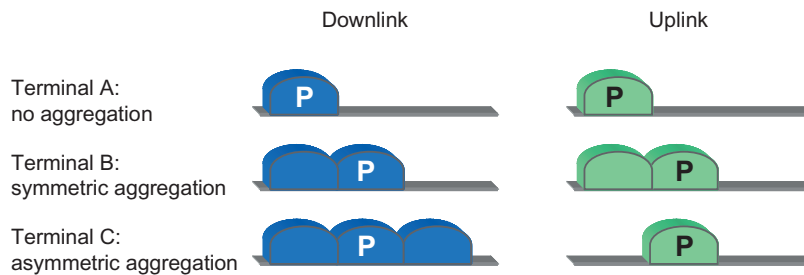
It should be noted that aggregated component carriers do not need to be contiguous in the frequency domain. Rather, with respect to the frequency location of the different component carriers, three different cases can be identified (see also Figure 9.5):

- Intra-band aggregation with frequency-contiguous component carriers
- Intra-band aggregation with non-contiguous component carriers
- Inter-band aggregation with non-contiguous component carriers.

The possibility to aggregate non-adjacent component carriers allows for exploitation of a fragmented spectrum; operators with a fragmented spectrum can provide high-data-rate services based on the availability of a wide overall bandwidth even though they do not possess a single wideband spectrum allocation. Except from an RF point of view there is no difference between the three different cases outlined in Figure 9.5 and they are all supported by the basic LTE release-10 specification. However, the complexity of RF implementation is vastly different, with the first case being the least complex. Thus, although spectrum aggregation is supported by the physical-layer and protocol specifications, the actual implementation will be strongly constrained, including specification of only a limited number of aggregation scenarios and aggregation over a dispersed spectrum only being supported by the most advanced terminals.

A terminal capable of carrier aggregation has one downlink *primary* component carrier and an associated uplink primary component carrier. In addition, it may have one or several *secondary* component carriers in each direction. Different terminals may have different carriers as their primary component carrier – that is, the configuration of the primary component carrier is terminal specific. The association between the downlink primary carrier and the corresponding uplink primary carrier is cell specific and signaled as part of the system information. This is similar to the case without carrier aggregation, although in the latter case the association is trivial. The reason for such an association is, for example, to determine to which uplink component carrier a certain scheduling grant transmitted on the downlink relates without having to explicitly signal the component-carrier number.

All idle mode procedures apply to the primary component carrier only or, expressed differently, carrier aggregation with additional secondary carriers configured only applies to terminals in the RRC_CONNECTED state. Upon connection to the network, the terminal performs the related procedures such

**FIGURE 9.6**

Examples of carrier aggregation ("P" denotes the primary component carrier).

as cell search and random access (see Chapter 14 for a detailed description of these procedures) following the same steps as in the absence of carrier aggregation. Once the communication between the network and the terminal is established, additional secondary component carriers can be configured.

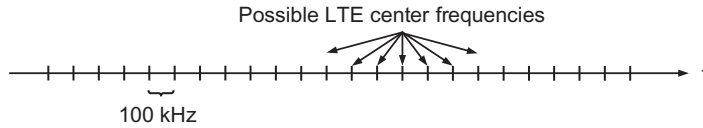
The fact that carrier aggregation is terminal specific – that is, different terminals may be configured to use different sets of component carriers – is useful not only from a network perspective to balance the load across component carriers, but also to handle different capabilities between terminals. Some terminals may be able to transmit/receive on multiple component carriers, while other terminals may do so on only a single carrier. This is an obvious consequence of being able to serve terminals from earlier releases at the same time as a carrier-aggregation-capable release-10 terminal, but it also allows for different capabilities in terms of carrier aggregation for different terminals as well as a differentiation between downlink and uplink carrier-aggregation capability. For example, a terminal may be capable of two component carriers in the downlink but of only a single component carrier – that is, no carrier aggregation – in the uplink, as is the case for terminal C in Figure 9.6. Note also that the primary component-carrier configuration can differ between terminals. Asymmetric carrier aggregation can also be useful to handle different spectrum allocations, for example if an operator has more spectrum available for downlink transmissions than uplink transmissions.

In release 10, only downlink-heavy asymmetries are supported – that is, the number of uplink component carriers configured for a terminal is always equal to or smaller than the number of configured downlink component carriers. Uplink-heavy asymmetries are less likely to be of practical interest and would also complicate the overall control signaling structure, as in such a case multiple uplink component carriers would need to be associated with the same downlink component carrier.

Carrier aggregation is supported for both FDD and TDD, although all component carriers need to have the same duplex scheme. Furthermore, in the case of TDD, the uplink–downlink configuration should be the same across component carriers. The special subframe configuration can be different for the different components carriers though, as long as the resulting downlink–uplink switch time is sufficiently large.

9.4 FREQUENCY-DOMAIN LOCATION OF LTE CARRIERS

In principle, an LTE carrier could be positioned anywhere within the spectrum and, actually, the basic LTE physical-layer specification does not say anything about the exact frequency location of an LTE

**FIGURE 9.7**

LTE carrier raster.

carrier, including the frequency band. However, in practice, there is obviously a need for restrictions on where an LTE carrier can be positioned in the frequency domain.

- In the end, an LTE terminal must be implemented and RF-wise such a terminal can obviously only support certain frequency bands. The frequency bands for which LTE is specified to operate are discussed in Chapter 17.
- After being activated, an LTE terminal has to search for a network-transmitted carrier within the frequency bands supported by the terminal. In order for that carrier search not to take an unreasonably long time, there is a need to limit the set of frequencies to be searched.

For this reason it is assumed that, within each supported frequency band, LTE carriers may exist on a 100kHz *carrier raster* or *carrier grid* – that is, the carrier center frequency can be expressed as $m \cdot 100\text{kHz}$, where m is an integer (see Figure 9.7).

In the case of carrier aggregation, multiple carriers can be transmitted to/from the same terminal. In order for the different component carriers to be accessible also by earlier-release terminals, each component carrier should fall on the 100kHz carrier grid. However, in the case of carrier aggregation, there is an additional constraint that the carrier spacing between adjacent component carriers should be a multiple of the 15kHz subcarrier spacing to allow transmission/reception with a single FFT.⁵ Thus, in the case of carrier aggregation, the carrier spacing between the different component carriers should be a multiple of 300kHz, the smallest carrier spacing being a multiple of both 100kHz (the raster grid) and 15kHz (the subcarrier spacing). A consequence of this is that there will always be a small number of unused subcarriers between two component carriers, even when they are located as close as possible to each other, as illustrated in Figure 9.8.

9.5 DUPLEX SCHEMES

Spectrum flexibility is one of the key features of LTE. In addition to the flexibility in transmission bandwidth, LTE also supports operation in both paired and unpaired spectrum by supporting both FDD- and TDD-based duplex operation with the time–frequency structures illustrated in Figure 9.9. Although the time-domain structure is, in most respects, the same for FDD and TDD, there are some differences, most notably the presence of a *special subframe* in the case of TDD. The special subframe is used to provide the necessary guard time for downlink–uplink switching, as discussed below.

⁵This is obviously only relevant for component carriers that are contiguous in the frequency domain. Furthermore, in the case of independent frequency errors between component carriers, separate FFTs may be needed at the receiver.

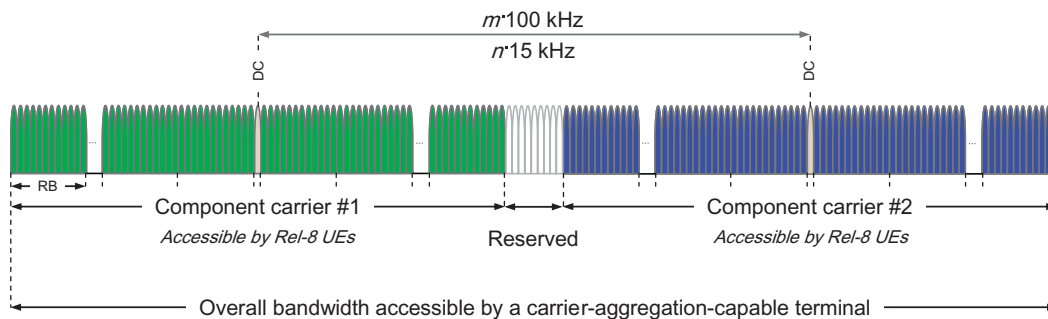


FIGURE 9.8

LTE carrier raster and carrier aggregation.

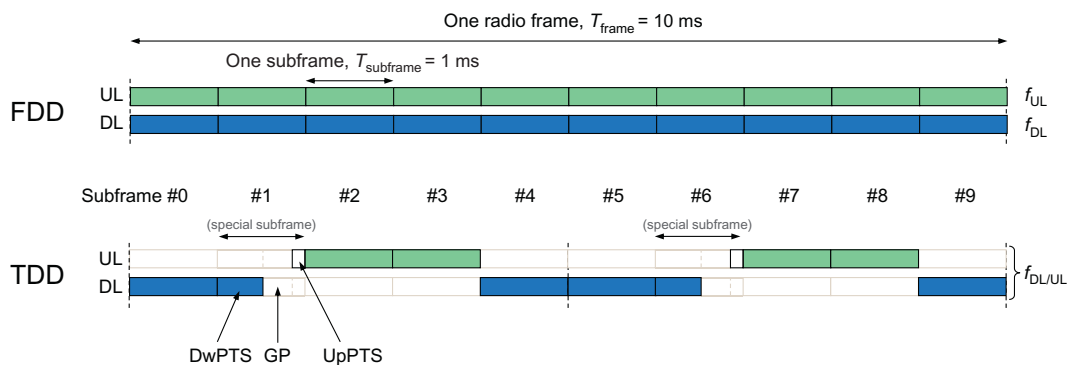


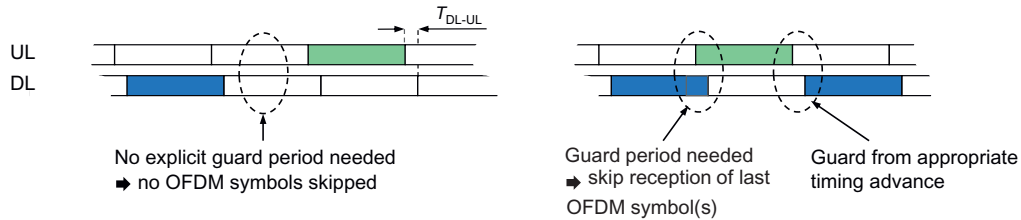
FIGURE 9.9

Uplink/downlink time-frequency structure for FDD and TDD.

9.5.1 Frequency-Division Duplex (FDD)

In the case of FDD operation (upper part of Figure 9.9), there are two carrier frequencies, one for uplink transmission (f_{UL}) and one for downlink transmission (f_{DL}). During each frame, there are thus ten uplink subframes and ten downlink subframes, and uplink and downlink transmission can occur simultaneously within a cell. Isolation between downlink and uplink transmissions is achieved by transmission/reception filters, known as duplex filters, and a sufficiently large *duplex separation* in the frequency domain.

Even if uplink and downlink transmission can occur simultaneously within a cell in the case of FDD operation, a terminal may be capable of *full-duplex* operation or only *half-duplex* operation for a certain frequency band, depending on whether or not it is capable of simultaneous transmission/reception. In the case of full-duplex capability, transmission and reception may also occur simultaneously at a terminal, whereas a terminal capable of only half-duplex operation cannot transmit and receive simultaneously. As mentioned in Chapter 7, supporting only half-duplex operation allows for simplified terminal implementation due to relaxed duplex-filter requirements. This applies especially

**FIGURE 9.10**

Guard time at the terminal for half-duplex FDD.

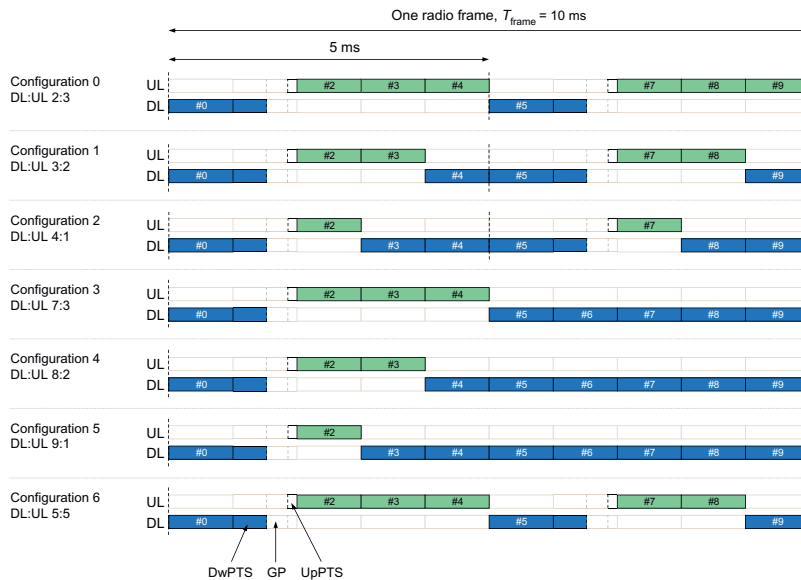
for certain frequency bands with a narrow duplex gap. Hence, full duplex support is *frequency-band dependent* such that a terminal may support only half-duplex operation in certain frequency bands while being capable of full-duplex operation in the remaining supported bands. It should be noted that full/half-duplex capability is a property of the *terminal*; the base station is operating in full duplex irrespective of the terminal capabilities. Hence, as the relevant transmission structures and timing relations are identical between full-duplex and half-duplex FDD, a single cell may simultaneously support a mixture of full-duplex and half-duplex FDD terminals. Half-duplex operation has an impact on the sustained data rates that can be provided to/from a single mobile terminal as it cannot transmit in all uplink subframes, but the cell capacity is hardly affected as typically it is possible to schedule different terminals in uplink and downlink in a given subframe.

Since a half-duplex terminal is not capable of simultaneous transmission and reception, the scheduling decisions must take this into account and half-duplex operation can be seen as a scheduling restriction, as will be discussed in more detail in Chapter 13. If a terminal is scheduled such that downlink reception in one subframe immediately precedes a subframe of uplink transmission, a guard time is necessary for the terminal to switch from reception to transmission. This is created in such cases by allowing the terminal to skip receiving⁶ the last OFDM symbol(s) in the downlink subframe, as illustrated in Figure 9.10.

9.5.2 Time-Division Duplex (TDD)

In the case of TDD operation (lower part of Figure 9.9), there is a single carrier frequency only and uplink and downlink transmissions are separated in the time domain on a cell basis. As seen in the figure, some subframes are allocated for uplink transmissions and some subframes for downlink transmission, with the switch between downlink and uplink occurring in the *special subframe* (subframe 1 and, in some cases, subframe 6). Different asymmetries in terms of the amount of resources – that is, subframes – allocated for uplink and downlink transmission respectively are provided through the seven different downlink/uplink configurations illustrated in Figure 9.11. As seen in the figure, subframes 0 and 5 are always allocated for downlink transmission while subframe 2 is always allocated for uplink transmissions. The remaining subframes (except the special subframe; see below)

⁶The impact on the decoding performance can be accounted for by the rate adaptation mechanism. For very high data rates, the performance impact may be somewhat larger due to the downlink interleaving structure, in which case it may be preferable not to schedule an uplink transmission immediately after a downlink reception.

**FIGURE 9.11**

Different downlink/uplink configurations in the case of TDD.

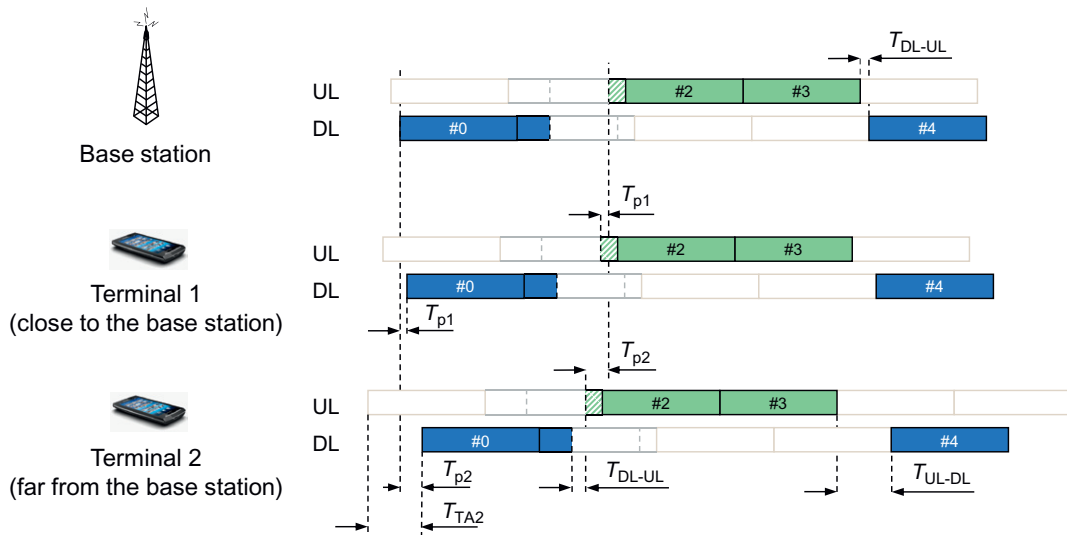
can then be flexibly allocated for downlink or uplink transmission depending on the configured downlink/uplink configuration. To avoid severe interference between downlink and uplink transmissions in different cells, neighboring cells typically have the same downlink/uplink configuration. This makes it difficult to change the downlink/uplink configuration dynamically, for example on a frame-by-frame basis. Therefore, the current LTE specification assumes that the downlink/uplink configuration is relatively static. It can be changed on a very slow basis though, in order to adapt to changing traffic patterns. It could also, in principle, be different in different areas, for example to match different traffic patterns, although inter-cell interference needs to be carefully addressed in this case.

As the same carrier frequency is used for uplink and downlink transmission, both the base station and the terminal need to switch from transmission to reception and vice versa. The switch between downlink and uplink occurs in the special subframe, which is split into three parts: a downlink part (DwPTS), a guard period (GP), and an uplink part (UpPTS).

The DwPTS is in essence treated as a normal downlink subframe,⁷ although the amount of data that can be transmitted is smaller due to the reduced length of the DwPTS compared to a normal subframe. The UpPTS, however, is not used for data transmission due to the very short duration. Instead, it can be used for channel sounding or random access. It can also be left empty, in which case it serves as extra guard period.

An essential aspect of any TDD system is the possibility to provide a sufficiently large *guard period* (or guard time), where neither downlink nor uplink transmissions occur. This guard period is necessary for switching from downlink to uplink transmission and vice versa and, as already mentioned, is

⁷For the shortest DwPTS duration of three OFDM symbols, DwPTS cannot be used for PDSCH transmission.

**FIGURE 9.12**

Timing relation for TDD operation.

obtained from the special subframe. The required length of the guard period depends on several factors. First, it should be sufficiently large to provide the necessary time for the circuitry in base stations and the terminals to switch from downlink to uplink. Switching is typically relatively fast, of the order of 20 microseconds, and in most deployments does not significantly contribute to the required guard time.

Secondly, the guard time should also ensure that uplink and downlink transmissions do not interfere at the base station. This is handled by advancing the uplink timing at the terminals such that, at the base station, the last uplink subframe before the uplink-to-downlink switch ends before the start of the first downlink subframe. The uplink timing of each terminal can be controlled by the base station by using the timing advance mechanism, as will be elaborated upon in Chapter 11. Obviously, the guard period must be large enough to allow the terminal to receive the downlink transmission and switch to transmission before it starts the (timing-advanced) uplink transmission. In essence, some of the guard period of the special subframe is “moved” from the downlink-to-uplink switch to the uplink-to-downlink switch by the timing-advance mechanism. This is illustrated in Figure 9.12. As the timing advance is proportional to the distance to the base station, a larger guard period is required when operating in large cells compared to small cells.

Finally, the selection of the guard period also needs to take interference between base stations into account. In a multi-cell network, inter-cell interference from downlink transmissions in neighboring cells must decay to a sufficiently low level before the base station can start to receive uplink transmissions. Hence, a larger guard period than is motivated by the cell size itself may be required as the last part of the downlink transmissions from distant base stations otherwise may interfere with uplink reception. The amount of guard period depends on the propagation environments, but in some cases the inter-base-station interference is a non-negligible factor when determining the guard period.

Table 9.1 Resulting Guard Period in OFDM Symbols for Different DwPTS and UpPTS Lengths (Normal Cyclic Prefix)									
	DwPTS								
	12	11		10		9		3	
GP	1	1	2	2	3	3	4	9	10
UpPTS	1	2	1	2	1	2	1	2	1

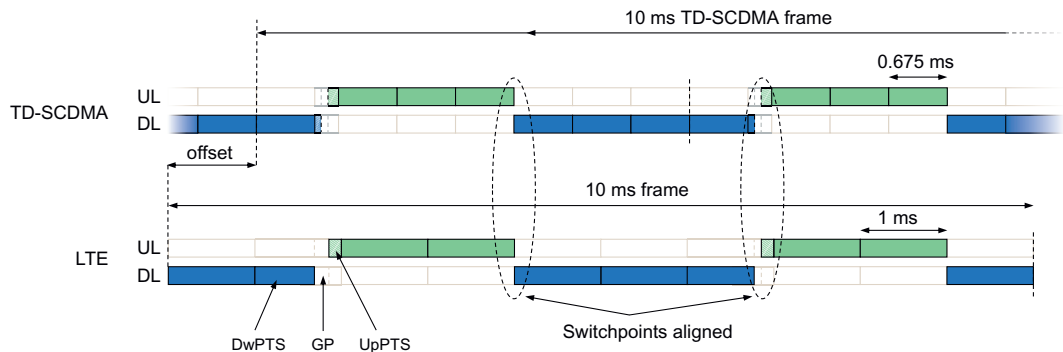


FIGURE 9.13

Coexistence between TD-SCDMA and LTE.

From the above discussion, it is clear that a sufficient amount of configurability of the guard period is needed to meet different deployment scenarios. Therefore, a set of DwPTS/GP/UpPTS configurations is supported as shown in [Table 9.1](#), where each configuration corresponds to a given length of the three fields in the special subframes. The DwPTS/GP/UpPTS configuration used in the cell is signaled as part of the system information.

9.5.3 LTE and TD-SCDMA Coexistence

In addition to supporting a wide range of different guard periods, an important aspect in the design of TDD in LTE was to simplify coexistence with and migration from systems based on the 3GPP TD-SCDMA standard (see Chapter 19 for details on TD-SCDMA). Basically, to handle inter-system interference from two different but co-sited TDD systems operating close in frequency, it is necessary to align the switch-points between the two systems. Since LTE supports configurable lengths of the DwPTS field, the switch-points of LTE and TD-SCDMA (or any other TDD system) can be aligned, despite the different subframe lengths used in the two systems. Aligning the switch-points between TD-SCDMA and LTE is the technical reason for splitting the special subframe into the three fields DwPTS/GP/UpPTS instead of locating the switch-point at the subframe boundary. An example of LTE/TD-SCDMA coexistence is given in [Figure 9.13](#).

The set of possible lengths of DwPTS/GP/UpPTS is selected to support common coexistence deployments, as well as to provide a high degree of guard-period flexibility for the reasons discussed earlier. The UpPTS length is one or two OFDM symbols and the DwPTS length can vary from three⁸ to twelve OFDM symbols, resulting in guard periods ranging from one to ten OFDM symbols. The resulting guard period for the different DwPTS and UpPTS configurations supported is summarized in Table 9.1 for the case of normal cyclic prefix. As discussed earlier, the DwPTS can be used for downlink data transmission, while the UpPTS can be used for sounding or random access only, due to its short duration.

⁸The smallest DwPTS length is motivated by the location of the primary synchronization signal in the DwPTS (see Chapter 14).

Downlink Physical-Layer Processing

10

In Chapter 8, the LTE radio-interface architecture was discussed with an overview of the functions and characteristics of the different protocol layers. Chapter 9 then gave an overview of the basic time–frequency structure of LTE transmissions, including the structure of the OFDM time–frequency grid being the fundamental physical resource on both uplink and downlink.

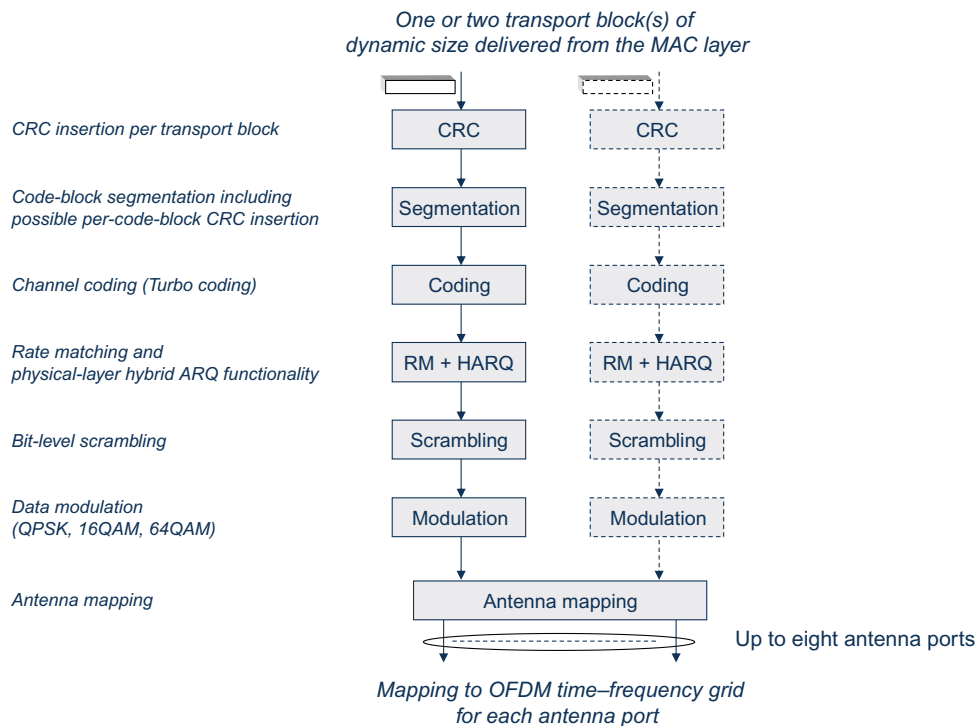
This chapter will provide a more detailed description of the downlink physical-layer functionality, including the transport-channel processing (Section 10.1), downlink reference signals (Section 10.2), details on downlink multi-antenna transmission (Section 10.3), and downlink L1/L2 control signaling (Section 10.4). Chapter 11 will provide a corresponding description for the *uplink* transmission direction. The later chapters will then go further into the details of some specific uplink and downlink functions and procedures.

10.1 TRANSPORT-CHANNEL PROCESSING

As described in Chapter 8, transport channels provide the interface between the MAC layer and the physical layer. As also described, for the LTE downlink there are four different types of transport channels defined, the Downlink Shared Channel (DL-SCH), the Multicast Channel (MCH), the Paging Channel (PCH), and the Broadcast Channel (BCH). This section provides a description of the physical-layer processing applied to DL-SCH transport channels, including the mapping to the physical resource – that is, to the resource elements of the OFDM time–frequency grid. DL-SCH is the main downlink transport-channel type in LTE and is used for transmission of user-specific higher-layer information, both user data and dedicated control information, as well as part of the downlink system information. The DL-SCH physical-layer processing is to a large extent applicable also to MCH and PCH transport channels, although with some additional constraints. On the other hand, the physical-layer processing, and the transmission structure in general, for the BCH is quite different. BCH transmission is described in Chapter 14 as part of the discussion on LTE system information.

10.1.1 Processing Steps

The different steps of the DL-SCH physical layer processing are outlined in [Figure 10.1](#). In the case of carrier aggregation – that is, transmission on multiple component carriers in parallel to the same terminal – the transmissions on the different carriers correspond to separate transport channels with separate and more or less independent physical-layer processing. The transport-channel processing outlined in [Figure 10.1](#) and the discussion below is thus valid also in the case of carrier aggregation.

**FIGURE 10.1**

Physical-layer processing for DL-SCH.

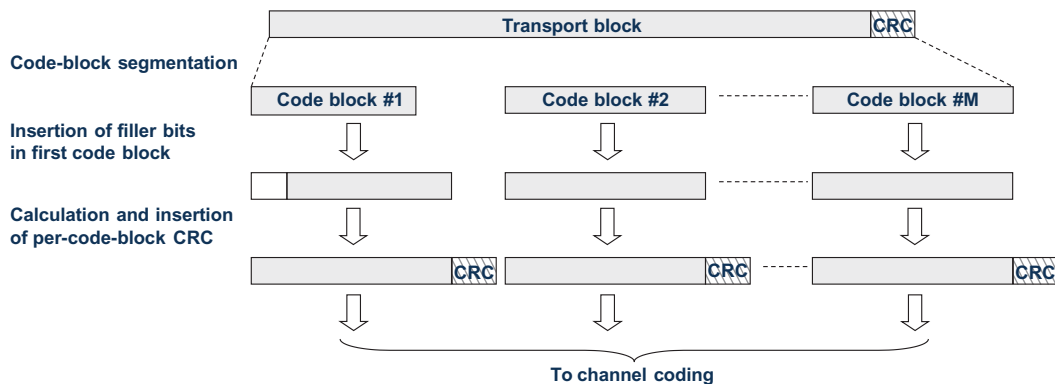
Within each *Transmission Time Interval* (TTI), corresponding to one subframe of length 1 ms, up to two transport blocks of dynamic size are delivered to the physical layer and transmitted over the radio interface for each component carrier. The number of transport blocks transmitted within a TTI depends on the configuration of the multi-antenna transmission scheme, as described in Section 10.3.

- In the case of no *spatial multiplexing* there is at most a single transport block in a TTI.
- In the case of spatial multiplexing, with transmission on multiple *layers* in parallel to the same terminal, there are two transport blocks within a TTI.¹

10.1.1.1 CRC Insertion Per Transport Block

In the first step of the physical-layer processing, a 24-bit CRC is calculated for and appended to each transport block. The CRC allows for receiver-side detection of errors in the decoded transport block. The corresponding error indication can, for example, be used by the downlink hybrid-ARQ protocol as a trigger for requesting retransmissions.

¹This is true for initial transmissions. In the case of hybrid-ARQ retransmissions there may also be cases when a single transport block is transmitted over multiple layers, as discussed, for example, in Section 10.3.2.

**FIGURE 10.2**

Code-block segmentation and per-code-block CRC insertion.

10.1.1.2 Code-Block Segmentation and Per-Code-Block CRC Insertion

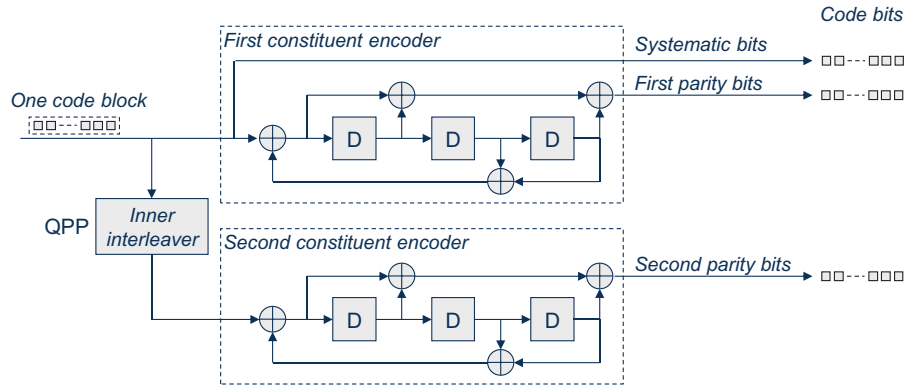
The LTE Turbo-coder internal interleaver is only defined for a limited number of code-block sizes, with a maximum block size of 6144 bits. If the transport block, including the transport-block CRC, exceeds this maximum code-block size, *code-block segmentation*, illustrated in Figure 10.2, is applied before the Turbo coding. Code-block segmentation implies that the transport block is segmented into smaller *code blocks*, the sizes of which should match the set of code-block sizes supported by the Turbo coder.

In order to ensure that a transport block of arbitrary size can be segmented into code blocks that match the set of available code-block sizes, the specification includes the possibility to insert “dummy” *filler bits* at the head of the first code block. However, the set of transport-block sizes currently defined for LTE has been selected so that filler bits are not needed.

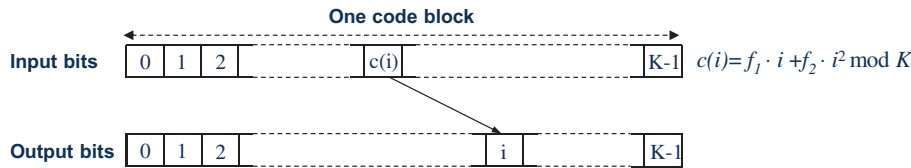
As can be seen in Figure 10.2, code-block segmentation also implies that an additional CRC (also of length 24 bits, but different compared to the transport-block CRC described above) is calculated for and appended to each code block. Having a CRC per code block allows for early detection of correctly decoded code blocks and correspondingly early termination of the iterative decoding of that code block. This can be used to reduce the terminal processing effort and corresponding energy consumption. In the case of a single code block no additional code-block CRC is applied.

One could argue that, in case of code-block segmentation, the transport-block CRC is redundant and implies unnecessary overhead as the set of code-block CRCs should indirectly provide information about the correctness of the complete transport block. However, code-block segmentation is only applied to large transport blocks for which the relative extra overhead due to the additional transport-block CRC is small. The transport-block CRC also adds additional error-detection capabilities and thus further reduces the risk for undetected errors in the decoded transport block.

Information about the transport-block size is provided to the terminal as part of the scheduling assignment transmitted on the PDCCH control channel, as described in Section 10.4.4. Based on this information, the terminal can determine the code-block size and number of code blocks. The terminal receiver can thus, based on the information provided in the scheduling assignment, straightforwardly undo the code-block segmentation and recover the decoded transport blocks.

**FIGURE 10.3**

LTE Turbo encoder.

**FIGURE 10.4**

Principles of QPP-based interleaving.

10.1.1.3 Channel Coding

Channel coding for DL-SCH (as well as for PCH and MCH) is based on Turbo coding [53], with encoding according to Figure 10.3. The encoding consists of two rate-1/2, eight-state constituent encoders, implying an overall code rate of 1/3, in combination with QPP-based² interleaving [69]. As illustrated in Figure 10.4, the QPP interleaver provides a mapping from the input (non-interleaved) bits to the output (interleaved) bits according to the function:

$$c(i) = f_1 \cdot i + f_2 \cdot i^2 \bmod K,$$

where i is the index of the bit at the output of the interleaver, $c(i)$ is the index of the same bit at the input of the interleaver, and K is the code-block/interleaver size. The values of the parameters f_1 and f_2 depend on the code-block size K . The LTE specification lists all supported code-block sizes, ranging from a minimum of 40 bits to a maximum of 6144 bits, together with the associated values for the parameters f_1 and f_2 . Thus, once the code-block size is known, the Turbo-coder inner interleaving, as well as the corresponding de-interleaving at the receiver side, can straightforwardly be carried out.

A QPP-based interleaver is *maximum contention free* [70], implying that the decoding can be parallelized without the risk for contention when the different parallel processes are accessing the

²QPP = Quadrature Permutation Polynomial.

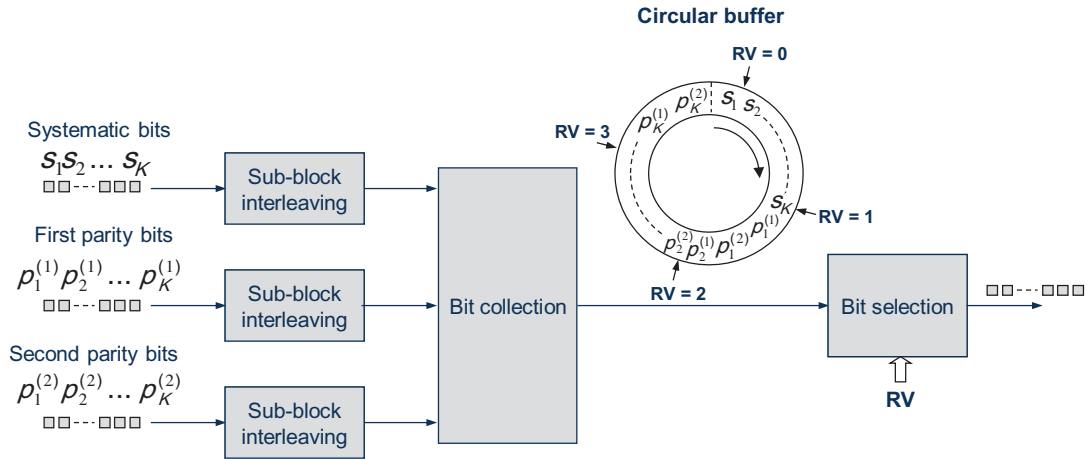


FIGURE 10.5

Rate matching and hybrid-ARQ functionality.

interleaver memory. For the very high data rates supported by LTE, the improved possibilities for parallel processing offered by QPP-based interleaving can substantially simplify the Turbo-encoder/decoder implementation.

10.1.1.4 Rate Matching and Physical-Layer Hybrid-ARQ Functionality

The task of the rate-matching and physical-layer hybrid-ARQ functionality is to extract, from the blocks of code bits delivered by the channel encoder, the exact set of code bits to be transmitted within a given TTI/subframe.

As illustrated in Figure 10.5, the outputs of the Turbo encoder (systematic bits, first parity bits, and second parity bits) are first separately interleaved. The interleaved bits are then inserted into what can be described as a circular buffer with the systematic bits inserted first, followed by alternating insertion of the first and second parity bits.

The bit selection then extracts consecutive bits from the circular buffer to an extent that matches the number of available resource elements in the resource blocks assigned for the transmission. The exact set of bits to extract depends on the *redundancy version* (RV) corresponding to different starting points for the extraction of coded bits from the circular buffer. As can be seen, there are four different alternatives for the redundancy version. The transmitter/scheduler selects the redundancy version and provides information about the selection as part of the scheduling assignment (see Section 10.4.4).

Note that the rate-matching and hybrid-ARQ functionality operates on the full set of code bits corresponding to one transport block and not separately on the code bits corresponding to a single code block.

10.1.1.5 Bit-Level Scrambling

LTE downlink scrambling implies that the block of code bits delivered by the hybrid-ARQ functionality is multiplied (*exclusive-or* operation) by a bit-level *scrambling sequence*. Without downlink scrambling,

the channel decoder at the terminal could, at least in principle, be equally matched to an interfering signal as to the target signal, thus being unable to properly suppress the interference. By applying different scrambling sequences for neighboring cells, the interfering signal(s) after descrambling is (are) randomized, ensuring full utilization of the processing gain provided by the channel code. Thus, the bit scrambling essentially serves the same purpose as the scrambling applied at chip level after the direct-sequence spreading in DS-CDMA-based systems such as WCDMA/HSPA. Fundamentally, channel coding can be seen as “advanced” spreading providing processing gain similar to direct-sequence spreading but also additional coding gain.

In LTE, downlink scrambling is applied to all transport channels as well as to the downlink L1/L2 control signaling. For all downlink transport-channel types except MCH, as well as for the L1/L2 control signaling, the scrambling sequences differ between neighboring cells (*cell-specific scrambling*) to ensure interference randomization between the cells. This is achieved by having the scrambling sequences depend on the physical-layer cell identity (Chapter 14). In contrast, in the case of MBSFN-based transmission using MCH, the same scrambling should be applied to all cells taking part in the MBSFN transmission – that is, all cells within the so-called *MBSFN area* (see Chapter 15).

10.1.1.6 Data Modulation

The downlink data modulation transforms the block of scrambled bits to a corresponding block of complex modulation symbols. The set of modulation schemes supported for the LTE downlink includes QPSK, 16QAM, and 64QAM, corresponding to two, four, and six bits per modulation symbol respectively.

10.1.1.7 Antenna Mapping

The antenna mapping jointly processes the modulation symbols corresponding to the one or two transport blocks and maps the result to different *antenna ports*. The antenna mapping can be configured in different ways corresponding to different multi-antenna transmission schemes, including transmit diversity, beam-forming, and spatial multiplexing. As indicated in Figure 10.1, LTE supports transmission using up to eight antenna ports depending on the exact multi-antenna transmission scheme. More details about LTE downlink multi-antenna transmission are provided in Section 10.3.

Note that the antenna ports referred to above do not necessarily correspond to specific physical antennas. Rather, an antenna port is a more general concept introduced, for example, to allow for beam-forming using multiple physical antennas without the terminal being aware of the beam-forming carried out at the transmitter side.

At least for the downlink, an antenna port can be seen as corresponding to the transmission of a *reference signal* (Section 10.2). Any data transmission from the antenna port can then rely on that reference signal for channel estimation for coherent demodulation. Thus, if the same reference signal is transmitted from multiple physical antennas, these physical antennas correspond to a *single* antenna port. Similarly, if two different reference signals are transmitted from the same set of physical antennas, this corresponds to *two separate antenna ports*.

It should be noted that the LTE specification actually has a somewhat more general definition of an antenna port, essentially just saying that two received signals can be assumed to have experienced the same overall channel, including any joint processing at the transmitter side, if and only if they have been transmitted on the same antenna port.

10.1.1.8 Resource-Block Mapping

The resource-block mapping takes the symbols to be transmitted on each antenna port and maps them to the resource elements of the set of resource blocks assigned by the MAC scheduler for the transmission. As described in Chapter 9, each resource block consists of 84 resource elements (twelve subcarriers during seven OFDM symbols).³ However, some of the resource elements within a resource block will not be available for the transport-channel transmission as they are occupied by:

- different types of downlink reference signals, as described in Section 10.2;
- downlink L1/L2 control signaling (one, two, or three OFDM symbols at the head of each subframe), as described in Section 10.4⁴.

Furthermore, as will be described in Chapter 14, within some resource blocks, additional resource elements are reserved for the transmission of *synchronization signals* as well as for the transmission of the BCH transport channel.

In the TDD special subframe (Section 9.5.2), mapping is limited to the DwPTS.

10.1.2 Localized and Distributed Resource Mapping

As already discussed in Chapter 7, when deciding what set of resource blocks to use for transmission to a specific terminal, the network may take the downlink channel conditions in both the time and frequency domains into account. Such time/frequency-domain channel-dependent scheduling, taking channel variations – for example, due to frequency-selective fading – into account, may significantly improve system performance in terms of achievable data rates and overall cell throughput.

However, in some cases downlink channel-dependent scheduling is not suitable to use or is not practically possible:

- For low-rate services such as voice, the feedback signaling associated with channel-dependent scheduling may lead to extensive relative overhead.
- At high mobility (high terminal speed), it may be difficult or even practically impossible to track the instantaneous channel conditions to the accuracy required for channel-dependent scheduling to be efficient.

In such situations, an alternative means to handle radio-channel frequency selectivity is to achieve frequency diversity by distributing a downlink transmission in the frequency domain.

One way to distribute a downlink transmission in the frequency domain, and thereby achieve frequency diversity, is to assign multiple non-frequency-contiguous resource blocks for the transmission to a terminal. LTE allows for such *distributed resource-block allocation* by means of *resource allocation types 0 and 1* (see Section 10.4.4). However, although sufficient in many cases, distributed resource-block allocation by means of these resource-allocation types has certain drawbacks:

- For both types of resource allocations, the minimum size of the allocated resource can be as large as four resource-block pairs and may thus not be suitable when resource allocations of smaller sizes are needed.
- In general, both these resource-allocation methods are associated with a relatively large PDCCH payload.

³72 resource elements in the case of extended cyclic prefix.

⁴In MBSFN subframes, the control region is limited to a maximum of two OFDM symbols.

In contrast, *resource-allocation type 2* (Section 10.4.4) always allows for the allocation of a single resource-block pair and is also associated with a relatively small PDCCH payload size. However, resource allocation type 2 only allows for the allocation of resource blocks that are contiguous in the frequency domain. In addition, regardless of the type of resource allocation, frequency diversity by means of distributed resource-block allocation will only be achieved in the case of resource allocations larger than one resource-block pair.

In order to provide the possibility for distributed resource-block allocation in the case of resource-allocation type 2, as well as to allow for distributing the transmission of a single resource-block pair in the frequency domain, the notion of a *Virtual Resource Block* (VRB) has been introduced for LTE.

What is being provided in the resource allocation is the resource allocation in terms of VRB pairs. The key to distributed transmission then lies in the mapping from VRB pairs to *Physical Resource Block* (PRB) pairs – that is, to the actual physical resource used for transmission.

The LTE specification defines two types of VRBs: *localized* VRBs and *distributed* VRBs. In the case of localized VRBs, there is a direct mapping from VRB pairs to PRB pairs, as illustrated in Figure 10.6. However, in the case of distributed VRBs, the mapping from VRB pairs to PRB pairs is more elaborate in the sense that:

- consecutive VRBs are not mapped to PRBs that are consecutive in the frequency domain; and
- even a single VRB pair is distributed in the frequency domain.

The basic principle of distributed transmission is outlined in Figure 10.7 and consists of two steps:

- A mapping from VRB pairs to PRB pairs such that consecutive VRB pairs are not mapped to frequency-consecutive PRB pairs (first step of Figure 10.7). This provides frequency diversity between consecutive VRB pairs. The spreading in the frequency domain is done by means of a block-based “interleaver” operating on resource-block pairs.

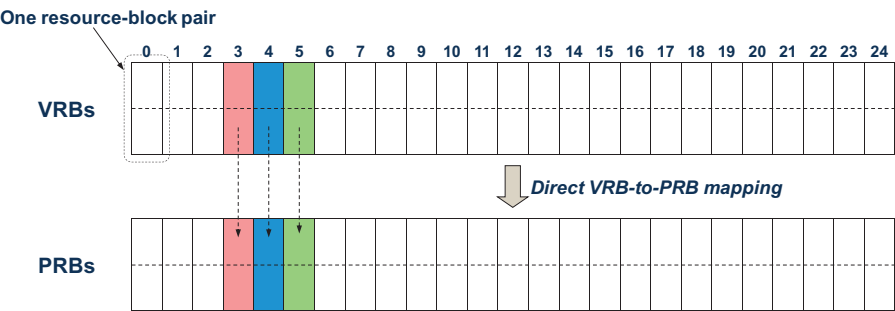


FIGURE 10.6

VRB-to-PRB mapping in the case of localized VRBs. A cell bandwidth corresponding to 25 resource blocks is assumed.

- A split of each resource-block pair such that the two resource blocks of the resource-block pair are transmitted with a certain frequency gap in between (second step of Figure 10.7). This also provides frequency diversity for a single VRB pair. This step can be seen as the introduction of frequency hopping on a slot basis.

Whether the VRBs are localized (and thus mapped according to Figure 10.6) or distributed (mapped according to Figure 10.7) is indicated on the associated PDCCH in the case of type 2 resource allocation. Thus, it is possible to dynamically switch between distributed and localized transmission and also mix distributed and localized transmission for different terminals within the same subframe.

The exact size of the frequency gap in Figure 10.7 depends on the overall downlink cell bandwidth according to Table 10.1. These gaps have been chosen based on two criteria:

1. The gap should be of the order of half the downlink cell bandwidth in order to provide good frequency diversity also in the case of a single VRB pair.
2. The gap should be a multiple of P^2 , where P is the size of a *resource-block group* as defined in Section 10.4.4 and used for resource allocation types 0 and 1. The reason for this constraint is to ensure a smooth coexistence in the same subframe between distributed transmission as described above and transmissions based on downlink allocation types 0 and 1.

Due to the constraint that the gap size should be a multiple of P^2 , the gap size will in most cases deviate from exactly half the cell bandwidth. In these cases, not all resource blocks within the cell bandwidth can be used for distributed transmission. As an example, for a cell bandwidth corresponding to 25 resource blocks (the example in Figure 10.7) and a corresponding gap size equal to 12 according to Table 10.1, the 25th resource-block pair cannot be used for distributed transmission.

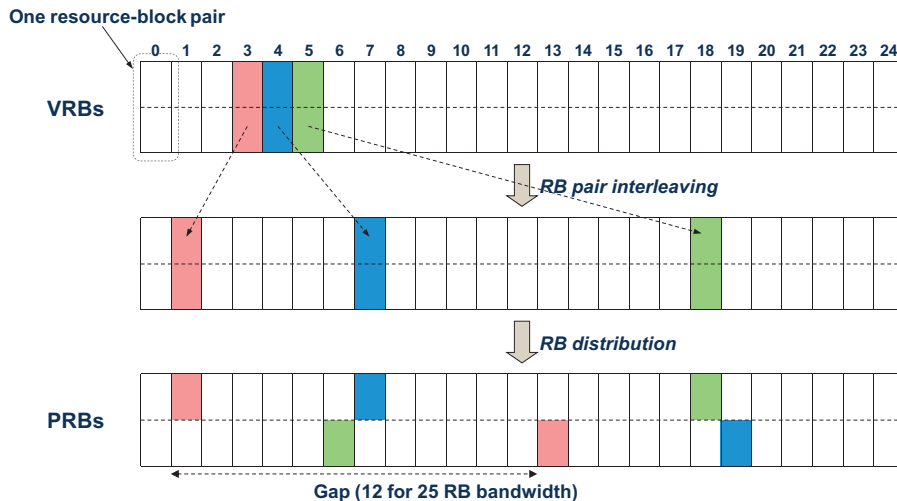


FIGURE 10.7

VRB-to-PRB mapping in the case of distributed VRBs. A cell bandwidth corresponding to 25 resource blocks is assumed.

Table 10.1 Gap Size for Different Cell Bandwidths (Number of Resource Blocks)										
	Bandwidth									
	6	7–8	9–10	11	12–19	20–26	27–44	45–63	64–79	80–110
<i>P</i>	1	1	1	2	2	2	3	3	4	4
Gap size	3	4	5	4	8	12	18	27	32	48

Table 10.2 Second Gap Size for Different Cell Bandwidths (Only Applicable for Cell Bandwidths of 50 RBs and Beyond)		
	Bandwidth	
	50–63	64–110
Gap size	9	16

As another example, for a cell bandwidth corresponding to 50 resource blocks (gap size equal to 27 according to [Table 10.1](#)), only 46 resource blocks would be available for distributed transmission.

In addition to the gap size outlined in [Table 10.1](#), for wider cell bandwidths (50 RBs and beyond), there is a possibility to use a second, smaller frequency gap with a size of the order of one-fourth of the cell bandwidth (see [Table 10.2](#)). The use of the smaller gap enables restriction of the distributed transmission to only a part of the overall cell bandwidth. Selection between the larger gap according to [Table 10.1](#) and the smaller gap according to [Table 10.2](#) is indicated by an additional bit in the resource allocation on PDCCH.

10.2 DOWNLINK REFERENCE SIGNALS

Downlink reference signals are predefined signals occupying specific resource elements within the downlink time–frequency grid. The LTE specification includes several types of downlink reference signals that are transmitted in different ways and used for different purposes by the receiving terminal:

- *Cell-specific reference signals* (CRS) are transmitted in every downlink subframe and in every resource block in the frequency domain, thus covering the entire cell bandwidth. The cell-specific reference signals can be used by the terminal for channel estimation for coherent demodulation of any downlink physical channel except for PMCH and for PDSCH in the case of *transmission modes 7, 8, or 9*. As described in Section 10.3, these transmission modes correspond to so-called *non-codebook-based precoding*. The cell-specific reference signals can also be used by the terminal to acquire *channel-state information* (CSI). Finally, terminal measurements on cell-specific reference signals are used as the basis for cell-selection and handover decisions.

- *Demodulation reference signals* (DM-RS), also sometimes referred to as *UE-specific reference signals*, are specifically intended to be used by terminals for channel estimation for PDSCH in the case of transmission modes 7, 8, or 9. The label “*UE-specific*” relates to the fact that each demodulation reference signal is intended for channel estimation by a single terminal. That specific reference signal is then only transmitted within the resource blocks assigned for PDSCH transmission to that terminal.
- *CSI reference signals* (CSI-RS) are specifically intended to be used by terminals to acquire *channel-state information* (CSI) in the case when demodulation reference signals are used for channel estimation.⁵ CSI-RS have a significantly lower time/frequency density, thus implying less overhead, compared to the cell-specific reference signals.
- *MBSFN reference signals* are intended to be used for channel estimation for coherent demodulation in the case of MCH transmission using so-called *MBSFN* (see Chapter 15 for more details on MCH transmission).
- *Positioning reference signals* were introduced in LTE release 9 to enhance *LTE positioning functionality*, more specifically to support the use of terminal measurements on multiple LTE cells to estimate the geographical position of the terminal. The positioning reference symbols of a certain cell can be configured to correspond to empty resource elements in neighboring cells, thus enabling high-SIR conditions when receiving neighbor-cell positioning reference signals.

10.2.1 Cell-Specific Reference Signals

Cell-specific reference signals, introduced in the first release of LTE (release 8), are the most basic downlink reference signals in LTE. There can be one, two, or four cell-specific reference signals in a cell, defining one, two, or four corresponding antenna ports.

10.2.1.1 Structure of a Single Reference Signal

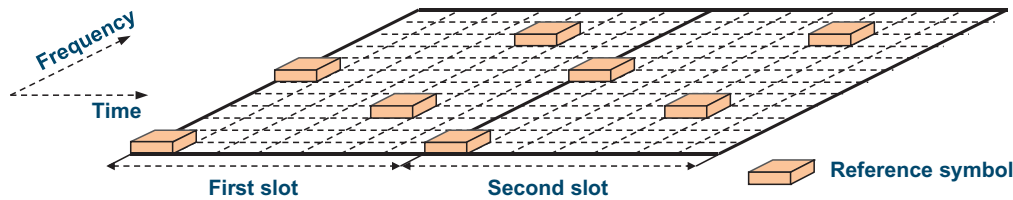
Figure 10.8 illustrates the structure of a single cell-specific reference signal. As can be seen, it consists of *reference symbols* of predefined values inserted within the first and third last⁶ OFDM symbol of each slot and with a frequency-domain spacing of six subcarriers. Furthermore, there is a frequency-domain staggering of three subcarriers for the reference symbols within the third last OFDM symbol. Within each resource-block pair, consisting of 12 subcarriers during one 1 ms subframe, there are thus eight reference symbols.

In general, the values of the reference symbols vary between different reference-symbol positions and also between different cells. Thus, a cell-specific reference signal can be seen as a *two-dimensional cell-specific sequence*. The period of this sequence equals one 10 ms frame. Furthermore, regardless of the cell bandwidth, the reference-signal sequence is defined assuming the maximum possible LTE carrier bandwidth corresponding to 110 resource blocks in the frequency domain. Thus, the basic reference-signal sequence has a length of 8880 symbols.⁷ For cell bandwidths less than the maximum possible value, only the reference symbols within that bandwidth are actually transmitted. The reference symbols in the center part of the band will therefore be the same, regardless of the actual cell bandwidth. This allows for the terminal to estimate the channel corresponding to the center

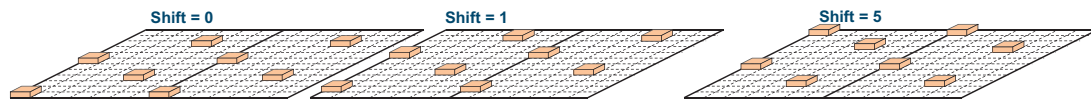
⁵More specifically, CSI-RS are only used in the case of transmission mode 9.

⁶This corresponds to the fifth and fourth OFDM symbols of the slot for normal and extended cyclic prefixes respectively.

⁷Four reference symbols per resource block, 110 resource blocks per slot, and 20 slots per frame.

**FIGURE 10.8**

Structure of cell-specific reference signal within a pair of resource blocks.

**FIGURE 10.9**

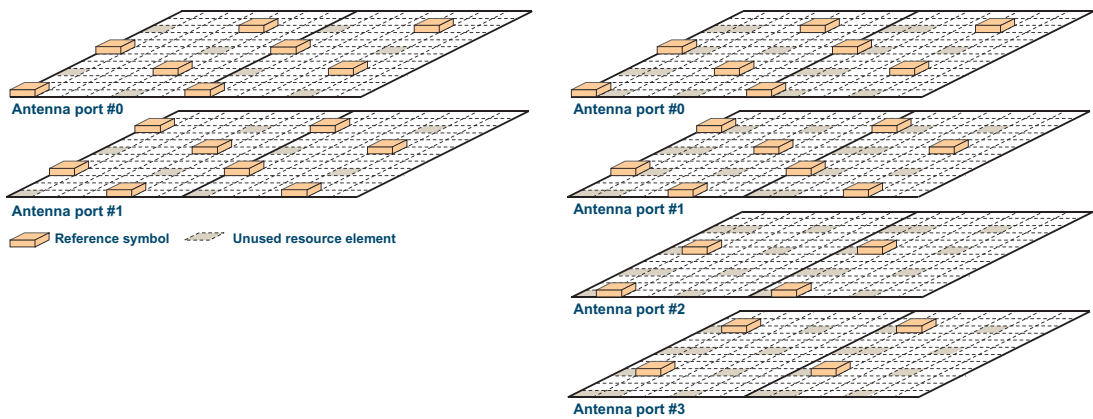
Different reference-signal frequency shifts.

part of the carrier, where, for example, the basic system information of the cell is transmitted on the BCH transport channel, without knowing the cell bandwidth. Information about the actual cell bandwidth, measured as number of resource blocks, is then provided on the BCH.

There are 504 different reference-signal sequences defined for LTE, where each sequence corresponds to one of 504 different *physical-layer cell identities*. As will be described in more detail in Chapter 14, during the so-called *cell-search procedure* the terminal detects the physical-layer identity of the cell as well as the cell frame timing. Thus, from the cell-search procedure, the terminal knows the reference-signal sequence of the cell (given by the physical-layer cell identity) as well as the start of the reference-signal sequence (given by the frame timing).

The set of reference-symbol positions outlined in Figure 10.8 is only one of six possible *frequency shifts* of the reference symbols, as illustrated in Figure 10.9. The frequency shift to use in a cell depends on the physical-layer identity of the cell such that each shift corresponds to 84 different cell identities. Thus, the six different frequency shifts jointly cover all 504 different cell identities. By properly assigning physical-layer cell identities to different cells, different reference-signal frequency shifts may be used in neighboring cells. This can be beneficial, for example, if the reference symbols are transmitted with higher energy compared to other resource elements, also referred to as *reference-signal power boosting*, in order to improve the reference-signal SIR. If reference signals of neighboring cells were transmitted using the same time/frequency resource, the boosted reference symbols of one cell would be interfered by equally boosted reference symbols of all neighboring cells,⁸ implying no gain in the reference-signal SIR. However, if different frequency shifts are used for the reference-signal transmissions of neighboring cells, the reference symbols of one cell will at least partly be interfered by non-reference symbols of neighboring cells, implying an improved reference-signal SIR in the case of reference-signal boosting.

⁸This assumes that the cell transmissions are frame-timing aligned.

**FIGURE 10.10**

Structure of cell-specific reference signals in the case of multiple reference signals. Two reference signals corresponding to two antenna ports (left). Four reference signals corresponding to four antenna ports (right).

10.2.1.2 Multiple Reference Signals

Figure 10.10 illustrates the reference-signal structure in the case of multiple, more specifically two and four, cell-specific reference signals, and corresponding multiple antenna ports, within a cell:⁹

- In the case of two reference signals within a cell (left part of Figure 10.10), the second reference signal is frequency multiplexed with the first reference signal, with a frequency-domain offset of three subcarriers.
- In the case of four reference signals (right part of Figure 10.10), the third and fourth reference signals are frequency multiplexed and transmitted within the *second* OFDM symbol of each slot, thus being time multiplexed with the first and second reference signals.

Obviously, the reference-symbol density for the third and fourth reference signals is lower, compared to the density of the first and second reference signals. The reason for this is to reduce the reference-signal overhead in the case of four reference signals. More specifically, while the first and second reference signals each correspond to a relative overhead of approximately 5% (four reference symbols within a resource block consisting of a total of 84 resource elements), the relative overhead of the third and fourth reference signals is only half of that or approximately 2.5%. This obviously has an impact on the potential of the terminal to track very fast channel variations. However, this can be justified based on an expectation that, for example, high-order spatial multiplexing will mainly be applied to scenarios with low mobility.

It can also be noted that in a resource element carrying reference signals for a certain transmission port, nothing is being transmitted on the antenna ports corresponding to the other reference signals. Thus, a cell-specific reference signal is not interfered by transmissions on other antenna ports. Multi-antenna transmission schemes, such as spatial multiplexing, to a large extent rely on good channel

⁹It is not possible to configure a cell with three cell-specific reference signals.

estimates to suppress interference between the different layers at the receiver side. However, in the channel estimation itself there is obviously no such suppression. Reducing the interference to the reference signals of an antenna port is therefore important in order to allow for good channel estimation, and corresponding good interference suppression, at the receiver side.

Note that, in MBSFN subframes, only the reference signals in the two first OFDM symbols of the subframe, corresponding to the control region of the MBSFN subframe, are actually transmitted. Thus, there is no transmission of cell-specific reference signals within the MBSFN part of the MBSFN subframe.

10.2.2 Demodulation Reference Signals

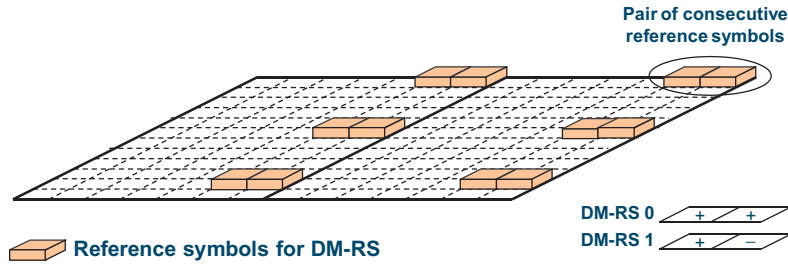
In contrast to cell-specific reference signals, a demodulation reference signal (DM-RS) is intended for a specific terminal and is only transmitted in the resource blocks assigned for transmission to that terminal. Demodulation reference signals are intended to be used for channel estimation for PDSCH transmissions for the case when cell-specific reference signals are not to be used – that is, for non-codebook-based precoding (see Section 10.3.3).

Demodulation reference signals were introduced in the first release of LTE (release 8). However, use of demodulation reference signals was then limited to single-layer transmission – that is, no spatial multiplexing – with at most one reference signal for each terminal. In LTE release 9, transmission based on demodulation reference signals was extended to support also dual-layer transmission, requiring up to two simultaneous reference signals (one for each layer). Transmission based on demodulation reference signals was then further extended in LTE release 10 to support up to eight-layer transmission, corresponding to up to eight reference signals.

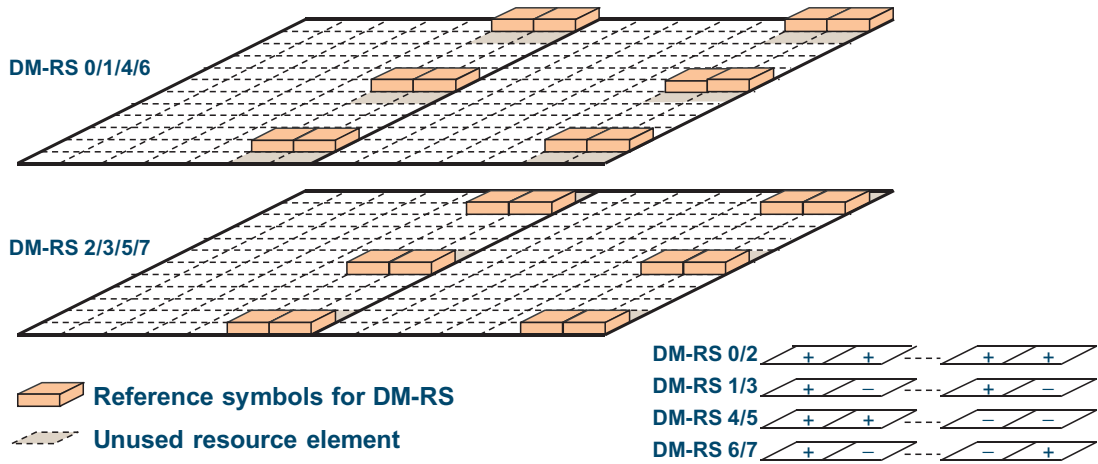
Actually, the DM-RS structure introduced in LTE release 9 was not a straightforward extension of the release-8 demodulation reference signals but rather a new structure, supporting both single-layer and dual-layer transmission. Already at the time of LTE release 9, it was relatively clear that the LTE radio-access technology should be further extended to support up to eight-layer spatial multiplexing in release 10. It was also quite clear that this extension would be difficult to achieve based on the release-8 DM-RS structure. Rather than extending the release-8 structure to support two reference signals and then introduce a completely new structure for release 10, it was decided instead to introduce a new, more future-proof structure already in release 9. Here we will focus on the DM-RS structure introduced in LTE release 9, including the release-10 extension to support up to eight simultaneous reference signals.

Figure 10.11 illustrates the DM-RS structure for the case of one or two reference signals.¹⁰ As can be seen, there are 12 reference symbols within a resource-block pair. In contrast to cell-specific reference signals, for which the reference symbols of one antenna port correspond to unused resource elements for the other antenna ports (see Figure 10.10), in the case of two demodulation reference signals all 12 reference symbols in Figure 10.11 are transmitted *for both reference signals* – that is, *on both antenna ports*. Interference between the reference signals is instead handled by applying mutually orthogonal patterns, also referred to as *orthogonal cover codes* (OCC), to pairs of consecutive reference symbols, as illustrated in the lower right corner of the figure.

¹⁰In the case of TDD, the DM-RS structure is slightly modified in the DwPTS due to the shorter duration of the DwPTS compared with normal downlink subframes.

**FIGURE 10.11**

Structure of demodulation reference signals (DM-RS) for the case of one or two reference signals, including orthogonal cover codes to separate the two reference signals.

**FIGURE 10.12**

Demodulation-reference-signal structure for the case of more than two reference signals.

In addition to the mutually orthogonal patterns, one can also apply a pseudo-random sequence to the reference symbols. This sequence is the same for both reference signals and thus does not impact the orthogonality between the transmitted reference signals. Rather, the pseudo-random sequence is intended to separate different demodulation reference signals targeting different terminals in the case of so-called MU-MIMO transmission, as described in Section 10.3.4.1.

Figure 10.12 illustrates the extended DM-RS structure introduced in LTE release 10 to support more than two reference signals. In this case, there are 24 reference symbols within a resource-block pair. The reference signals are frequency multiplexed in groups of four reference signals while, within each group, the reference signals are separated by means of mutually orthogonal patterns covering four reference symbols (two pairs of consecutive reference symbols). It should be noted that orthogonality between the full set of eight reference signals requires that the channel does not vary over the set of reference symbols over which the orthogonal pattern spans. As the four reference symbols are

not consecutive in time, this implies somewhat stronger limitations on the amount of channel variations that can be tolerated without losing the reference-signal orthogonality. However, more than four demodulation reference signals is only applicable to the case of spatial multiplexing with more than four layers, a transmission configuration that is typically anyway only applicable to low-mobility scenarios. Also note that the orthogonal patterns are selected such that, for four or less reference signals, orthogonality is achieved already over pairs of reference symbols. Thus, for three and four reference signals, the constraints on channel variations are the same as for two reference signals according to Figure 10.11.

When demodulation reference signals are transmitted within a resource block, PDSCH mapping to the time–frequency grid of the resource block will be modified to avoid the resource elements in which the reference signals are transmitted. Although this modified mapping is not “understood” by earlier-release terminals not supporting demodulation reference signals, this is not a problem as demodulation reference signals will only be transmitted in resource blocks that are scheduled for PDSCH transmission to terminals of later releases supporting demodulation reference signals and thus “understanding” the modified PDSCH mapping. As will be seen below, the situation is somewhat different in the case of CSI-RS.

As the number of transmitted layers may vary dynamically, the number of transmitted demodulation reference signals may also vary. Thus, the transmission may dynamically change between the DM-RS structures in Figures 10.11 and 10.12. The terminal is informed about the number of transmitted layers (the “transmission rank”) as part of the scheduling assignment and will thus know the DM-RS structure and associated PDSCH mapping for each subframe.

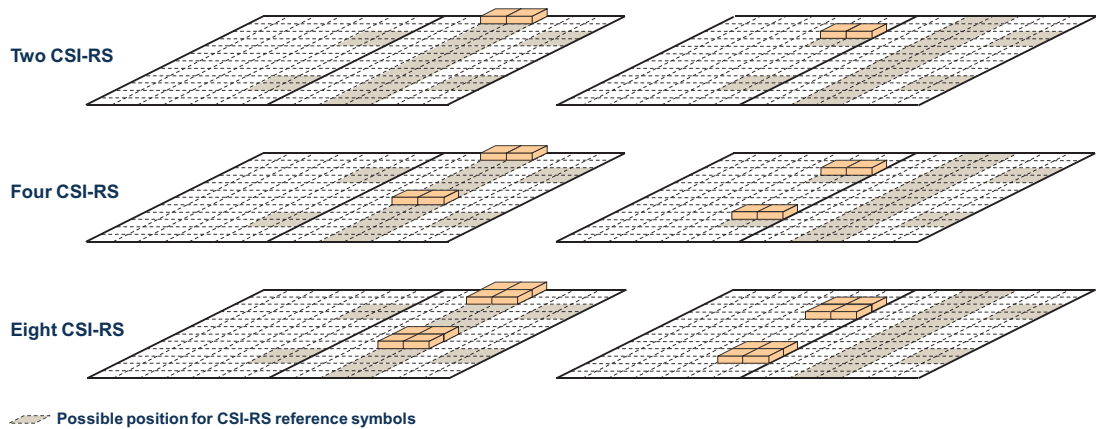
10.2.3 CSI Reference Signals

Support for CSI reference signals (CSI-RS) was introduced in LTE release 10. CSI-RS is intended to be used by terminals to acquire channel-state information when demodulation reference signals are used for channel estimation. More specifically, CSI-RS are to be used in transmission mode 9.¹¹

As described in Section 10.2.1, the cell-specific reference signals, already available in the first release of LTE, can also be used to acquire channel-state information. There are at least two reasons why CSI-RS was introduced in release 10 as a complement to the already supported cell-specific reference signals:

- In LTE release 8 there could be at most four cell-specific reference signals in a cell. As already mentioned, multi-antenna transmission was extended in LTE release 10 to support downlink spatial multiplexing with up to eight layers, corresponding to eight transmit antennas at the base station. Thus, there was a need to extend the LTE CSI capabilities beyond what was possible with the release 8 cell-specific reference signals.
- The time-frequency density of the cell-specific reference signals was selected to allow for channel estimation for coherent demodulation even with the most extreme channel conditions, including very fast channel variations in both the time and frequency domains. As a consequence, each

¹¹The reason why CSI-RS are not used for transmission modes 7 and 8 is simply that these transmission modes were introduced in LTE releases 8 and 9 respectively, and CSI-RS was not introduced until LTE release 10, together with the introduction of transmission mode 9.

**FIGURE 10.13**

Examples of reference-signal positions for different number of CSI-RS within a cell. In the case of a single CSI-RS, the same structure as for two CSI-RS is used.

cell-specific reference signal implies a relatively high overhead (approximately 5% and 2.5% for the first/second and third/fourth reference signals respectively). On the other hand, in the case of a reference signal only targeting CSI, such high time-frequency density, and correspondingly relatively large overhead, is not necessary. Thus, rather than extending the cell-specific reference signals it was found to be more efficient to introduce a new type of reference signals (CSI-RS) only targeting CSI, with a flexible and, in general, lower time/frequency density and corresponding lower overhead per reference signal.

10.2.3.1 CSI-RS Structure

A cell can be configured with one, two, four, or eight CSI-RS. The exact CSI-RS structure, including the exact set of resource elements used for CSI-RS in a resource block, depends on the number of CSI-RS configured within the cell and may also be different for different cells. More specifically, within a resource-block pair there are 40 *possible* positions for the reference symbols of CSI-RS and, in a given cell, a subset of the corresponding resource elements is used for CSI-RS transmission, as described below and exemplified in Figure 10.13.¹²

- In the case of two CSI-RS within a cell, the CSI-RS consists of two consecutive reference symbols per resource-block pair, as illustrated in the upper part of Figure 10.13. The two CSI-RS are separated by applying mutually *orthogonal cover codes* (OCC) to the two reference symbols, similar to demodulation reference signals described in Section 10.2.2. Thus, there is a possibility for 20 different CSI-RS configurations, two of which are shown in Figure 10.13, based on the 40 *possible* reference-symbol positions.

¹²For TDD, there are additional *possible* locations for reference symbols for CSI-RS beyond what is illustrated in Figure 10.13.

- In the case of four/eight CSI-RS, the CSI-RS are pair-wise frequency multiplexed, as illustrated in the middle/lower part of Figure 10.13. Thus, there is a possibility for ten/five different CSI-RS configurations, each supporting four/eight CSI-RS.

In the case of a single CSI-RS, the same structure as for two CSI-RS (upper part of Figure 10.13) is used, although with only one of the orthogonal cover codes. The overhead of a single CSI-RS, in terms of occupied resource elements, is thus the same as for two CSI-RS.

In the time domain, the CSI-RS can be transmitted with different periods, ranging from a period of 5 ms (every fifth subframe) to 80 ms (every eighth frame). With a 5 ms period, the overhead per CSI-RS is thus roughly 0.12% per CSI-RS (one resource element per CSI-RS per resource-block pair, but only in every fifth subframe), with longer periods implying correspondingly less overhead.

The exact subframe in which CSI-RS is transmitted in a cell can also be configured, allowing for separation of CSI-RS transmissions between cells also in the time domain, in addition to using different sets of resource elements within the same subframe, as illustrated in Figure 10.13.

In subframes in which CSI-RS is transmitted, it is transmitted in every resource block in the frequency domain. In other words, a CSI-RS transmission covers the entire cell bandwidth.

As mentioned above, when demodulation reference signals are transmitted within a resource block, the corresponding resource elements on which the reference signals are transmitted are explicitly avoided when mapping PDSCH symbols to the resource block. This “modified” PDSCH mapping, which is obviously not “understood” by earlier-release terminals, is possible as demodulation reference signals can be assumed to be transmitted only in resource blocks in which terminals supporting such reference signals are scheduled – that is, terminals of release 10 or later.¹³ Expressed alternatively, an earlier-release terminal can be assumed never to be scheduled in a resource block in which demodulation reference signals are transmitted and thus in which the modified PDSCH mapping is used.

The situation is somewhat different for CSI-RS. As CSI-RS is transmitted within all resource blocks in the frequency domain, it would imply a strong scheduler constraint to assume that release-8/9 terminals could never be scheduled in a resource block in which CSI-RS is transmitted. If the PDSCH mapping were modified to explicitly avoid the resource elements in which CSI-RS is transmitted, the mapping would not be recognized by a release-8/9 terminal. Instead, in the case of resource blocks scheduled to release-8/9 terminals, the PDSCH is mapped exactly according to release 8 – that is, the mapping is not modified to avoid the resource elements on which CSI-RS is to be transmitted. The CSI-RS is then simply transmitted on top of the corresponding PDSCH symbols.¹⁴ This will obviously impact the PDSCH demodulation performance, as some PDSCH symbols will be highly corrupted. However, the remaining PDSCH symbols will not be impacted and the PDSCH will still be decodable, although with somewhat reduced performance.

On the other hand, if a release-10 terminal is scheduled in a resource block in which CSI-RS is transmitted, the PDSCH mapping is modified to explicitly avoid the resource elements on which the CSI-RS is transmitted, similar to demodulation reference signals. Thus, if CSI-RS is transmitted in a resource block, the PDSCH mapping to that resource block will be somewhat different depending on the release of the terminal being scheduled in the resource block.

¹³Partly also for terminals of release 9, but then only for a maximum of two DM-RS.

¹⁴In practice, the base station would probably not transmit PDSCH in these resource elements in order to avoid interference to the CSI-RS transmission. The key thing is that the mapping of the remaining PDSCH symbols is in line with release 8.

It should be noted that release-8 mapping also has to be used for transmission of, for example, system information and paging messages, as such transmissions must be possible to receive also by release-8/9 terminals.

10.2.3.2 Muted CSI-RS

Figure 10.13 illustrated the set of resource elements that can *potentially* be used for CSI-RS, a subset of which is used to create the different CSI-RS within a cell. In normal operation, the remaining resource elements highlighted in Figure 10.13 are used for PDSCH transmission within the cell.

However, it is also possible to additionally configure one or several subsets of the resource elements outlined in Figure 10.13 as *muted CSI-RS*. A muted CSI-RS has the same structure as a normal (non-muted) CSI-RS except that nothing is actually transmitted on the corresponding resource elements. A muted CSI-RS can thus be seen as a normal (non-muted) CSI-RS *with zero power*.

The intention with muted CSI-RS is to be able to create “transmission holes” corresponding to actual CSI-RS transmissions in other (neighboring) cells. This serves two purposes:

1. To make it possible for a terminal to receive CSI-RS of neighboring cells, without being severely interfered by transmissions in its own cell. CSI estimation on neighboring cells would be of interest if support for different multi-cell-transmission techniques such as *CoMP* (see Chapter 20) were introduced in future LTE releases.
2. To reduce the interference to CSI-RS transmissions in other cells. This is especially applicable to so-called *heterogeneous network deployments* where overlapping cell layers with cells of substantially different power are to coexist in the same spectrum. Interference handling in heterogeneous deployments is further discussed in Chapter 13.

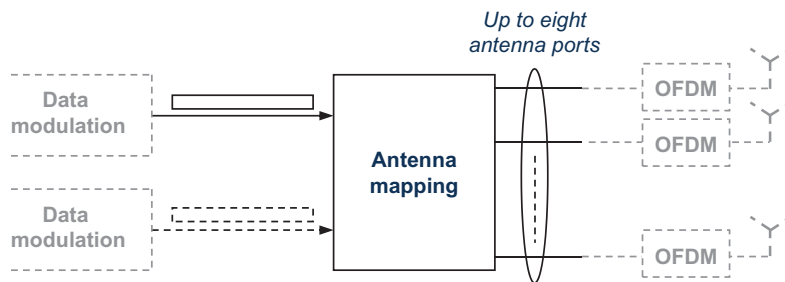
There can be one or multiple sets of muted CSI-RS in a cell. In the case of interference avoidance to neighboring cells (corresponding to point 2 above), there is typically sufficient with a single set of muted CSI-RS as one should typically avoid interference to a single lower-power cell layer. When muted CSI-RS are used to remove own-cell interference in order to be able to better receive CSI-RS from neighboring cells (point 1 above), multiple sets of muted CSI-RS are typically needed as different neighboring cells can be assumed to use mutually different CSI-RS configurations.

10.3 MULTI-ANTENNA TRANSMISSION

As illustrated in Figure 10.14, multi-antenna transmission in LTE can, in general, be described as a mapping from the output of the data modulation to the different antennas ports. The input to the antenna mapping thus consists of the modulation symbols (QPSK, 16QAM, 64QAM) corresponding to the one or two transport blocks. To be more specific, there is one transport block per TTI except for spatial multiplexing, in which case there may be two transport blocks per TTI (see also Section 10.3.2).

The output of the antenna mapping is a set of symbols for each antenna port. The symbols of each antenna port are subsequently applied to the OFDM modulator – that is, mapped to the basic OFDM time–frequency grid corresponding to that antenna port.

The different multi-antenna transmission schemes correspond to different so-called *transmission modes*. There are currently nine different transmission modes defined for LTE. They differ in terms of the specific structure of the antenna mapping of Figure 10.14 but also in terms of what reference

**FIGURE 10.14**

General structure for LTE downlink multi-antenna transmission. Modulation symbols corresponding to one or two transport blocks mapped to up to eight antenna ports.

signals are assumed to be used for demodulation (cell-specific reference signals or demodulation reference signals respectively) and the type of CSI feedback they rely on. Transmission mode 1 corresponds to single-antenna transmission while the remaining transmission modes correspond to different multi-antenna transmission schemes, including transmit diversity, beam-forming, and spatial multiplexing. Actually, LTE supports both beam-forming and spatial multiplexing as part of more general *antenna precoding*. There are two approaches to downlink antenna precoding, which differ depending on whether cell-specific reference signals or demodulation reference signals are used for channel estimation. These two approaches can also be referred to as *codebook-based precoding* and *non-codebook-based precoding* respectively, the reason for these specific names being further clarified below.

The list below summarizes the currently defined transmission modes and the associated multi-antenna transmission schemes. The different multi-antenna transmission schemes are described in more detail in the sections below.

- **Transmission mode 1:** Single-antenna transmission.
- **Transmission mode 2:** Transmit diversity.
- **Transmission mode 3:** *Open-loop* codebook-based precoding in the case of more than one layer, transmit diversity in the case of rank-one transmission.
- **Transmission mode 4:** *Closed-loop* codebook-based precoding.
- **Transmission mode 5:** Multi-user-MIMO version of transmission mode 4.
- **Transmission mode 6:** Special case of closed-loop codebook-based precoding limited to single-layer transmission.
- **Transmission mode 7:** Release-8 non-codebook-based precoding supporting only single-layer transmission.
- **Transmission mode 8:** Release-9 non-codebook-based precoding supporting up to two layers.
- **Transmission mode 9:** Release-10 non-codebook-based precoding supporting up to eight layers.

Strictly speaking, transmission modes are only applicable for DL-SCH transmission. Thus, a certain transmission mode should not be seen as identical to a certain multi-antenna transmission configuration. Rather, a certain multi-antenna transmission scheme is applied to DL-SCH transmission

according to the list above when the corresponding transmission mode is configured for DL-SCH transmission. However, the same multi-antenna transmission scheme may also be applied to other types of transmissions, such as transmission of BCH and L1/L2 control signaling.¹⁵ However, this does not mean that the corresponding transmission mode is applied to such transmissions.

It should also be mentioned that, although a certain multi-antenna transmission scheme can be seen as being associated with a certain transmission mode, for transmission modes 3–9 there is a possibility for fall-back to transmit diversity without implying that the configured transmission mode is changed.

10.3.1 Transmit Diversity

Transmit diversity can be applied to any downlink physical channel. However, it is especially applicable to transmissions that cannot be adapted to varying channel conditions by means of link adaptation and/or channel-dependent scheduling, and thus for which diversity is more important. This includes transmission of the BCH and PCH transport channels, as well as L1/L2 control signaling. Actually, as already mentioned, transmit diversity is the *only* multi-antenna transmission scheme applicable to these channels. Transmit diversity is also used for transmission of DL-SCH when transmission mode 2 is configured. Furthermore, as also already mentioned, transmit diversity is a fall-back “mode” for transmission mode 3 and higher. More specifically, a scheduling assignment using DCI format 1A (see Section 10.4.4) implies the use of transmit diversity regardless of the configured transmission mode.

Transmit diversity assumes the use of cell-specific reference signals for channel estimation. Thus, a transmit-diversity signal is always transmitted on the same antenna ports as the cell-specific reference signals. Actually, if a cell is configured with two cell-specific reference signals, transmit diversity for two antenna ports *must be used* for BCH and PCH, as well as for the L1/L2 control signaling. Similarly, if four cell-specific reference signals are configured for the cell, transmit diversity for four antenna ports has to be used for the transmission of these channels. In this way, a terminal does not have to be explicitly informed about what multi-antenna transmission scheme is used for these channels. Rather, this is given implicitly from the number of cell-specific reference signals configured for a cell.¹⁶

10.3.1.1 Transmit Diversity for Two Antenna Ports

In the case of two antenna ports, LTE transmit diversity is based on *Space-Frequency Block Coding* (SFBC), as described in Chapter 5. As can be seen from Figure 10.15, SFBC implies that two consecutive modulation symbols S_i and S_{i+1} are mapped directly to frequency-adjacent resource elements on the first antenna port. On the second antenna port the frequency-swapped and transformed symbols $-S_{i+1}^*$ and S_i^* are mapped to the corresponding resource elements, where “*” denotes complex conjugate.

Figure 10.15 also indicates how the antenna ports on which a transmit-diversity signal is being transmitted correspond to the cell-specific reference signals, more specifically CRS 0 and CRS 1 in

¹⁵Actually, only single-antenna transmission and transmit diversity is applicable to BCH and L1/L2 control signaling.

¹⁶Actually, the situation is partly the opposite, i.e. the terminal blindly detects the number of antenna ports used for BCH transmission and, from that, decides on the number of cell-specific reference signals configured within the cell.

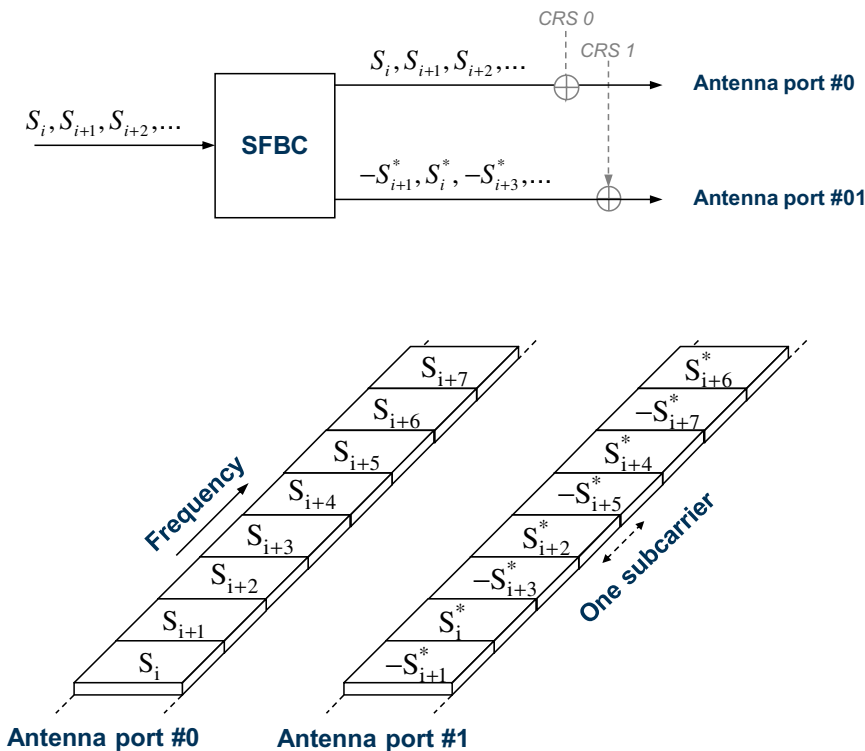


FIGURE 10.15

Transmit diversity for two antenna ports – SFBC.

the case of two antenna ports. Note that one should not interpret this such that the CRS is specifically transmitted for this transmit-diversity signal. The cell-specific reference signals are always transmitted and can be jointly used by multiple transmissions within a cell.

10.3.1.2 Transmit Diversity for Four Antenna Ports

In the case of four antenna ports, LTE transmit diversity is based on a combination of SFBC and *Frequency-Switched Transmit Diversity* (FSTD). As can be seen in Figure 10.16, combined SFBC/FSTD implies that pairs of modulation symbols are transmitted by means of SFBC with transmission alternating between pairs of antenna ports (antenna ports 0 and 2 and antenna ports 1 and 3 respectively). For the resource elements where transmission is on one pair of antenna ports, there is no transmission on the other pair of antenna ports. Thus, combined SFBC/FSTD in some sense operates on groups of four modulation symbols and corresponding groups of four frequency-consecutive resource elements on each antenna port. As mentioned in Section 10.4.1, this is the reason for the use of *resource-element groups*, each consisting of four resource elements, when defining the mapping of the L1/L2 control signaling to the physical resource.

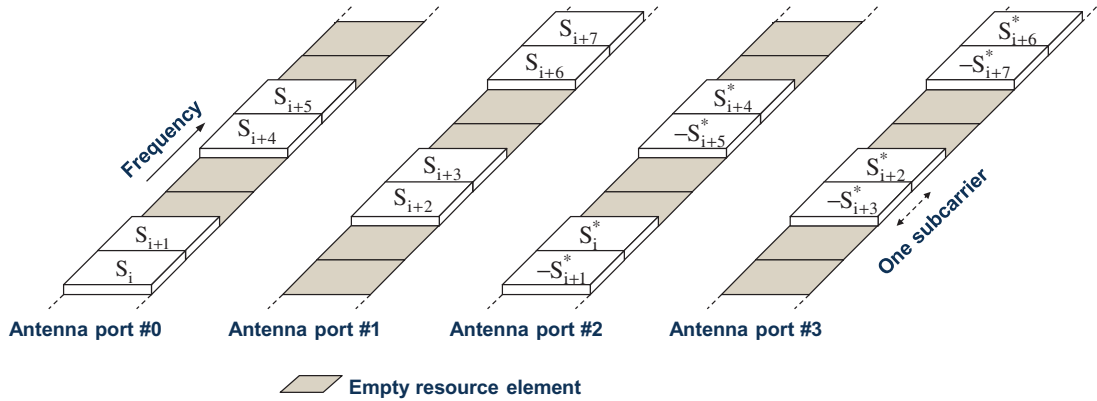


FIGURE 10.16

Transmit diversity for four antenna ports – combined SFBC/FSTD.

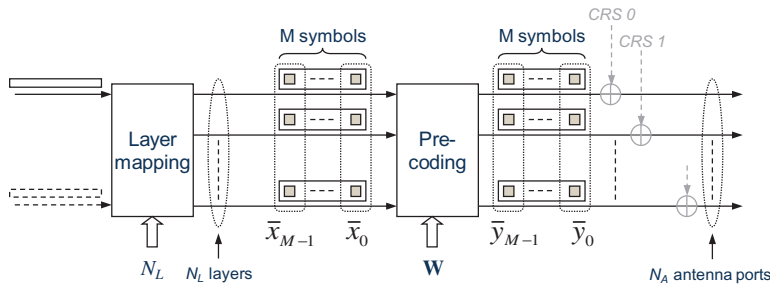


FIGURE 10.17

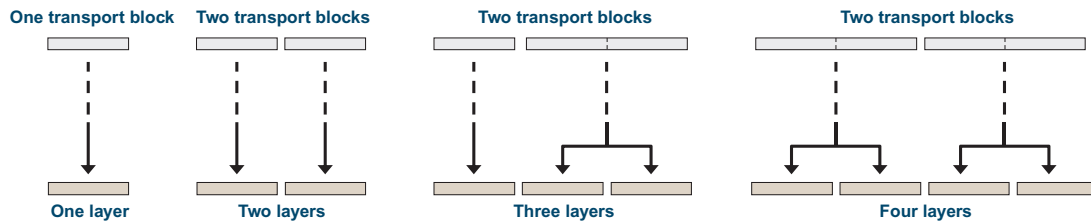
The basic structure of LTE codebook-based antenna precoding. The figure also indicates how cell-specific reference signals (CRS) are applied after precoding.

10.3.2 Codebook-Based Precoding

The basic processing for codebook-based precoding is illustrated in Figure 10.17. The modulation symbols corresponding to one or two transport blocks are first mapped to N_L layers. The number of layers may range from a minimum of one layer up to a maximum number of layers equal to the number of antenna ports.¹⁷ The layers are then mapped to the antenna ports by means of the *precoder functionality*. As codebook-based precoding relies on the cell-specific reference signals for channel estimation, and there are at most four cell-specific reference signals in a cell, codebook-based precoding allows for a maximum of four antenna ports and, as a consequence, a maximum of four layers.

The mapping to layers is outlined in Figure 10.18 for the case of an initial transmission. There is one transport block in the case of a single layer ($N_L = 1$) and two transport blocks for two or more

¹⁷In practice, the number of layers is also limited by, and should not exceed, the number of receive antennas available at the terminal.

**FIGURE 10.18**

Transport-block-to-layer mapping for codebook-based antenna precoding (initial transmission).

layers ($N_L > 1$). In the case of a hybrid-ARQ retransmission, if only one of two transport blocks needs to be retransmitted and that transport block was mapped to two layers for the initial transmission, the retransmission may also be carried out on two layers. Thus, in the case of a retransmission, a single transport block may also be transmitted using two layers.

The mapping to layers is such that the number of modulation symbols on each layer is the same and equal to the number of symbols to be transmitted on each antenna port. Thus, in the case of three layers, there should be twice as many modulation symbols corresponding to the second transport block (mapped to the second and third layers) compared to the first transport block (mapped to the first layer). This is ensured by the selection of an appropriate transport-block size in combination with the rate-matching functionality. In the case of four layers, the first transport block is mapped to the first and second layers while the second transport block is mapped to the third and fourth layers. In this case, the number of modulation symbols should thus be the same for the two transport blocks. For one transport block being mapped to two layers, the modulated symbols corresponding to the transport block are mapped to the layers in an alternating fashion – that is, every second modulation symbol is mapped to the first and second layer respectively.

In the case of multi-antenna precoding the number of layers is also often referred to as the *transmission rank*.¹⁸ The transmission rank can vary dynamically, for example based on the number of layers that can be supported by the channel. The latter is sometimes also referred to as the *channel rank*.

After layer mapping, a set of N_L symbols (one symbol from each layer) is linearly combined and mapped to the antenna ports. This combination/mapping can be described by a *precoder matrix* $\mathbf{W}$ of size $N_A \times N_L$, where N_A is the number of antenna ports. More specifically, the vector $\bar{y}_i$ of size N_A , consisting of one symbol for each antenna port, is given by $\bar{y}_i = \mathbf{W} \cdot \bar{x}_i$, where the vector $\bar{x}_i$ of size N_L consists of one symbol from each layer. As the number of layers can vary dynamically, also the number of columns of the precoder matrix will vary dynamically. Specifically, in the case of a single layer, the precoder matrix $\mathbf{W}$ is a vector of size $N_A \times 1$ that provides beam-forming for a single modulation symbol.

Figure 10.17 also indicates how the cell-specific reference signals are applied after antenna precoding. Channel estimation based on the cell-specific reference signals will thus reflect the channel for each antenna port *not including the precoding*. As a consequence, the terminal receiver must have

¹⁸In the LTE specification, transmit diversity is actually also described as transmission using *multiple layers*. However, transmit diversity is still a *single-rank* transmission scheme.

explicit knowledge about what precoding has been applied at the transmitter side in order to properly process the received signal and recover the different layers. Once again, the figure should not be interpreted such that cell-specific reference signals are inserted specifically for a given PDSCH transmission.

There are two operational modes for codebook-based precoding, *closed-loop operation* and *open-loop operation*. These two modes differ in terms of the exact structure of the precoder matrix and how the matrix is selected by the network and made known to the terminal.

10.3.2.1 Closed-Loop Operation

In case of closed-loop precoding it is assumed that the network selects the precoder matrix based on feedback from the terminal.

Based on measurements on the cell-specific reference signals, the terminal selects a suitable transmission rank and corresponding precoder matrix. Information about the selected rank and precoder matrix is then reported to the network in the form of a *Rank Indication* (RI) and a *Precoder-Matrix Indication* (PMI), as described in Section 11.4. It is important to understand though that the RI and PMI are only recommendations and the network does not need to follow the RI/PMI provided by the terminal when selecting the actual transmission rank and precoder matrix to be used for transmission to the terminal. When not following the terminal recommendation, the network must explicitly inform the terminal what precoder matrix is used for the downlink transmission. On the other hand, if the network uses the precoder matrix recommended by the terminal, only a confirmation that the network is using the recommended matrix is signaled.

To limit the signaling on both uplink and downlink only a limited set of precoder matrices, also referred to as the *codebook*, is defined for each transmission rank for a given number of antenna ports. Both the terminal (when reporting PMI) and the network (when selecting the actual precoder matrix to use for the subsequent downlink transmission to the terminal) should select a precoder matrix from the corresponding codebook. Thus, for terminal PMI reporting, as well as when the network informs the terminal about the actual precoder matrix used for the downlink transmission, only the index of the selected matrix needs to be signaled.

As LTE supports multi-antenna transmission using two and four antenna ports, codebooks are defined for:

- Two antenna ports and one and two layers, corresponding to precoder matrices of size 2×1 and 2×2 respectively.
- Four antenna ports and one, two, three, and four layers, corresponding to precoder matrices of size 4×1 , 4×2 , 4×3 , and 4×4 respectively.

As an example, the precoder matrices specified for the case of two antenna ports are illustrated in Table 10.3. As can be seen, there are four 2×1 precoder matrices for single-layer transmission and three 2×2 precoder matrices for two-layer transmission. In the same way, sets of 4×1 , 4×2 , 4×3 , and 4×4 matrices are defined for the case of four antenna ports and one, two, three, and four layers respectively. It should be pointed out that the first rank-2 (2×2) matrix in Table 10.3 is not used in closed-loop operation but only for *open-loop precoding*, as described in the next section.

Even if the network is following the precoder-matrix recommendation provided by the terminal, the network may, for different reasons, decide to use a lower rank for the transmission, so-called *rank*

Table 10.3 Precoder Matrices for Two Antenna Ports and One and Two Layers
(the First 2×2 Matrix Is Only Used for Open-Loop Precoding)

One layer	$\frac{1}{\sqrt{2}} \begin{bmatrix} +1 \\ +1 \end{bmatrix}$	$\frac{1}{\sqrt{2}} \begin{bmatrix} +1 \\ -1 \end{bmatrix}$	$\frac{1}{\sqrt{2}} \begin{bmatrix} +1 \\ +j \end{bmatrix}$	$\frac{1}{\sqrt{2}} \begin{bmatrix} +1 \\ -j \end{bmatrix}$
Two layers	$\frac{1}{\sqrt{2}} \begin{bmatrix} +1 & 0 \\ 0 & +1 \end{bmatrix}$	$\frac{1}{2} \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$	$\frac{1}{2} \begin{bmatrix} +1 & +1 \\ +j & -j \end{bmatrix}$	

override. In that case the network will use a subset of the columns of the recommended precoder matrix. The network precoder confirmation will then include explicit information about the set of columns being used or, equivalently, about the set of layers being transmitted.

There is also a possibility to apply closed-loop precoding strictly limited to single-layer (rank-1) transmission. This kind of multi-antenna transmission is associated with *transmission mode 6*. The reason for defining an additional transmission mode limited to single-layer transmission rather than relying on the general closed-loop precoding associated with transmission mode 4 is that, by strictly limiting to single-layer transmission, the signaling overhead on both downlink and uplink can be reduced. Transmission mode 6 can, for example, be configured for terminals with low SINR for which multi-layer transmission would anyway not apply.

10.3.2.2 Open-Loop Operation

Open-loop precoding does not rely on any detailed precoder recommendation being reported by the terminal and does not require any explicit network signaling of the actual precoder used for the downlink transmission. Instead, the precoder matrix is selected in a predefined and deterministic way known to the terminal in advance. One use of open-loop precoding is in high-mobility scenarios where accurate feedback is difficult to achieve due to the latency in the PMI reporting.

The basic transmission structure for open-loop precoding is aligned with the general codebook-based precoding outlined in Figure 10.17 and only differs from closed-loop precoding in the structure of the precoding matrix $\mathbf{W}$.

In the case of open-loop precoding, the precoder matrix can be described as the product of two matrices $\mathbf{W}'$ and $\mathbf{P}$, where $\mathbf{W}'$ and $\mathbf{P}$ are of size $N_A \times N_L$ and $N_L \times N_L$ respectively:

$$\mathbf{W} = \mathbf{W}' \cdot \mathbf{P}. \quad (10.1)$$

In the case of two antenna ports, the matrix $\mathbf{W}'$ is the normalized 2×2 identity matrix:¹⁹

$$\mathbf{W}' = \frac{1}{\sqrt{2}} \begin{bmatrix} +1 & 0 \\ 0 & +1 \end{bmatrix}. \quad (10.2)$$

¹⁹As non-codebook-based precoding is not used for rank-1 transmission (see below), there is no need for any matrix $\mathbf{W}'$ of size 2×1 .

In the case of four antenna ports, $\mathbf{W}'$ is given by cycling through four of the defined $4 \times N_L$ precoder matrices and is different for consecutive resource elements.

The matrix $\mathbf{P}$ can be expressed as $\mathbf{P} = \mathbf{D}_i \cdot \mathbf{U}$, where $\mathbf{U}$ is a constant matrix of size $N_L \times N_L$ and $\mathbf{D}_i$ is a matrix of size $N_L \times N_L$ that varies between subcarriers (indicated by the index i). As an example, the matrices $\mathbf{U}$ and $\mathbf{D}_i$ for the case of two layers ($N_L = 2$) are given by:

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & e^{-j2\pi/2} \end{bmatrix}, \quad \mathbf{D}_i = \begin{bmatrix} 1 & 0 \\ 0 & e^{-j2\pi i/2} \end{bmatrix}. \quad (10.3)$$

The basic idea with the matrix $\mathbf{P}$ is to average out any differences in the channel conditions as seen by the different layers.

Similar to closed-loop precoding, the transmission rank for open-loop precoding can also vary dynamically down to a minimum of two layers. Transmission mode 3, associated with open-loop precoding, also allows for rank-1 transmission. In that case, transmit diversity as described in Section 10.3.1 is used – that is, SFBC for two antenna ports and combined SFBC/FSTD for four antenna ports.

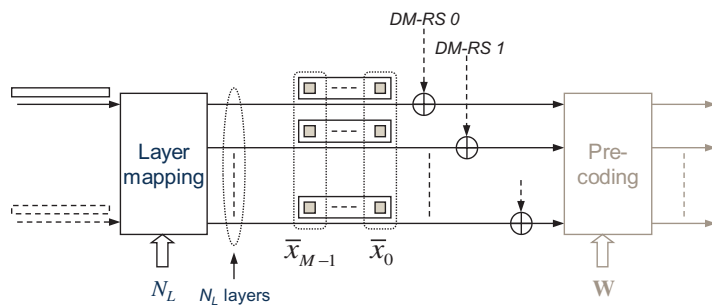
10.3.3 Non-Codebook-Based Precoding

Similar to codebook-based precoding, non-codebook-based precoding is only applicable to DL-SCH transmission. Non-codebook-based precoding was introduced in LTE release 9 but was then limited to a maximum of two layers. The extension to eight layers was then introduced as part of release 10. The release-9 scheme, associated with transmission mode 8, is a subset of the extended release-10 scheme (transmission mode 9).

There is also a release-8 non-codebook-based precoding defined, associated with transmission mode 7. Transmission mode 7 relies on the release-8 demodulation reference signals mentioned but not described in detail in Section 10.2.2 and only supports single-layer transmission. In this description we will focus on non-codebook-based precoding corresponding to transmission modes 8 and 9.

The basic principles for non-codebook-based precoding can be explained based on [Figure 10.19](#) (where the precoder is intentionally shaded; see below). As can be seen, this figure is very similar to the corresponding figure illustrating codebook-based precoding ([Figure 10.17](#)), with layer mapping of modulation symbols corresponding to one or two transport blocks followed by precoding. The layer mapping also follows the same principles as that of codebook-based precoding (see [Figure 10.18](#)) but is extended to support up to eight layers. In particular, at least for an initial transmission, there are two transport blocks per TTI except for the case of a single layer, in which case there is only one transport block within the TTI. Similar to codebook-based precoding, for hybrid-ARQ retransmissions there may in some cases be a single transport block also in the case of multi-layer transmission.

The main difference in [Figure 10.19](#) compared to [Figure 10.17](#) (codebook-based precoding) is the presence of demodulation reference signals (DM-RS) before the precoding. The transmission of precoded reference signals allows for demodulation and recovery of the transmitted layers at the receiver side *without explicit receiver knowledge of the precoding applied at the transmitter side*. Put simply, channel estimation based on precoded demodulation reference signals will reflect the channel experienced by the layers, *including the precoding*, and can thus be used directly for coherent demodulation of the different layers. There is thus no need to signal any precoder-matrix information

**FIGURE 10.19**

Basic principles for non-codebook-based antenna precoding.

to the terminal, which only needs to know the number of layers – that is, the transmission rank. As a consequence, the network can select an arbitrary precoder and there is no need for any explicit codebook to select from. This is the reason for the term *non-codebook-based* precoding. It should be noted though that non-codebook-based precoding may still rely on codebooks for the terminal feedback, as described below.

The possibility to select an arbitrary precoder matrix for the transmission is also the reason why the precoder is shaded in Figure 10.19. The precoder part of Figure 10.19 is not visible in the LTE specification and, strictly speaking, in the case of non-codebook-based precoding the antenna mapping defined according to Figure 10.14 consists of only the layer mapping. This also means that the antenna ports defined in Figure 10.14 correspond to the different layers in Figure 10.19 or, expressed differently, precoding occurs *after* the antenna ports.

Still, there must be a way for the network to select a suitable precoder matrix for the transmission. There are essentially two ways by which this can be done in the case of non-codebook-based precoding.

The network may estimate the uplink channel state, for example based on transmission of uplink sounding reference signals as described in the next chapter, and rely on uplink/downlink channel reciprocity when selecting the precoder matrix to use for the downlink transmission. This is especially of interest for TDD operation for which the use of the same frequency for uplink and downlink transmission typically leads to a higher degree of downlink/uplink channel reciprocity. It should be noted though that if the terminal uses multiple receive antennas, it also has to transmit on multiple antennas in order for uplink measurements to fully reflect the downlink channel state.

Alternatively, the network may rely on terminal feedback for precoder-matrix selection. For transmission modes 8 and 9, this feedback is actually very similar to the corresponding feedback for closed-loop codebook-based precoding. In particular, the feedback is based on the same precoder matrices as for codebook-based precoding (see Table 10.3 for an example) but extended to handle up to eight antenna ports. Furthermore, for transmission mode 9, the terminal measurements should be based on CSI-RS, as described in Section 10.2.3, rather than the cell-specific reference signals.

Thus, despite the name, non-codebook-based precoding may also use defined codebooks. However, in contrast to codebook-based precoding, the codebooks are then only used for the terminal PMI reporting and not for the actual downlink transmission.

10.3.4 Downlink Multi-User MIMO

Spatial multiplexing implies the transmission of multiple layers – that is, multiple parallel transmissions on the same time–frequency resource, to the same terminal. The presence of multiple antennas at both the transmitter and receiver sides in combination with transmitter and/or receiver signal processing is then used to suppress interference between the different layers.

Spatial multiplexing is also often referred to as *Multi-Input-Multi-Output* (MIMO) transmission, reflecting the fact that the channel in the case of spatial multiplexing can be seen as having *multiple inputs*, corresponding to the multiple transmit antennas, and *multiple outputs*, corresponding to the multiple receive antennas. The more specific term *Single-User MIMO* (SU-MIMO) is also often used for reasons that will become obvious below.

The term *Multi-User MIMO* (MU-MIMO) is, in 3GPP, used to denote transmission to *different* terminals using the same time–frequency resource, in practice relying on multiple antennas at the transmitter (network) side to separate the two transmissions.

In principle, one could realize MU-MIMO as a direct extension to spatial multiplexing, with the different layers simply being intended for different terminals. The set of terminals would demodulate and perhaps also decode the full set of layers. The data on a layer not intended for a specific terminal would then just be discarded by that terminal after demodulation/decoding.

However, such an approach would imply that all terminals involved in the MU-MIMO transmission would need to know about the full set of layers being transmitted. It would also imply that one would need to have exactly the same resource assignment for all terminals involved in the MU-MIMO transmission. All terminals would also need to include the full set of receive antennas necessary to receive the overall multi-layer transmission.

Instead, 3GPP has chosen an MU-MIMO approach that does not require terminal knowledge about the presence of the other transmissions, allows for only partly overlapping resource assignments, and, at least in principle, does not require the presence of multiple receive antennas at the mobile terminal.²⁰ There are two approaches to MU-MIMO specified in 3GPP, one being an integrated part of transmission modes 8 and 9 corresponding to non-codebook-based precoding, and one being based on codebook-based precoding but associated with a special transmission mode, *transmission mode 5*.

10.3.4.1 MU-MIMO within Transmission Modes 8/9

The only thing that is needed to support MU-MIMO as part of transmission modes 8 and 9 – that is, within the concept of non-codebook precoding – is to be able to assign different demodulation reference signals for the transmissions to the different terminals. As discussed in Section 10.2.2, for transmission mode 8 it is possible to define two orthogonal reference signals by means of two mutually orthogonal patterns, or *orthogonal cover codes* (OCC). In the case of spatial multiplexing, these OCC and corresponding reference signals are used for the transmission of the two layers. However, they can also be assigned to different users to enable MU-MIMO transmission within transmission mode 8.

In addition, as also mentioned in Section 10.2.2, a pseudo-random sequence is applied to the demodulation reference signals. This sequence is not terminal specific in the sense that each terminal

²⁰Note, though, that the LTE performance requirements in general assume the presence of at least two receive antennas at the mobile terminal.

has its own sequence. Rather, there are only two sequences available and information on what pseudo-random sequence is used for a certain demodulation reference signal to a certain terminal is provided in the scheduling assignment together with information about the OCC.

In total, this allows for four different reference signals that can be assigned for four different transmissions being transmitted in parallel to different terminals using MU-MIMO. This assumes no spatial multiplexing on a per-link basis – that is, the transmission to each terminal consists of a single layer. One can also combine MU-MIMO with spatial multiplexing by transmitting two layers to a single terminal, in which case the number of multiplexed terminals is reduced correspondingly as the OCC are then used to separate the two layers transmitted to a single terminal. Obviously, the transmission of multiple layers to the same terminal requires the use of multiple receive antennas at the terminal.

Finally, although MU-MIMO does not rely on any terminal knowledge of the other, parallel, transmissions, nothing prevents a terminal from assuming that there are transmissions on the other “layers”, corresponding to the other demodulation reference signals, and applying receiver-side signal processing to suppress the potential interference from those *possibly* present transmissions. Such suppression would then, once again, be based on the presence of multiple receive antennas at the mobile terminal.

10.3.4.2 MU-MIMO Based on CRS

The above-described MU-MIMO transmission is part of transmission modes 8 and 9 and thus became available in LTE release 9, with further extension in release 10. However, already in LTE release 8, MU-MIMO was possible by a minor modification of transmission mode 4 – that is, closed-loop codebook-based beam-forming, leading to *transmission mode 5*. The only difference between transmission modes 4 and 5 is the signaling of an additional power offset between PDSCH and the cell-specific reference signals (CRS).

In general, for transmission modes relying on CRS (as well as when relying on DM-RS) for channel estimation the terminal will use the reference signal as a phase reference but also as a power/amplitude reference for the demodulation of signals transmitted by means of higher-order modulation (16QAM and 64QAM). Thus, for proper demodulation of higher-order modulation, the terminal needs to know the power offset between the CRS and the PDSCH.

The terminal is informed about this power offset by means of higher-layer signaling. However, what is then provided is the offset between CRS power and the overall PDSCH power, including all layers. In the case of spatial multiplexing, the overall PDSCH power has to be divided between the different layers, and it is the relation between the CRS power and the per-layer PDSCH power that is relevant for demodulation.

In the case of pure spatial multiplexing (no MU-MIMO) – that is, transmission modes 3 and 4 – the terminal knows about the number of layers and thus indirectly about the offset between the CRS power and the per-layer PDSCH power.

In the case of MU-MIMO, the total available power will typically also be divided between the transmissions to the different terminals, with less PDSCH power being available for each transmission. However, terminals are not aware of the presence of parallel transmissions to other terminals and are thus not aware of any per-PDSCH power reduction. For this reason, transmission mode 5 includes the explicit signaling of an *additional power offset* of -3 dB to be used by the terminal in addition to the CRS/PDSCH power offset signaled by higher layers.

Transmission mode 5 is limited to single-rank transmission and, in practice, limited to two users being scheduled in parallel, as there is only a single -3 dB offset defined.

Note that the above power-offset signaling is not needed for MU-MIMO based on demodulation reference signals (transmission modes 8 and 9) as, in this case, each transmission has its own set of reference signals. The power per reference signal will thus scale with the number of layers and transmissions, similar to the PDSCH power and the reference-signal to per-layer PDSCH power ratio will remain constant.

10.4 DOWNLINK L1/L2 CONTROL SIGNALING

To support the transmission of downlink and uplink transport channels, there is a need for certain *associated downlink control signaling*. This control signaling is often referred to as *downlink L1/L2 control signaling*, indicating that the corresponding information partly originates from the physical layer (Layer 1) and partly from Layer 2 MAC. Downlink L1/L2 control signaling consists of downlink scheduling assignments, including information required for the terminal to be able to properly receive, demodulate, and decode the DL-SCH²¹ on a component carrier, uplink scheduling grants informing the terminal about the resources and transport format to use for uplink (UL-SCH) transmission, and hybrid-ARQ acknowledgements in response to UL-SCH transmissions. In addition, the downlink control signaling can also be used for the transmission of power-control commands for power control of uplink physical channels, as well as for certain special purposes such as MBSFN notifications.

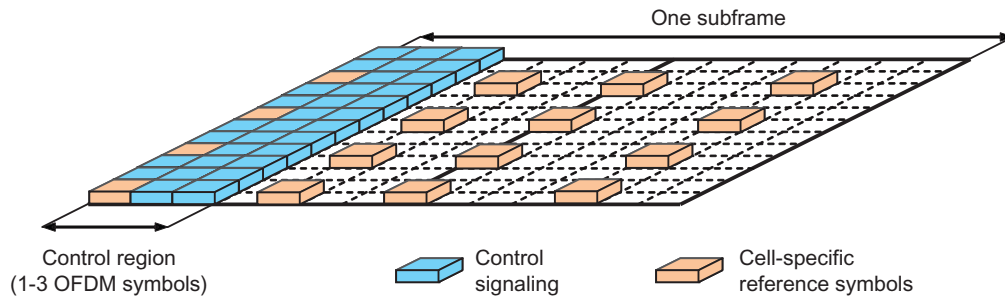
As illustrated in [Figure 10.20](#), the downlink L1/L2 control signaling is transmitted within the first part of each subframe. Thus, each subframe can be said to be divided into a *control region* followed by a *data region*, where the control region corresponds to the part of the subframe in which the L1/L2 control signaling is transmitted. To simplify the overall design, the control region always occupies an integer number of OFDM symbols, more specifically one, two, or three OFDM symbols (for narrow cell bandwidths, 10 resource blocks or less, the control region consists of two, three, or four OFDM symbols to allow for a sufficient amount of control signaling). In the case of carrier aggregation, there is one control region per component carrier.

The size of the control region expressed in number of OFDM symbols, or, equivalently, the start of the data region, can be dynamically varied on a per-subframe basis independently for each component carrier.²² Thus, the amount of radio resources used for control signaling can be dynamically adjusted to match the instantaneous traffic situation. For a small number of users being scheduled in a subframe, the required amount of control signaling is small and a larger part of the subframe can be used for data transmission (larger data region).

The reason for transmitting the control signaling at the beginning of the subframe is to allow for terminals to decode downlink scheduling assignments as early as possible. Processing of the data region – that is, demodulation and decoding of the DL-SCH transmission – can then begin before the end of the subframe. This reduces the delay in the DL-SCH decoding and thus the overall downlink transmission delay. Furthermore, by transmitting the L1/L2 control channel at the beginning of the subframe – that is, by allowing for early decoding of the L1/L2 control information – terminals that

²¹L1/L2 control signaling is also needed for the reception, demodulation, and decoding of the PCH transport channel.

²²The start of the data region is semi-statically configured when using cross-carrier scheduling (see Section 10.4.1).

**FIGURE 10.20**

LTE time–frequency grid illustrating the split of the subframe into (variable-sized) control and data regions.

are not scheduled in the subframe may power down the receiver circuitry for a part of the subframe, allowing for a reduced terminal power consumption.

The downlink L1/L2 control signaling consists of four different physical-channel types:

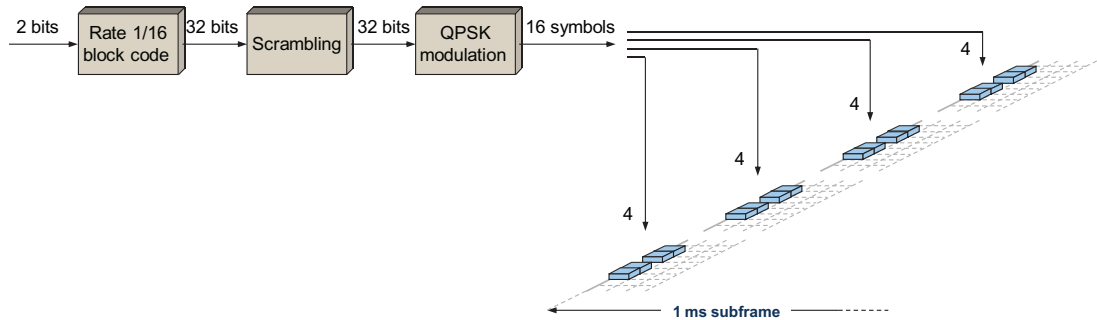
- The *Physical Control Format Indicator Channel* (PCFICH), informing the terminal about the size of the control region (one, two, or three OFDM symbols). There is one and only one PCFICH on each component carrier or, equivalently, in each cell.
- The *Physical Downlink Control Channel* (PDCCH), used to signal downlink scheduling assignments and uplink scheduling grants. Each PDCCH typically carries signaling for a single terminal, but can also be used to address a group of terminals. Multiple PDCCHs can exist in each cell.
- The *Physical Hybrid-ARQ Indicator Channel* (PHICH), used to signal hybrid-ARQ acknowledgements in response to uplink UL-SCH transmissions. Multiple PHICHs can exist in each cell.
- The *Relay Physical Downlink Control Channel* (R-PDCCH), used for relaying. A detailed discussion can be found in Chapter 16 in conjunction with the overall description of relays; at this stage it suffices to note that the R-PDCCH is not transmitted in the control region.

The maximum size of the control region is normally three OFDM symbols (four in the case of narrow cell bandwidths), as mentioned above. However, there are a few exceptions to this rule. When operating in TDD mode, the control region in subframes 1 and 6 is restricted to at most two OFDM symbols since, for TDD, the primary synchronization signal (see Chapter 14) occupies the third OFDM symbol in those subframes. Similarly, for MBSFN subframes (see Chapter 9), the control region is restricted to a maximum of two OFDM symbols.

10.4.1 Physical Control Format Indicator Channel

The PCFICH indicates the size of the control region in terms of the number of OFDM symbols – that is, indirectly where in the subframe the data region starts. Correct decoding of the PCFICH information is thus essential. If the PCFICH is incorrectly decoded, the terminal will neither know how to process the control channels nor where the data region starts for the corresponding subframe.²³

²³Theoretically, the terminal could blindly try to decode all possible control channel formats and, from which format that was correctly decoded, deduce the starting position of the data region, but this can be a very complex procedure.

**FIGURE 10.21**

Overview of the PCFICH processing.

The PCFICH consists of two bits of information, corresponding to the three²⁴ control-region sizes of one, two, or three OFDM symbols (two, three or four for narrow bandwidths), which are coded into a 32-bit codeword. The coded bits are scrambled with a cell- and subframe-specific scrambling code to randomize inter-cell interference, QPSK modulated, and mapped to 16 resource elements. As the size of the control region is unknown until the PCFICH is decoded, the PCFICH is always mapped to the first OFDM symbol of each subframe.

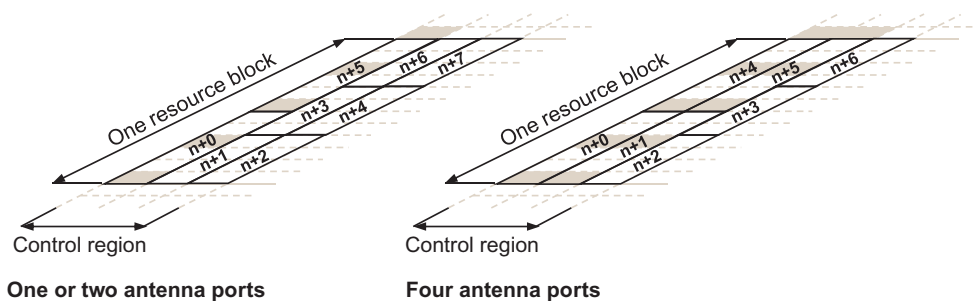
The mapping of the PCFICH to resource elements in the first OFDM symbol in the subframe is done in groups of four resource elements, with the four groups being well separated in frequency to obtain good diversity. Furthermore, to avoid collisions between PCFICH transmissions in neighboring cells, the location of the four groups in the frequency domain depends on the physical-layer cell identity.

The transmission power of the PCFICH is under control of the eNodeB. If necessary for coverage in a certain cell, the power of the PCFICH can be set higher than for other channels by “borrowing” power from, for example, simultaneously transmitted PDCCHs. Obviously, increasing the power of the PCFICH to improve the performance in an interference-limited system depends on the neighboring cells not increasing their transmit power on the interfering resource elements. Otherwise, the interference would increase as much as the signal power, implying no gain in received SIR. However, as the PCFICH-to-resource-element mapping depends on the cell identity, the probability of (partial) collisions with PCFICH in neighboring cells in synchronized networks is reduced, thereby improving the performance of PCFICH power boosting as a tool to control the error rate.

The overall PCFICH processing is illustrated in Figure 10.21.

To describe the mapping of the PCFICH, and L1/L2 control signaling in general, to resource elements, some terminology is required. As mentioned above, the mapping is specified in terms of groups of four resource elements, so-called *resource-element groups*. To each resource-element group, a *symbol quadruplet* consisting of four (QPSK) symbols is mapped. The main motivation behind this, instead of simply mapping the symbols one by one, is the support of transmit diversity. As discussed in Section 10.3, transmit diversity with up to four antenna ports is specified for L1/L2

²⁴The fourth combination is reserved for future use.

**FIGURE 10.22**

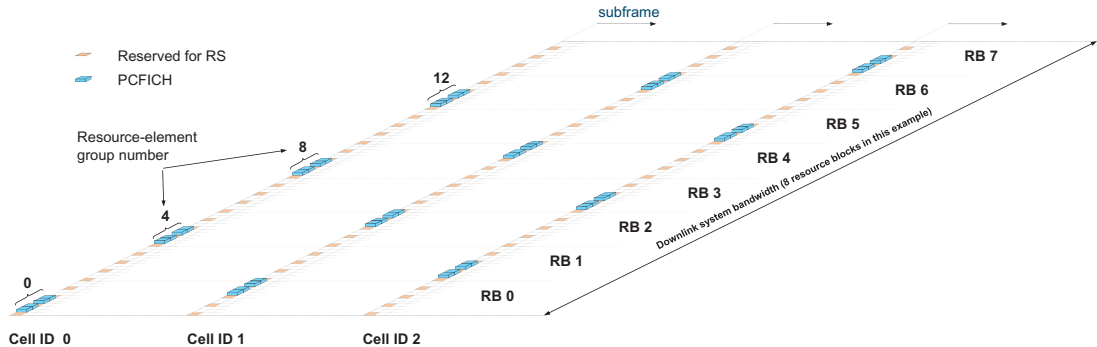
Numbering of resource-element groups in the control region (assuming a size of three OFDM symbols).

control signaling. Transmit diversity for four antenna ports is specified in terms of groups of four symbols (resource elements) and, consequently, the L1/L2 control-channel processing is also defined in terms of symbol quadruplets.

The definition of the resource-element groups assumes that reference symbols corresponding to two antenna ports are present in the first OFDM symbol, regardless of the actual number of antenna ports configured in the cell. This simplifies the definition and reduces the number of different structures to handle. Thus, as illustrated in Figure 10.22, in the first OFDM symbol there are two resource-element groups per resource block, as every third resource element is reserved for reference signals (or non-used resource elements corresponding to reference symbols on the other antenna port). As also illustrated in Figure 10.22, in the second OFDM symbol (if part of the control region) there are two or three resource-element groups depending on the number of antenna ports configured. Finally, in the third OFDM symbol (if part of the control region) there are always three resource-element groups per resource block. Figure 10.22 also illustrates how resource-element groups are numbered in a time-first manner within the size of the control region.

Returning to the PCFICH, four resource-element groups are used for the transmission of the 16 QPSK symbols. To obtain good frequency diversity the resource-element groups should be well spread in frequency and cover the full downlink cell bandwidth. Therefore, the four resource-element groups are separated by one-fourth of the downlink cell bandwidth in the frequency domain, with the starting position given by physical-layer cell identity. This is illustrated in Figure 10.23, where the PCFICH mapping to the first OFDM symbol in a subframe is shown for three different physical-layer cell identities in the case of a downlink cell bandwidth of eight resource blocks. As seen in the figure, the PCFICH mapping depends on the physical-layer cell identity to reduce the risk of inter-cell PCFICH collisions. The cell-specific shifts of the reference symbols, described in Section 10.2.1, are also seen in the figure.

For terminals capable of carrier aggregation, there is one PCFICH per component carrier. Independent signaling of the control-region size on the different component carriers is used, implying that the control region may be of different size on different component carriers. Hence, in principle the terminal needs to receive the PCFICH on each of the component carriers it is scheduled upon. Furthermore, as different component carriers may have different physical-layer cell identities, the location and scrambling may differ across component carriers.

**FIGURE 10.23**

Example of PCFICH mapping in the first OFDM symbol for three different physical-layer cell identities.

If cross-carrier scheduling is used (see Section 10.4.6) – that is, control signaling related to a certain PDSCH transmission is transmitted on a component carrier other than the PDSCH itself – the terminal needs to know the starting position for the data region on the carrier upon which the PDSCH is transmitted. Using the PCFICH on the component carrier carrying the PDSCH would be possible in principle, although it would increase the probability of incorrectly decoding the PDSCH since there are two PCFICH instances, one for the PDCCH decoding on one component carrier and one for PDSCH reception on the other component carrier. This would be problematic, especially since the main use of cross-carrier scheduling is enhanced support of heterogeneous deployments (see Chapter 13), where some of the component carriers may be subject to strong interference. Therefore, for cross-carrier scheduled transmissions, the start of the data region is not obtained from the PCFICH on that component carrier, but is configured on a semi-static basis. The semi-statically configured value may differ from the value signaled on the PCFICH on the component carrier carrying the PDSCH transmission.

10.4.2 Physical Hybrid-ARQ Indicator Channel

The PHICH is used for transmission of hybrid-ARQ acknowledgements in response to UL-SCH transmission. There is one PHICH transmitted per received transport block and TTI – that is, when uplink spatial multiplexing is used on a component carrier, two PHICHs are used to acknowledge the transmission, one per transport block.

For proper operation of the hybrid-ARQ protocol, as discussed in Chapter 12, the error rate of the PHICH should be sufficiently low. The operating point of the PHICH is not specified and is up to the network operator to decide upon, but typically ACK-to-NAK and NAK-to-ACK error rates of the order of 10^{-2} and 10^{-3} – 10^{-4} respectively are targeted. The reason for the asymmetric error rates is that an NAK-to-ACK error would imply loss of a transport block at the MAC level, a loss that has to be recovered by RLC retransmissions with the associated delays, while an ACK-to-NAK error only implies an unnecessary retransmission of an already correctly decoded transport block. To meet these error-rate targets without excessive power, it is beneficial to control the PHICH transmission power as a function of the radio-channel quality of the terminal to which the PHICH is directed. This has influenced the design of the PHICH structure.

In principle, a PHICH could be mapped to a set of resource elements exclusively used by this PHICH. However, taking the dynamic PHICH power setting into account, this could result in significant variations in transmission power between resource elements, which can be challenging from an RF implementation perspective. Therefore, it is preferable to spread each PHICH on multiple resource elements to reduce the power differences while at the same time providing the energy necessary for accurate reception. To fulfill this, a structure where several PHICHs are code multiplexed on to a set of resource elements is used in LTE. The hybrid-ARQ acknowledgement (one single bit of information per transport block) is repeated three times, followed by BPSK modulation on either the I or the Q branch and spreading with a length-four orthogonal sequence. A set of PHICHs transmitted on the same set of resource elements is called a PHICH group, where a PHICH group consists of eight PHICHs in the case of normal cyclic prefix. An individual PHICH can thus be uniquely represented by a single number from which the number of the PHICH group, the number of the orthogonal sequence within the group, and the branch, I or Q, can be derived.

For extended cyclic prefix, which is typically used in time-dispersive environments, the radio channel may not be flat over the frequency spanned by a length-four sequence. A non-flat channel would negatively impact the orthogonality between the sequences. Hence, for extended cyclic prefix, orthogonal sequences of length two are used for spreading, implying only four PHICHs per PHICH group. However, the general structure remains the same as for the normal cyclic prefix.

After forming the composite signal representing the PHICHs in a group, cell-specific scrambling is applied and the 12 scrambled symbols are mapped to three resource-element groups. Similarly to the other L1/L2 control channels, the mapping is described using resource-element groups to be compatible with the transmit diversity schemes defined for LTE.

The overall PHICH processing is illustrated in Figure 10.24.

The requirements on the mapping of PHICH groups to resource elements are similar to those for the PCFICH, namely to obtain good frequency diversity and to avoid collisions between neighboring cells in synchronized networks. Hence, each PHICH group is mapped to three resource-element groups, separated by approximately one-third of the downlink cell bandwidth. In the first OFDM

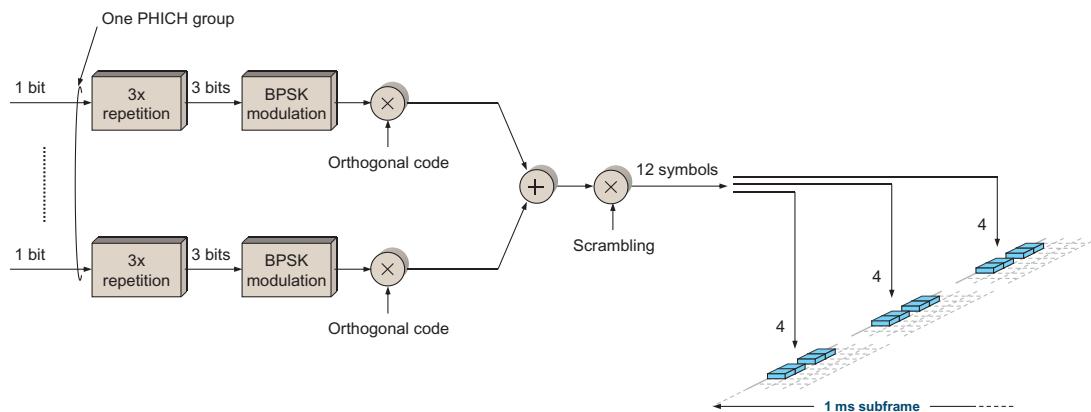


FIGURE 10.24

PHICH structure.

symbol in the control region, resources are first allocated to the PCFICH, the PHICHs are mapped to resource elements not used by the PCFICH, and finally, as will be discussed further below, the PDCCHs are mapped to the remaining resource elements.

Typically, the PHICH is transmitted in the first OFDM symbol only, which allows the terminal to attempt to decode the PHICH even if it failed decoding of the PCFICH. This is advantageous as the error requirements on the PHICH typically are stricter than for PCFICH. However, in some propagation environments, having a PHICH duration of a single OFDM symbol would unnecessarily restrict the coverage. To alleviate this, it is possible to semi-statically configure a PHICH duration of three OFDM symbols. In this case, the control region is three OFDM symbols long in all subframes to fulfill the general principle of separating the control region from data in the time domain only. Obviously, the value transmitted on the PCFICH will be fixed (and can be ignored) in this case. For narrow bandwidths, where the control region can be up to four OFDM symbols long, there is still a need to use the PCFICH to differentiate between a size-three and size-four control region.

As a general principle, LTE transmits the PHICH on the same component carrier that was used for the scheduling grant for 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 monitor the component carriers it monitors for uplink scheduling grants (especially as the PDCCH may override the PHICH to support adaptive retransmissions, as discussed in Chapter 12).

The association between PUSCH transmission in an uplink subframe and the corresponding downlink PHICH resource is discussed in Chapter 12. At this stage, it is sufficient to note that, for TDD configuration zeros, uplink transmissions in multiple subframes may need to be acknowledged in one downlink subframe. Consequently, the amount of resources for PHICH may differ between subframes in the case of TDD.

The PHICH configuration is part of the system information transmitted on the PBCH; one bit indicates whether the duration is one or three OFDM symbols and two bits indicate the amount of resources in the control region reserved for PHICHs, expressed as a fraction of the downlink cell bandwidth in terms of resource blocks. Having the possibility to configure the amount of PHICH resources is useful as the PHICH capacity depends on, for example, whether the network uses multi-user MIMO or not. The PHICH configuration must reside on the PBCH, as it needs to be known in order to properly process the PDCCHs for reception of the part of the system information on the DL-SCH. For TDD, the PHICH information provided on the PBCH is not sufficient for the terminal to know the exact set of resources used by PHICH, as there is also a dependency on the uplink/downlink allocation, provided as part of the system information transmitted on the PDSCH. In order to receive the system information on the DL-SCH, which contains the uplink/downlink allocation, the terminal therefore has to blindly process the PDCCHs under different PHICH configuration hypotheses.

10.4.3 Physical Downlink Control Channel

The PDCCH is used to carry downlink control information (DCI) such as scheduling decisions and power-control commands. More specifically, the DCI includes:

- Downlink scheduling assignments, including PDSCH resource indication, transport format, hybrid-ARQ information, and control information related to spatial multiplexing (if applicable). A downlink scheduling assignment also includes a command for power control of the PUCCH used for transmission of hybrid-ARQ acknowledgements in response to downlink scheduling assignments.

- Uplink scheduling grants, including PUSCH resource indication, transport format, and hybrid-ARQ-related information. An uplink scheduling grant also includes a command for power control of the PUSCH.
- Power-control commands for a set of terminals as a complement to the commands included in the scheduling assignments/grants.

The different types of control information, both between the groups above as well as within the groups, correspond to different DCI message sizes. For example, supporting spatial multiplexing with non-contiguous allocation of resource blocks in the frequency domain requires a larger scheduling message in comparison with an uplink grant allowing for frequency-contiguous allocations only. The DCI is therefore categorized into different *DCI formats*, where a format corresponds to a certain message size and usage. The DCI formats are summarized in [Table 10.4](#), sorted in approximate order of the relative size. The actual message size depends on, among other factors, the cell bandwidth as, for larger bandwidths, a larger number of bits is required to indicate the resource-block allocation. The number of cell-specific reference signals in the cell and whether cross-carrier scheduling is configured or not will also affect the absolute size of most DCI formats. Hence, a given DCI format may have different sizes depending on the overall configuration of the cell. This will be discussed further below; at this stage it suffices to note that formats 0, 1A, 3, and 3A have the same message size.²⁵

Table 10.4 DCI Formats					
Size	Usage				
	Uplink Grant		Downlink Assignment		Power Control
Small	–		1C	Special purpose compact assignment	–
...	0	Single layer	1A	Contiguous allocations only	3, 3A
	–		1B	Codebook-based beam-forming using CRS	–
	–		1D	Multi-user MIMO using CRS	–
	4	Spatial multiplexing	–		–
	–		1	Flexible allocations	–
	–		2A	Open-loop spatial multiplexing using CRS	–
	–		2B	Dual-layer transmission using DM-RS	–
	–		2C	Multi-layer transmission using DM-RS	–
Large	–		2	Closed-loop spatial multiplexing using CRS	–

²⁵The smaller of DCI formats 0 and 1A is padded to ensure the same payload size. Which of 0 and 1A is padded depends on the uplink and downlink cell bandwidths; in the case of identical uplink and downlink bandwidths in a cell there is a single bit of padding in format 0.

One PDCCH carries one DCI message with one of the formats above. As multiple terminals can be scheduled simultaneously, on both downlink and uplink, there must be a possibility to transmit multiple scheduling messages within each subframe. Each scheduling message is transmitted on a separate PDCCH, and consequently there are typically multiple simultaneous PDCCH transmissions within each cell. Furthermore, to support different radio-channel conditions, link adaptation can be used, where the code rate of the PDCCH is selected to match the radio-channel conditions.

For carrier aggregation, scheduling assignments/grants are transmitted individually per component carrier. PDCCH transmission for carrier aggregation is further elaborated upon in Section 10.4.6.

10.4.4 Downlink Scheduling Assignment

Downlink scheduling assignments are valid for the same subframe in which they are transmitted. The scheduling assignments use one of the DCI formats 1, 1A, 1B, 1C, 1D, 2, 2A, 2B, or 2C and the DCI formats used depend on the transmission mode configured (the relation between transmission modes and DCI formats is further described in Section 10.4.9). The reason for supporting multiple formats with different message sizes for the same purpose is to allow for a trade-off in control-signaling overhead and scheduling flexibility. Parts of the contents are the same for the different DCI formats, as seen in Table 10.5, but obviously there are also differences due to the different capabilities.

DCI format 1 is the basic downlink assignment format in the absence of spatial multiplexing (transmission modes 1, 2, and 7). It supports non-contiguous allocations of resource blocks and the full range of modulation-and-coding schemes.

DCI format 1A, also known as the “compact” downlink assignment, supports allocation of frequency-contiguous resource blocks only and can be used in all transmission modes. Contiguous allocations reduce the payload size of the control information with a somewhat reduced flexibility in resource allocations. The full range of modulation-and-coding schemes is supported.

DCI format 1B is used to support codebook-based beam-forming described in Section 10.3.2, with a low control-signaling overhead (transmission mode 6). The content is similar to DCI format 1A with the addition of bits for signaling of the precoding matrix. As codebook-based beam-forming can be used to improve the data rates for cell-edge terminals, it is important to keep the related DCI message size small so as not to unnecessarily limit the coverage.

DCI format 1C is used for various special purposes such as random-access response, paging, and transmission of system information. Common for these applications is simultaneous reception of a relatively small amount of information by *multiple* users. Hence, DCI format 1C supports QPSK only, has no support for hybrid-ARQ retransmissions, and does not support closed-loop spatial multiplexing. Consequently, the message size for DCI format 1C is very small, which is beneficial for coverage and efficient transmission of the type of system messages for which it is intended. Furthermore, as only a small number of resource blocks can be indicated, the size of the corresponding indication field in DCI format 1C is independent of the cell bandwidth.

DCI format 1D is used to support multi-user MIMO (transmission mode 5) scheduling of one codeword with precoder information. To support dynamic sharing of the transmission power between the terminals sharing the same resource block in multi-user MIMO, one bit of power offset information is included in DCI format 1D, as described in Section 10.3.4.2.

DCI format 2 is an extension for DCI format 1 to support closed-loop spatial multiplexing (transmission mode 4). Thus, information about the number of transmission layers and the index of the

Table 10.5 DCI Formats Used for Downlink Scheduling										
Field		DCI Format								
		1	1A	1B	1C	1D	2	2A	2B	2C
Resource information	Carrier indicator	•	•	•		•	•	•	•	•
	Resource block assignment type	0/1	2	2	2'	2	0/1	0/1	0/1	0/1
HARQ process number		•	•	•		•	•	•	•	•
First transport block	MCS	•	•	•	•	•	•	•	•	•
	RV	•	•	•		•	•	•	•	•
	NDI	•	•	•		•	•	•	•	•
Second transport block	MCS						•	•	•	•
	RV						•	•	•	•
	NDI						•	•	•	•
Multi-antenna information	PMI confirmation			•						
	Precoding information			•		•	•	•		
	Transport block swap flag						•	•		
	Power offset					•				
	DM-RS scrambling								•	
	# Layers/DM-RS scrambling									•
Downlink assignment index		•	•	•		•	•	•	•	•
PUCCH power control		•	•	•		•	•	•	•	•
Flag for 0/1A differentiation			•							
Padding (only if needed)		(•)	(•)	(•)		(•)	(•)	(•)	(•)	(•)
Identity		•	•	•	•	•	•	•	•	•

precoder matrix used are jointly encoded in the precoding information field. Some of the fields in DCI format 1 have been duplicated to handle the two transport blocks transmitted in parallel in the case of spatial multiplexing.

DCI format 2A is similar to DCI format 2 except that it supports open-loop spatial multiplexing (transmission mode 3) instead of closed-loop spatial multiplexing. The precoder information field is used to signal the number of transmission layers only, hence the field has a smaller size than in DCI format 2. Furthermore, since DCI format 2A is used for scheduling of multi-layer transmissions only, the precoder information field is only necessary in the case of four transmit antenna ports (a two-antenna-port setup cannot be used for more than two transmission layers).

DCI format 2B was introduced in release 9 in order to support dual-layer spatial multiplexing in combination with beam-forming using DM-RS (transmission mode 8). Since scheduling with DCI format 2B relies on DM-RS, precoding/beam-forming is transparent to the terminal and there is no need to signal a precoder index. The number of layers can be controlled by disabling one of the transport blocks. Two different scrambling sequences for the DM-RS can be used, as described in Section 10.2.2.

DCI format 2C was introduced in release 10 and is used to support spatial multiplexing using DM-RS (transmission mode 9). To some extent, it can be seen as a generalization of format 2B to support spatial multiplexing of up to eight layers. DM-RS scrambling and the number of layers are jointly signaled by a single three-bit field.

Many information fields in the different DCI formats are, as already mentioned, common among several of the formats, while some types of information exist only in certain formats. The information in the DCI formats used for downlink scheduling can be organized into different groups, as shown in Table 10.5, with the fields present varying between the DCI formats. The contents of the different DCI formats are explained in more detail below:

- Resource information, consisting of:
 - Carrier indicator (0 or 3 bit). This field is only present if cross-carrier scheduling is enabled via RRC signaling and is used to indicate the component carrier the downlink control information relates to (see Section 10.4.6).
 - Resource-block allocation. This field indicates the resource blocks on one component carrier upon which the terminal should receive the PDSCH. The size of the field depends on the cell bandwidth and on the DCI format, more specifically on the resource indication type, as discussed in Section 10.4.4.1. Resource allocation types 0 and 1, which are the same size, support non-contiguous resource-block allocations, while resource allocation type 2 has a smaller size but supports contiguous allocations only. DCI format 1C uses a restricted version of type 2 in order to further reduce control signaling overhead.
- Hybrid-ARQ process number (3 bit for FDD, 4 bit for TDD), informing the terminal about the hybrid-ARQ process to use for soft combining. Not present in DCI format 1C.
- For the first (or only) transport block:²⁶
 - Modulation-and-coding scheme (5 bit), used to provide the terminal with information about the modulation scheme, the code rate, and the transport-block size, as described further below. DCI format 1C has a restricted size of this field as only QPSK is supported.

²⁶A transport block can be disabled by setting the modulation-and-coding scheme to zero and the redundancy version to one in the DCI.

- New-data indicator (1 bit), used to clear the soft buffer for initial transmissions. Not present in DCI format 1C as this format does not support hybrid ARQ.
- Redundancy version (2 bit).
- For the second transport block (only present in DCI format supporting spatial multiplexing):
 - Modulation-and-coding scheme (5 bit).
 - New-data indicator (1 bit).
 - Redundancy version (2 bit).
- Multi-antenna information. The different DCI formats are intended for different multi-antenna schemes and which of the fields below that are included depends on the DCI format shown in Table 10.5.
 - PMI confirmation (1 bit), present in format 1B only. Indicates whether the eNodeB uses the (frequency-selective) precoding matrix recommendation from the terminal or if the recommendation is overridden by the information in the PMI field.
 - Precoding information, provides information about the index of the precoding matrix used for the downlink transmission and, indirectly, about the number of transmission layers.
 - Transport block swap flag (1 bit), indicating whether the two codewords should be swapped prior to being fed to the hybrid-ARQ processes. Used for averaging the channel quality between the codewords.
 - Power offset between the PDSCH and cell-specific reference signals used to support dynamic power sharing between multiple terminals for multi-user MIMO.
 - Reference-signal scrambling sequence, used to control the generation of quasi-orthogonal reference-signal sequences, as discussed in Section 10.2.2.
 - Number of layers and reference-signal scrambling sequence (jointly encoded information in release 10).
- Downlink assignment index (2 bit), informing the terminal about the number of downlink transmissions for which a single hybrid-ARQ acknowledgement should be generated according to Section 12.1.3. Present for TDD only.
- Transmit-power control for PUCCH (2 bit). For scheduling of a secondary component carrier in the case of carrier aggregation, these bits are reused as *acknowledgement resource indicator* (ARI) – see Chapter 11.
- DCI format 0/1A indication (1 bit), used to differentiate between DCI formats 1A and 0 as the two formats have the same message size. This field is present in DCI formats 0 and 1A only. DCI formats 3 and 3A, which have the same size, are separated from DCI formats 0 and 1A through the use of a different RNTI.
- Padding; the smaller of DCI formats 0 and 1A is padded to ensure the same payload size irrespective of the uplink and downlink cell bandwidths. Padding is also used to ensure that the DCI size is different for different DCI formats (this is rarely required in practice as the payload sizes are different due to the different amounts of information). Finally, padding is used to avoid certain DCI sizes that may cause ambiguous decoding.²⁷

²⁷For a small set of specific payload sizes, the control signaling may be correctly decoded at an aggregation level other than the one used by the transmitter. Since the PHICH resource is derived from the first CCE used for the PDCCH, this may result in incorrect PHICH being monitored by the terminal. To overcome this, padding is used if necessary to avoid the problematic payload sizes.

- Identity (RNTI) of the terminal for which the PDSCH transmission is intended (16 bit). As described in Section 10.4.8, the identity is not explicitly transmitted but implicitly included in the CRC calculation. There are different RNTIs defined depending on the type of transmission (unicast data transmission, paging, power-control commands, etc.).

10.4.4.1 Signaling of Downlink Resource-Block Allocations

Focusing on the signaling of resource-block allocations, there are three different possibilities, types 0, 1, and 2, as indicated in Table 10.5. Resource-block allocation types 0 and 1 both support non-contiguous allocations of resource blocks in the frequency domain, whereas type 2 supports contiguous allocations only. A natural question is why multiple ways of signaling the resource-block allocations are supported, a question whose answer lies in the number of bits required for the signaling. The most flexible way of indicating the set of resource blocks the terminal is supposed to receive the downlink transmission upon is to include a bitmap with size equal to the number of resource blocks in the cell bandwidth. This would allow for an arbitrary combination of resource blocks to be scheduled for transmission to the terminal but would, unfortunately, also result in a very large bitmap for the larger cell bandwidths. For example, in the case of a downlink cell bandwidth corresponding to 100 resource blocks, the downlink PDCCH would require 100 bits for the bitmap alone, to which the other pieces of information need to be added. Not only would this result in a large control-signaling overhead, but it could also result in downlink coverage problems as more than 100 bits in one OFDM symbol correspond to a data rate exceeding 1.4 Mbit/s. Consequently, there is a need for a resource allocation scheme requiring a smaller number of bits while keeping sufficient allocation flexibility.

In resource allocation type 0, the size of the bitmap has been reduced by pointing not to individual resource blocks in the frequency domain, but to groups of contiguous resource blocks, as shown at the top of Figure 10.25. The size of such a group is determined by the downlink cell bandwidth; for the smallest bandwidths there is only a single resource block in a group, implying that an arbitrary set of resource blocks can be scheduled, whereas for the largest cell bandwidths groups of four resource blocks are used (in the example in Figure 10.25, the cell bandwidth is 25 resource blocks, implying a group size of two resource blocks). Thus, the bitmap for the system with a downlink cell bandwidth of 100 resource blocks is reduced from 100 to 25 bits. A drawback is that the scheduling granularity is reduced; single resource blocks cannot be scheduled for the largest cell bandwidths using allocation type 0.

However, also in large cell bandwidths, frequency resolution of a single resource block is sometimes useful, for example to support small payloads. Resource allocation type 1 addresses this by dividing the total number of resource blocks in the frequency domain into dispersed subsets, as shown in the middle of Figure 10.25. The number of subsets is given from the cell bandwidth with the number of subsets in type 1 being equal to the group size in type 0. Thus, in Figure 10.25, there are two subsets, whereas for a cell bandwidth of 100 resource blocks there would have been four different subsets. Within a subset, a bitmap indicates the resource blocks in the frequency domain upon which the downlink transmission occurs.

To inform the terminal whether resource allocation type 0 or 1 is used, the resource allocation field includes a flag for this purpose, denoted “type” to the left in Figure 10.25. For type 0, the only additional information is the bitmap discussed above. For type 1, on the other hand, in addition to the bitmap

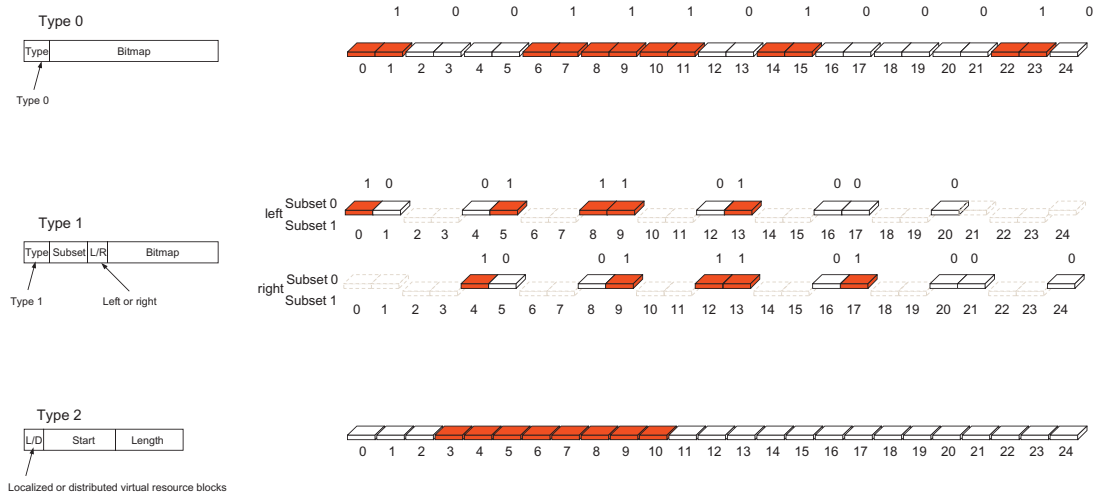
**FIGURE 10.25**

Illustration of resource-block allocation types (cell bandwidth corresponding to 25 resource blocks used in this example).

itself, information about the subset for which the bitmap relates is also required. As one of the requirements in the design of resource allocation type 1 was to maintain the same number of bits in the allocation as for type 0 without adding unnecessary overhead,²⁸ the bitmap in resource allocation type 1 is smaller than in type 0 to allow for the signaling of the subset number. However, a consequence of a smaller bitmap is that not all resource blocks in the subset can be addressed simultaneously. To be able to address all resources with the bitmap, there is a flag indicating whether the bitmap relates to the “left” or “right” part of the resource blocks, as depicted in the middle part of Figure 10.25.

Unlike the other two types of resource-block allocation signaling, type 2 does not rely on a bitmap. Instead, it encodes the resource allocation as a start position and length of the resource-block allocation. Thus, it does not support arbitrary allocations of resource blocks but only frequency-contiguous allocations, thereby reducing the number of bits required for signaling the resource-block allocation. The number of bits required for resource-signaling type 2 compared to type 0 or 1 is shown in Figure 10.26 and, as shown, the difference is fairly large for the larger cell bandwidths.

All three resource-allocation types refer to *virtual* resource blocks (see Section 10.1.1.8 for a discussion of resource-block types). For resource-allocation types 0 and 1, the virtual resource blocks are of localized type and the virtual resource blocks are directly mapped to physical resource blocks. For resource-allocation type 2, on the other hand, both localized and distributed virtual resource blocks are supported. One bit in the resource allocation field indicates whether the allocation signaling refers to localized or distributed resource blocks.

²⁸ Allowing different sizes would result in an increase in the number of blind decoding attempts required in the terminal.

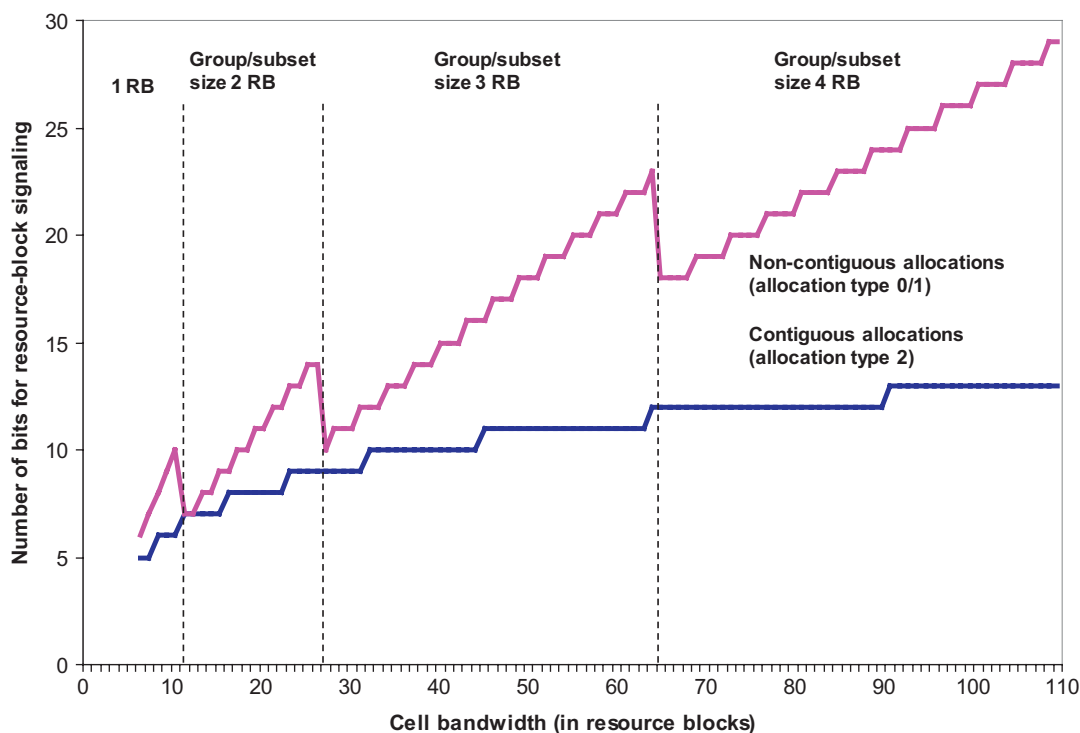


FIGURE 10.26

Number of bits used for downlink resource-allocation signaling for downlink allocation types 0/1 and 2.

10.4.4.2 Signaling of Transport-Block Sizes

Proper reception of a downlink transmission requires, in addition to the set of resource blocks, knowledge about the modulation scheme and the transport-block size, information (indirectly) provided by a five-bit field in the different DCI formats. Of the 32 combinations, 29 are used to signal the modulation-and-coding scheme whereas three are reserved, the purpose of which is described later. Together, the modulation-and-coding scheme and the number of resource blocks assigned provide the transport-block size on the DL-SCH. Thus, the possible transport-block sizes can be described as a table with 29 rows and 110 columns, one column for each number of resource blocks possible to transmit upon (the number of columns follows from the maximum downlink component carrier bandwidth of 110 resource blocks).

Each modulation-and-coding scheme represents a particular combination of modulation scheme and channel-coding rate or, equivalently, a certain spectral efficiency measured in the number of information bits per modulation symbol. Although the 29-by-110 table of transport-block sizes in principle could be filled directly from the modulation-and-coding scheme and the number of resource blocks, this would result in arbitrary transport-block sizes, which is not desirable. First, as all the higher-layer protocol layers are byte aligned, the resulting transport-block sizes should be an integer

number of bytes. Secondly, common payloads, for example RRC signaling messages and VoIP, should be possible to transmit without padding. Aligning with the QPP interleaver sizes is also beneficial, as this would avoid the use of filler bits (see Section 10.1.1.3). Finally, the same transport-block size should ideally appear for several different resource-block allocations, as this allows the number of resource blocks to be changed between retransmission attempts, providing increased scheduling flexibility. Therefore, a “mother table” of transport-block sizes is first defined, fulfilling the requirements above. Each entry in the 29-by-110 table is picked from the mother table such that the resulting spectral efficiency is as close as possible to the spectral efficiency of the signaled modulation-and-coding scheme. The mother table spans the full range of transport-block sizes possible, with an approximately constant worst-case padding.

From a simplicity perspective, it is desirable if the transport-block sizes do not vary with the configuration of the system. The set of transport-block sizes is therefore independent of the actual number of antenna ports and the size of the control region.²⁹ The design of the table assumes a control region of three OFDM symbols and two antenna ports, the “reference configuration”. If the actual configuration is different, the resulting code rate for the DL-SCH will be slightly different as a result of the rate-matching procedure. However, the difference is small and of no practical concern. Also, if the actual size of the control region is smaller than the three-symbol assumption in the reference configuration, the spectral efficiencies will be somewhat smaller than the range indicated by the modulation-and-coding scheme signaled as part of the DCI. Thus, information about the modulation scheme used is obtained directly from the modulation-and-coding scheme, whereas the exact code rate and rate matching is obtained from the implicitly signaled transport-block size together with the number of resource elements used for DL-SCH transmission.

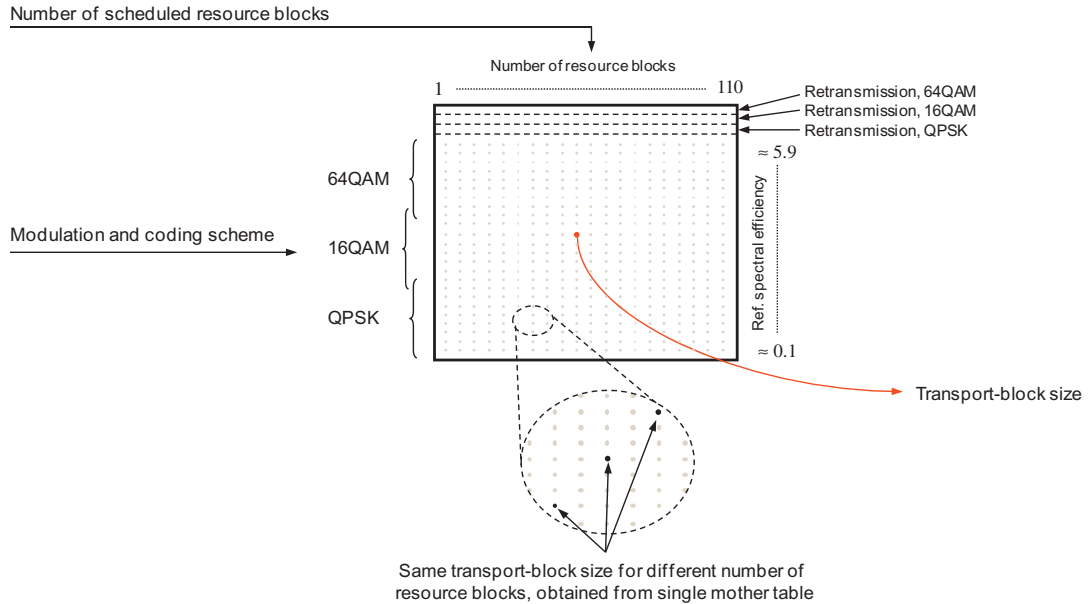
For bandwidths smaller than the maximum of 110 resource blocks, a subset of the table is used. More specifically, in case of a cell bandwidth of N resource blocks, the first N columns of the table are used. Also, in the case of spatial multiplexing, a single transport block can be mapped to up to four layers. To support the higher data rates this facilitates, the set of supported transport-block sizes needs to be extended beyond what is possible in the absence of spatial multiplexing. The additional entries are in principle obtained by multiplying the sizes with the number of layers to which a transport block is mapped and adjusting the result to match the QPP interleaver size.

The 29 combinations of modulation-and-coding schemes each represent a reference spectral efficiency in the approximate range of 0.1–5.9 bits/s/symbol.³⁰ There is some overlap in the combinations in the sense that some of the 29 combinations represent the same spectral efficiency. The reason is that the best combination for realizing a specific spectral efficiency depends on the channel properties; sometimes higher-order modulation with a low code rate is preferable over lower-order modulation with a higher code rate, sometimes the opposite is true. With the overlap, the eNodeB can select the best combination, given the propagation scenario. As a consequence of the overlap, two of the rows in the 29-by-110 table are duplicates and result in the same spectral efficiency but with different modulation schemes, and there are only 27 unique rows of transport-block sizes.

Returning to the three reserved combinations in the modulation-and-coding field mentioned at the beginning, those entries can be used for retransmissions only. In the case of a retransmission, the

²⁹For DwPTS, the transport-block size is scaled by a factor of 0.75 compared to the values found in the table, motivated by the DwPTS having a shorter duration than a normal subframe.

³⁰The exact values vary slightly with the number of resource blocks allocated due to rounding.

**FIGURE 10.27**

Computing the transport-block size.

transport-block size is, by definition, unchanged and fundamentally there is no need to signal this piece of information. Instead, the three reserved values represent the modulation scheme, QPSK, 16QAM or 64QAM, which allows the scheduler to use an (almost) arbitrary combination of resource blocks for the retransmission. Obviously, using any of the three reserved combinations assumes that the terminal properly received the control signaling for the initial transmission; if this is not the case, the retransmission should explicitly indicate the transport-block size.

The derivation of the transport-block size from the modulation-and-coding scheme and the number of scheduled resource blocks is illustrated in Figure 10.27.

10.4.5 Uplink Scheduling Grants

Uplink scheduling grants use one of DCI formats 0 or 4, where DCI format 4 was added in release 10 to support uplink spatial multiplexing. The basic resource-allocation scheme for the uplink is single-cluster allocations where the resource blocks are contiguous in the frequency domain, although release 10 added support for multi-cluster transmissions of up to two clusters on a single component carrier.

DCI format 0 is used for scheduling uplink transmissions not using spatial multiplexing on one component carrier. It has the same size control-signaling message as the “compact” downlink assignment (DCI format 1A). A flag in the message is used to inform the terminal whether the message is an uplink scheduling grant (DCI format 0) or a downlink scheduling assignment (DCI format 1A).

Table 10.6 DCI Formats for Uplink Scheduling Grants			
Field		DCI Format	
		0	4
Resource information	Carrier indicator	•	•
	Multi-cluster flag	•	•
	Resource block assignment	0	1
First transport block	MCS/RV	•	•
	NDI	•	•
Second transport block	MCS/RV		•
	NDI		•
DMRS phase rotation		•	•
Precoding information			•
CSI request		•	•
SRS request			•
Uplink index/DAI		•	•
PUSCH power control		•	•
Flag for 0/1A differentiation		•	
Padding (only if needed)		(•)	(•)
Identity		•	•

DCI format 4 is used for uplink transmissions using spatial multiplexing on one component carrier. Consequently, the size of DCI format 4 is larger than that of DCI format 0, as additional information fields are required.

Many information fields are common to the two DCI formats, but there are also differences, as shown in Table 10.6. The contents of the different DCI formats are explained in more detail below:

- Resource information, consisting of:
 - Carrier indicator (0 or 3 bit). This field is only present if cross-carrier scheduling is enabled via RRC signaling and is used to indicate the uplink component carrier the grant relates to (see Section 10.4.6).
 - Multi-cluster flag (1 bit), indicating whether one or two clusters of resource blocks are used for the uplink transmission. This flag is not present in releases prior to release 10. In previous releases, the downlink bandwidth is, in practice, always at least as large as the uplink bandwidth, implying that one padding bit is used for DCI format 0 in those releases to align with the size of format 1A. The padding bit could therefore be replaced by the multi-cluster flag in release 10 without sacrificing backwards compatibility. In DCI format 4, supported in release 10 only, the multi-cluster flag is always present.
 - Resource-block allocation, including hopping indication. This field indicates the resource blocks upon which the terminal should transmit the PUSCH using uplink resource-allocation type 0 (DCI format 0) or type 1 (DCI format 4), as described in Section 10.4.5.1. The size of

the field depends on the cell bandwidth. For single-cluster allocations, uplink frequency hopping, as described in Chapter 11, can be applied to the uplink PUSCH transmission.

- For the first (or only) transport block:
 - Modulation-and-coding scheme including redundancy version (5 bit), used to provide the terminal with information about the modulation scheme, the code rate, and the transport-block size. The signaling of the transport-block size uses the same transport-block table as for the downlink – that is, the modulation-and-coding scheme together with the number of scheduled resource blocks provides the transport-block size. However, as the support of 64QAM in the uplink is not mandatory for all terminals, terminals not capable of 64QAM use 16QAM when 64QAM is indicated in the modulation-and-coding field. The use of the three reserved combinations is slightly different than for the downlink; the three reserved values are used for implicit signaling of the redundancy version, as described below. A transport block can be disabled by signaling a specific combination of modulation-and-coding scheme and number of resource blocks. Disabling one transport block is used when retransmitting a single transport block only.
 - New-data indicator (1 bit), used to indicate to the terminal whether transmission of a new transport block or retransmission of the previous transport block is granted.
- For the second transport block (DCI format 4 only):
 - Modulation-and-coding scheme including redundancy version (5 bit).
 - New-data indicator (1 bit).
- Phase rotation of the uplink demodulation reference signal³¹ (3 bit), used to support multi-user MIMO, as described in Chapter 11. By assigning different reference-signal phase rotations to terminals scheduled on the same time–frequency resources, the eNodeB can estimate the uplink channel response from each terminal and suppress the inter-terminal interference by the appropriate processing.
- Precoding information, used to signal the precoder to use for the uplink transmission.
- Channel-state request flag (1 or 2 bit). The network can explicitly request an aperiodic channel-state report to be transmitted on the UL-SCH by setting this bit(s) in the uplink grant. In the case of carrier aggregation, two bits are used to indicate which downlink component carrier the CSI should be reported for (see Chapter 12).
- SRS request (2 bit), used to trigger aperiodic sounding using one of up to three preconfigured settings, as discussed in Chapter 11.
- Uplink index/DAI (2 bit). This field is present only when operating in TDD. For uplink–downlink configuration 0 (uplink-heavy configuration), it is used as an uplink index to signal for which uplink subframe(s) the grant is valid, as described in Chapter 13. For other uplink–downlink configurations, it is used as downlink assignment index to indicate the number of downlink transmissions the eNodeB expects hybrid-ARQ acknowledgement for.
- Transmit-power control for PUSCH (2 bit).
- DCI format 0/1A indication (1 bit), used to differentiate between DCI formats 1A and 0 as the two formats have the same message size. This field is present in DCI formats 0 and 1A only.

³¹In release 10, this field can also be used to select the orthogonal cover sequence used for uplink reference signals (see Chapter 11 for details).

- Padding; the smaller of DCI formats 0 and 1A is padded to ensure the same payload size irrespective of the uplink and downlink cell bandwidths. Padding is also used to ensure that the DCI size is different for different DCI formats (this is rarely required in practice as the payload sizes are different due to the different amounts of information). Finally, padding is used to avoid certain DCI sizes that may cause ambiguous decoding.
- Identity (RNTI) of the terminal for which the grant is intended (16 bit). As described in Section 10.4.8, the identity is not explicitly transmitted but implicitly included in the CRC calculation.

There is no explicit signaling of the redundancy version in the uplink scheduling grants. This is motivated by the use of a synchronous hybrid-ARQ protocol in the uplink; retransmissions are normally triggered by a negative acknowledgement on the PHICH and not explicitly scheduled as for downlink data transmissions. However, as described in Chapter 12, there is a possibility to explicitly schedule retransmissions. This is useful in a situation where the network will explicitly move the retransmission in the frequency domain by using the PDCCH instead of the PHICH. Three values of the modulation-and-coding field are reserved to mean redundancy versions one, two, and three. If one of those values is signaled, the terminal should assume that the same modulation and coding as the original transmission is used. The remaining entries are used to signal the modulation-and-coding scheme to use and also imply that redundancy version zero should be used. The difference in usage of the reserved values compared to the downlink scheduling assignments implies that the modulation scheme, unlike the downlink case, cannot change between uplink (re)transmission attempts.

10.4.5.1 Signaling of Uplink Resource-Block Allocations

The basic uplink resource-allocation scheme is single-cluster allocations – that is, allocations contiguous in the frequency domain – but release 10 also provides the possibility for multi-cluster uplink transmissions.

Single-cluster allocations use uplink resource-allocation type 0, which is identical to downlink resource allocation type 2 described in Section 10.4.4.1 except that the single-bit flag indicating localized/distributed transmission is replaced by a single-bit hopping flag. The resource allocation field in the DCI provides the set of virtual resource blocks to use for uplink transmission. The set of physical resource blocks to use in the two slots of a subframe is controlled by the hopping flag, as described in Chapter 11.

Multi-cluster allocations with up to two clusters were introduced in release 10, using uplink resource-allocation type 1. In resource-allocation type 1, the starting and ending positions of two clusters of frequency-contiguous resource blocks are encoded into an index. Uplink resource-allocation type 1 does not support frequency hopping (diversity is achieved through the use of two clusters instead). Indicating two clusters of resources naturally requires additional bits compared to the single-cluster case. At the same time, the total number of bits used for resource-allocation type 1 should be identical to that of type 0 (this is similar to the situation for the downlink allocation types 0 and 1; without aligning the sizes a new DCI format with a corresponding negative impact on the number of blind decodings is necessary). Since frequency hopping is not supported for allocation type 1, the bit otherwise used for the hopping flag can be reused for extending the resource-allocation field. However, despite the extension of the resource-allocation field by one bit, the number of bits is not sufficient to provide a single-resource-block resolution in the two clusters for all bandwidths. Instead, similar to downlink resource allocation type 0, groups of resource blocks are used and the starting and ending positions of the two clusters are

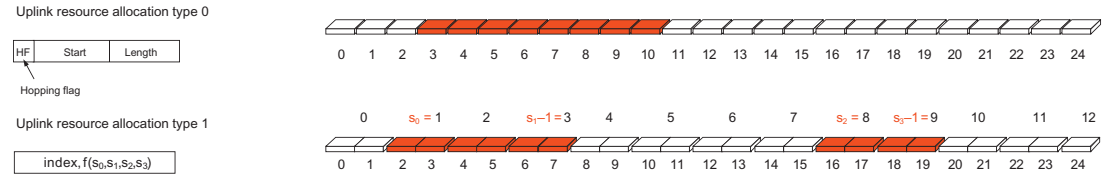


FIGURE 10.28

Illustration of uplink resource-block allocation types (uplink bandwidth corresponding to 25 resource blocks used in this example).

given in terms of group numbers. The size of such a group is determined by the uplink carrier bandwidth in a similar way as for the downlink; for the smallest bandwidths there is only a single resource block in a group, implying that an arbitrary (as long as the limit of at most two clusters is observed) set of resource blocks can be scheduled, whereas for the largest cell bandwidths groups of four resource blocks are used (in the example in Figure 10.28, the cell bandwidth is 25 resource blocks, implying a group size of two resource blocks).

10.4.6 Carrier Aggregation and Cross-Carrier Scheduling

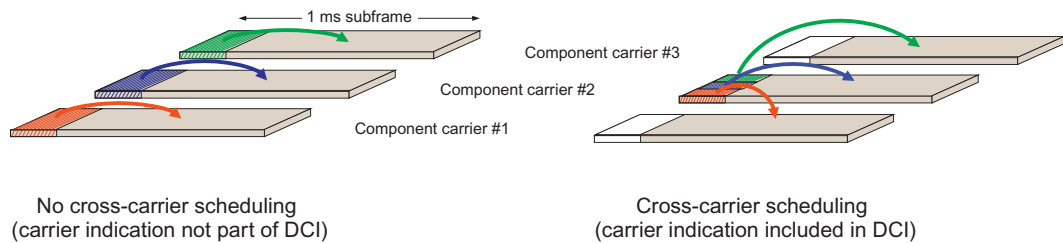
Carrier aggregation, where a terminal receives or transmits on multiple component carriers, is an integral part of LTE from release 10 onwards. Clearly, for a terminal supporting multiple component carriers, the terminal needs to know to which component carrier a certain DCI relates. This information can either be implicit or explicit, depending on whether cross-carrier scheduling is used or not, as illustrated in Figure 10.29. Enabling cross-carrier scheduling is done individually via RRC signaling on a per-terminal and per-component-carrier basis.

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

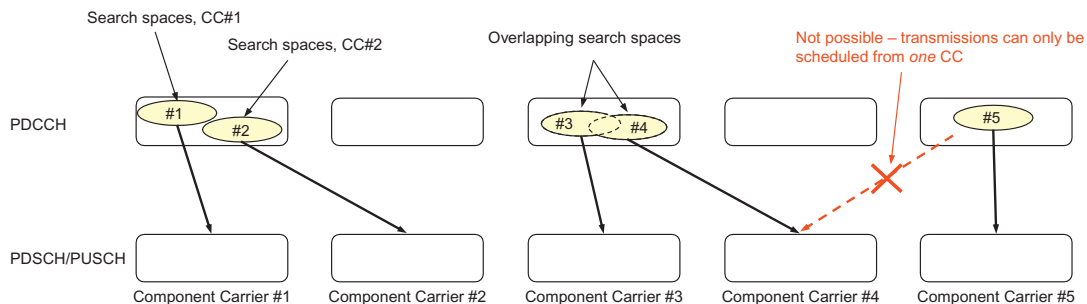
In the presence of cross-carrier scheduling, where downlink PDSCH or uplink PUSCH is transmitted on an (associated) component carrier other than the PDCCH, the carrier indicator in the PDCCH provides information about the component carrier used for the PDSCH or PUSCH.

Whether cross-carrier scheduling is used or not is, as already mentioned, configured using higher-layer signaling; if cross-carrier scheduling is not configured then no carrier indication field is included in the DCI. Thus, most of the DCI formats come in two “flavors”, with and without the carrier indication field, and which “flavor” the terminal is supposed to monitor is determined by enabling/disabling support for cross-carrier scheduling.

To signal which component carrier a grant relates to, the component carriers are numbered. The primary component carrier is always given the number zero, while the different secondary component carriers are assigned a unique number each through UE-specific RRC signaling. Hence, even if the

**FIGURE 10.29**

Cross-carrier scheduling.

**FIGURE 10.30**

Example of scheduling of multiple component carriers (search spaces are discussed in Section 10.4.9).

terminal and the eNodeB may have different understandings of the component carrier numbering during a brief period of reconfiguration, at least transmissions on the primary component carrier can be scheduled.

Irrespective of whether cross-carrier scheduling is used or not, PDSCH/PUSCH on a component carrier can only be scheduled from *one* component. Thus, for each PDSCH/PUSCH component carrier there is an associated component carrier, configured via UE-specific RRC signaling, where the corresponding DCI can be transmitted. Figure 10.30 illustrates one example hereof, where PDSCH/PUSCH transmissions on component carrier 1 are scheduled using PDCCHs transmitted on component carrier 1. In this case, as cross-carrier scheduling is not used, there is no carrier indicator in the corresponding DCI formats. PDSCH/PUSCH transmissions on component carrier 2 are cross-carrier scheduled from PDCCHs transmitted on component carrier 1. Hence, the DCI formats in the UE-specific search space for component carrier 2 include the carrier indicator.

Note also, as transmissions on a component carrier can be scheduled by PDCCHs on *one* component carrier only, component carrier 4 cannot be scheduled by PDCCHs on component carrier 5 as the semi-static association between the component carriers used for PDCCH transmission and the actual data transmission has associated data on component carrier 4, with PDCCHs on component carrier 3 in this example.

10.4.7 Power-Control Commands

As a complement to the power-control commands provided as part of the downlink scheduling assignments and the uplink scheduling grants, there is the potential to transmit a power-control command using DCI formats 3 (single-bit command per terminal) or 3A (two-bit command per terminal). The main motivation for DCI format 3/3A is to support power control for semi-persistent scheduling. The power-control message is directed to a group of terminals using an RNTI specific for that group. Each terminal can be allocated two power-control RNTIs, one for PUCCH power control and the other for PUSCH power control. Although the power-control RNTIs are common to a group of terminals, each terminal is informed through RRC signaling which bit(s) in the DCI message it should follow. No carrier indicator is used for formats 3 and 3A.

10.4.8 PDCCH Processing

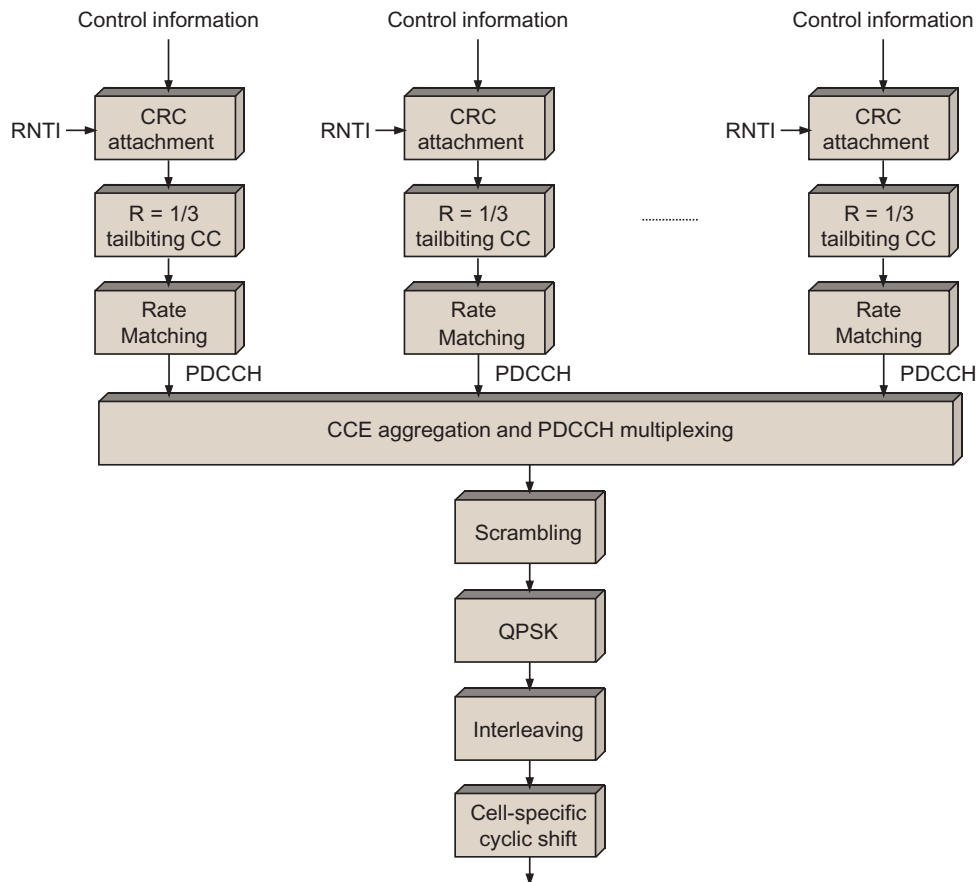
Having discussed the contents of L1/L2 control signaling in terms of the different DCI formats, the transmission of the DCI message on a PDCCH can be described. The processing of downlink control signaling is illustrated in Figure 10.31. A CRC is attached to each DCI message payload. The identity of the terminal (or terminals) addressed – that is, the RNTI – is included in the CRC calculation and not explicitly transmitted. Depending on the purpose of the DCI message (unicast data transmission, power-control command, random-access response, etc.), different RNTIs are used; for normal unicast data transmission, the terminal-specific C-RNTI is used.

Upon reception of DCI, the terminal will check the CRC using its set of assigned RNTIs. If the CRC checks, the message is declared to be correctly received and intended for the terminal. Thus, the identity of the terminal that is supposed to receive the DCI message is implicitly encoded in the CRC and not explicitly transmitted. This reduces the amount of bits necessary to transmit on the PDCCH as, from a terminal point of view, there is no difference between a corrupt message whose CRC will not check and a message intended for another terminal.

After CRC attachment, the bits are coded with a rate-1/3 tail-biting convolutional code and rate-matched to fit the amount of resources used for PDCCH transmission. Tail-biting convolutional coding is similar to conventional convolutional coding with the exception that no tail bits are used. Instead, the convolutional encoder is initialized with the last bits of the message prior to the encoding process. Thus, the starting and ending states in the trellis in an MLSE (Viterbi) decoder are identical.

After the PDCCHs to be transmitted in a given subframe have been allocated to the desired resource elements (the details of which are given below), the sequence of bits corresponding to all the PDCCH resource elements to be transmitted in the subframe, including the unused resource elements, is scrambled by a cell- and subframe-specific scrambling sequence to randomize inter-cell interference, followed by QPSK modulation and mapping to resource elements.

To allow for simple yet efficient processing of the control channels in the terminal, the mapping of PDCCHs to resource elements is subject to a certain structure. This structure is based on so-called *Control-Channel Elements* (CCEs), which in essence is a convenient name for a set of 36 useful resource elements (nine resource-element groups as defined in Section 10.4.1). The number of CCEs, one, two, four, or eight, required for a certain PDCCH depends on the payload size of the control information (DCI payload) and the channel-coding rate. This is used to realize link adaptation for the PDCCH; if the channel conditions for the terminal to which the PDCCH is intended are

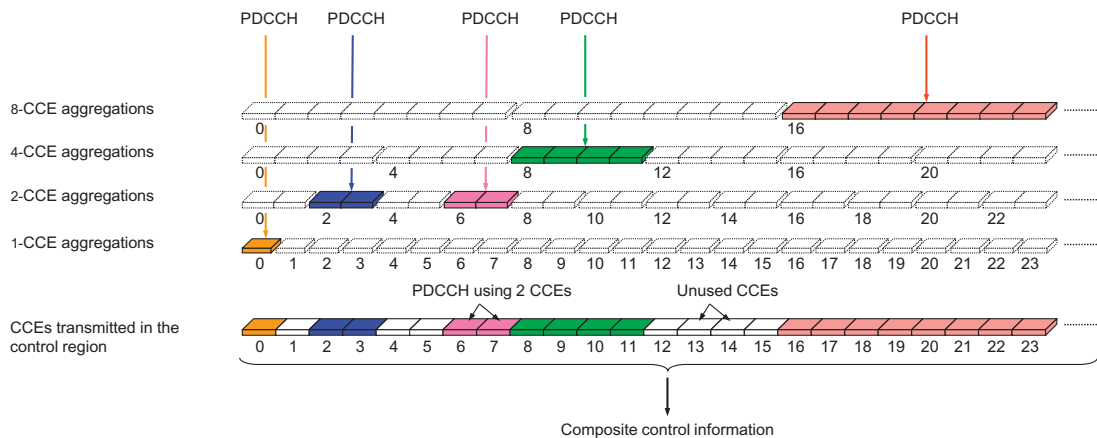
**FIGURE 10.31**

Processing of L1/L2 control signaling.

disadvantageous, a larger number of CCEs needs to be used compared to the case of advantageous channel conditions. The number of CCEs used for a PDCCH is also referred to as the aggregation level.

The number of CCEs available for PDCCHs depends on the size of the control region, the cell bandwidth, the number of downlink antenna ports, and the amount of resources occupied by PHICH. The size of the control region can vary dynamically from subframe to subframe as indicated by the PCFICH, whereas the other quantities are semi-statically configured. The CCEs available for PDCCH transmission can be numbered from zero and upward, as illustrated in Figure 10.32. A specific PDCCH can thus be identified by the numbers of the corresponding CCEs in the control region.

As the number of CCEs for each of the PDCCHs may vary and is not signaled, the terminal has to blindly determine the number of CCEs used for the PDCCH it is addressed upon. To reduce the complexity of this process somewhat, certain restrictions on the aggregation of contiguous CCEs have been specified. For example, an aggregation of eight CCEs can only start on CCE numbers evenly

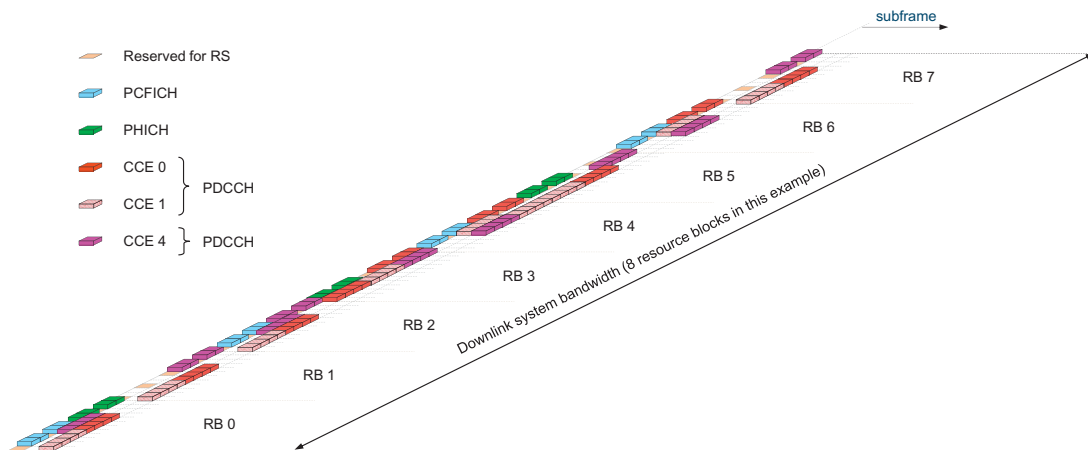
**FIGURE 10.32**

CCE aggregation and PDCCH multiplexing.

divisible by 8, as illustrated in Figure 10.32. The same principle is applied to the other aggregation levels. Furthermore, some combinations of DCI formats and CCE aggregations that result in excessively high channel-coding rates are removed.

The sequence of CCEs should match the amount of resources available for PDCCH transmission in a given subframe – that is, the number of CCEs varies according to the value transmitted on the PCFICH. In many cases, not all the PDCCHs that can be transmitted in the control region are used. Nevertheless, unused PDCCHs are part of the interleaving and mapping process in the same way as any other PDCCH. At the terminal, the CRC will not check for those “dummy” PDCCHs. Preferably, the transmission power is set to zero for those unused PDCCHs; the power can be used by other control channels.

The mapping of the modulated composite control information is, for the same reason as for the other control channels, described in terms of symbol quadruplets being mapped to resource-element groups. Thus, the first step of the mapping stage is to group the QPSK symbols into symbol quadruplets, each consisting of four consecutive QPSK symbols. In principle, the sequence of quadruplets could be mapped directly to the resource elements in sequential order. However, this would not exploit all the frequency diversity available in the channel and diversity is important for good performance. Furthermore, if the same CCE-to-resource-element mapping is used in all neighboring cells, a given PDCCH will persistently collide with one and the same PDCCH in the neighboring cells assuming a fixed PDCCH format and inter-cell synchronization. In practice, the number of CCEs per PDCCH varies in the cell as a function of the scheduling decisions, which gives some randomization to the interference, but further randomization is desirable to obtain a robust control-channel design. Therefore, the sequence of quadruplets is first interleaved using a block interleaver to allow exploitation of the frequency diversity, followed by a cell-specific cyclic shift to randomize the interference between neighboring cells. The output from the cell-specific shift is mapped to resource-element groups in a time-first manner, as illustrated in Figure 10.22, skipping resource-element groups used for PCFICH and PHICH. Time-first mapping

**FIGURE 10.33**

Example of mapping of PCFICH, PHICH, and PDCCH.

preserves the interleaving properties; with frequency-first over multiple OFDM symbols, resource-element groups that are spread far apart after the interleaving process may end up close in frequency, although on different OFDM symbols.

The interleaving operation described above, in addition to enabling exploitation of the frequency diversity and randomizing the inter-cell interference, also serves the purpose of ensuring that each CCE spans virtually all the OFDM symbols in the control region. This is beneficial for coverage as it allows flexible power balancing between the PDCCHs to ensure good performance for each of the terminals addressed. In principle, the energy available in the OFDM symbols in the control region can be balanced arbitrarily between the PDCCHs. The alternative of restricting each PDCCH to a single OFDM symbol would imply that power cannot be shared between PDCCHs in different OFDM symbols.

Similarly to the PCFICH, the transmission power of each PDCCH is under the control of the eNodeB. Power adjustments can therefore be used as a complementary link adaptation mechanism in addition to adjusting the code rate. Relying on power adjustments alone might seem a tempting solution but, although possible in principle, it can result in relatively large power differences between resource elements. This may have implications on the RF implementation and may violate the out-of-band emission masks specified. Hence, to keep the power differences between the resource elements reasonable, link adaptation through adjusting the channel code rate, or equivalently the number of CCEs aggregated for a PDCCH, is necessary. The two mechanisms for link adaptation, power adjustments and different code rates, complement each other.

To summarize and to illustrate the mapping of PDCCHs to resource elements in the control region, consider the example shown in Figure 10.33. In this example, the size of the control region in the subframe considered equals three OFDM symbols. Two downlink antenna ports are configured (but, as explained above, the mapping would be identical in the case of a single antenna port). One PHICH group is configured and three resource-element groups are therefore used by the PHICHs. The cell identity is assumed to be identical to zero in this case.

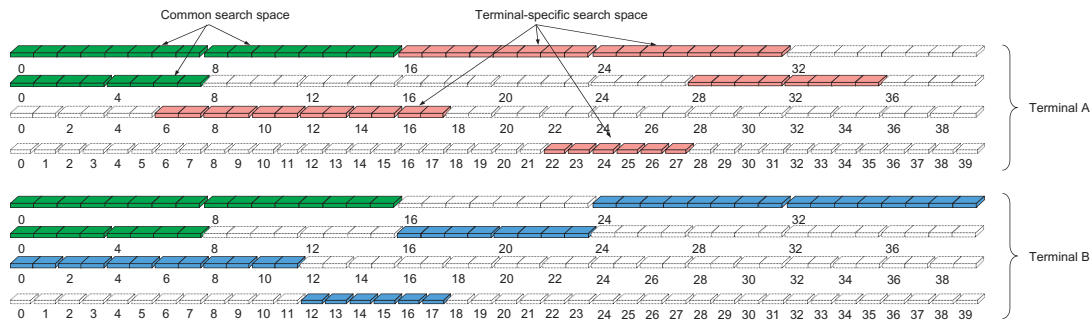
The mapping can then be understood as follows: First, the PCFICH is mapped to four resource-element groups, followed by allocating the resource-element groups required for the PHICH. The resource-element groups left after the PCFICH and PHICH are used for the different PDCCHs in the system. In this particular example, one PDCCH is using CCE numbers 0 and 1, while another PDCCH is using CCE number 4. Consequently, there is a relatively large amount of unused resource-element groups in this example; either they can be used for additional PDCCHs or the power otherwise used for the unused CCEs could be allocated to the PDCCHs in use (as long as the power difference between resource elements is kept within the limits set by the RF requirements). Furthermore, depending on the inter-cell interference situation, fractional loading of the control region may be desirable, implying that some CCEs are left unused to reduce the average inter-cell interference.

10.4.9 Blind Decoding of PDCCHs

As described above, each PDCCH supports multiple formats and the format used is a priori unknown to the terminal. Therefore, the terminal needs to blindly detect the format of the PDCCHs. The CCE structure described in the previous section helps in reducing the number of blind decoding attempts, but is not sufficient. Hence, it is required to have mechanisms to limit the CCE aggregations that the terminal is supposed to monitor. Clearly, from a scheduling point of view, restrictions in the allowed CCE aggregations are undesirable as they may influence the scheduling flexibility and require additional processing at the transmitter side. At the same time, requiring the terminal to monitor all possible CCE aggregations, also for the larger cell bandwidths, is not attractive from a terminal-complexity point of view. To impose as few restrictions as possible on the scheduler while at the same time limit the maximum number of blind decoding attempts in the terminal,³² LTE defines so-called *search spaces*, which describe the set of CCEs the terminal is supposed to monitor for scheduling assignments/grants relating to a certain component carrier. Search spaces are discussed below, starting with the case of a single component carrier and later extended to the case of multiple component carriers.

A search space is a set of candidate control channels formed by CCEs on a given aggregation level, which the terminal is supposed to attempt to decode. As there are multiple aggregation levels, corresponding to one, two, four, and eight CCEs, a terminal has multiple search spaces. In each subframe, the terminals will attempt to decode all the PDCCHs that can be formed from the CCEs in each of its search spaces. If the CRC checks, the content of the control channel is declared as valid for this terminal and the terminal processes the information (scheduling assignment, scheduling grants, etc.). Clearly, the network can only address a terminal if the control information is transmitted on a PDCCH formed by the CCEs in one of the terminal's search spaces. For example, terminal A in Figure 10.34 cannot be addressed on a PDCCH starting at CCE number 20, whereas terminal B can. Furthermore, if terminal A is using CCEs 16–23, terminal B cannot be addressed on aggregation level 4 as all CCEs in its level-4 search space are blocked by the use for the other terminals. From this it can be intuitively understood that for efficient utilization of the CCEs in the system, the search spaces should differ between terminals. Each terminal in the system therefore has a *terminal-specific* search space at each aggregation level.

³²In release 8/9, the number of blind decoding attempts is 44 per subframe, while for release 10 with uplink spatial multiplexing the number is 60 (assuming a single component carrier).

**FIGURE 10.34**

Principal illustration of search spaces in two terminals.

As the terminal-specific search space is typically smaller than the number of PDCCHs the network could transmit at the corresponding aggregation level, there must be a mechanism determining the set of CCEs in the terminal-specific search space for each aggregation level. One possibility would be to let the network configure the terminal-specific search space in each terminal. However, this would require explicit signaling to each of the terminals and possibly reconfiguration at handover. Instead, the terminal-specific search spaces in LTE are defined without explicit signaling through a function of the terminal identity and implicitly the subframe number. Dependence on the subframe number results in the terminal-specific search spaces being time varying, which helps resolve blocking between terminals. If a given terminal cannot be scheduled in a subframe as all the CCEs that the terminal is monitoring have already been used for scheduling other terminals in the same subframe, the time-varying definition of the terminal-specific search spaces is likely to resolve the blocking in the next subframe.

In several situations, there is a need to address a group of, or all, terminals in the system. One example hereof is dynamic scheduling of system information, another is transmission of paging messages, both described in Chapter 14. Transmission of explicit power-control commands to a group of terminals is a third example. To allow all terminals to be addressed at the same time, LTE has defined *common search spaces* in addition to the terminal-specific search spaces. A common search space is, as the name implies, common, and all terminals in the cell monitor the CCEs in the common search spaces for control information. Although the motivation for the common search space is primarily transmission of various system messages, it can be used to schedule individual terminals as well. Thus, it can be used to resolve situations where scheduling of one terminal is blocked due to lack of available resources in the terminal-specific search space. Unlike unicast transmissions, where the transmission parameters of the control signaling can be tuned to match the channel conditions of a specific terminal, system messages typically need to reach the cell border. Furthermore, the data rate of the associated DL-SCH transmission is typically modest. Consequently, the common search spaces are only defined for aggregation levels of four and eight CCEs and only for the smallest DCI formats, 0/1A/3/3A and 1C. There is no support for DCI formats with spatial multiplexing in the common search space. This helps to reduce the number of blind decoding attempts in the terminal used for monitoring the common search space.

Table 10.7 Downlink DCI Formats Monitored in Different Search Spaces for C-RNTI

Mode	Monitored DCI Formats in Search Space			Description
	Common	UE-Specific	UE-Specific	
1	1A	1A	1	Single antenna transmission
2	1A	1A	1	Transmit diversity
3	1A	1A	2A	Open-loop spatial multiplexing
4	1A	1A	2	Closed-loop spatial multiplexing
5	1A	1A	1D	Multi-user MIMO
6	1A	1A	1B	Single layer codebook-based precoding
7	1A	1A	1	Single-layer transmission using DM-RS
8	1A	1A	2B	Dual-layer transmission using DM-RS
9	1A	1A	2C	Multi-layer transmission using DM-RS

Figure 10.34 illustrates the terminal-specific and common search spaces in two terminals in a certain subframe. The terminal-specific search spaces are different in the two terminals and will, as described above, vary from subframe to subframe. Furthermore, the terminal-specific search spaces partially overlap between the two terminals in this subframe (CCEs 24–31 on aggregation level 8) but, as the terminal-specific search space varies between subframes, the overlap in the next subframe is most likely different.

The downlink DCI formats to decode in the terminal-specific search space depend on the *transmission mode* configured for the terminal. Transmission modes are described in Section 10.3 and, in principle, correspond to different multi-antenna configurations. As an example, there is no need to attempt to decode DCI format 2 when the terminal has not been configured for spatial multiplexing, which helps to reduce the number of blind decoding attempts. Configuration of the transmission mode is done via RRC signaling. As the exact subframe number when this configuration takes effect in the terminal is not specified and may vary depending on, for example, RLC retransmissions, there is a (short) period when the network and the terminal may have different understandings of which transmission mode is configured. Therefore, in order not to lose the possibility of communicating with the terminal, it is necessary to have at least one DCI format that is decoded irrespective of the transmission mode. For downlink transmissions, DCI format 1A serves this purpose and the network can therefore always transmit data to the terminal using this DCI format. Another function of format 1A is to reduce overhead for transmissions when full flexibility in resource-block assignment is not needed.

In Table 10.7, the DCI monitoring is described for the case of a single component carrier. For multiple component carriers, the above procedures are applied in principle to each of the activated³³ downlink component carriers. Hence, in principle there is one UE-specific search space per aggregation level and per (activated) component carrier upon which PDSCH can be received (or PUSCH transmitted), although there are some carrier-aggregation-specific modifications. Clearly, for terminals configured to use carrier aggregation, this results in an increase in the number of blind decoding

³³Individual component carriers can be activated/deactivated as discussed in Chapter 12.

attempts compared to a terminal not using carrier aggregation, as scheduling assignments/grants for each of the component carriers need to be monitored.

The common search space is only defined for transmissions on the primary component carrier. As the main function of the common search space is to handle scheduling of system information intended for multiple terminals, and such information must be receivable by all terminals in the cell, scheduling in this case uses the common search space. For this reason, the carrier indication field is never present in DCI formats monitored in the common search space.

As mentioned before, there is one UE-specific search space per aggregation level and component carrier³⁴ used for the PDSCH/PUSCH. This is illustrated in Figure 10.30, where PDSCH/PUSCH transmissions on component carrier 1 are scheduled using PDCCHs transmitted on component carrier 1. No carrier indicator is assumed in the UE-specific search space for component carrier 1 as cross-carrier scheduling is not used. For component carrier 2, on the other hand, a carrier indicator is assumed in the UE-specific search space as component carrier 2 is cross-carrier scheduled from PDCCHs transmitted on component carrier 1.

Search spaces for different component carriers may overlap in some subframes. In Figure 10.30, this happens for the UE-specific search spaces for component carriers 3 and 4. The terminal will handle the two search spaces independently, assuming (in this example) a carrier indicator for component carrier 4 but not for component carrier 3. If the UE-specific and common search spaces relating to different component carriers happen to overlap for some aggregation level when cross-carrier scheduling is configured, the terminal only needs to monitor the common search space. The reason for this is to avoid ambiguities; if the component carriers have different bandwidths a DCI format in the common search space may have the same payload size as another DCI format in the UE-specific search space relating to another component carrier.

³⁴If the DCI formats for two different component carriers have the same size, then search spaces can be shared such that, in the example in Figure 10.30, component carrier 2 can be scheduled from search spaces 2 and 1. This comes “for free” as the DCI formats in search space 1 need to be monitored anyway and the carrier-indicator field points to the component carrier for which the grant/assignment is intended.

Uplink Physical-Layer Processing

11

This chapter provides a description of the basic physical-layer functionality of the LTE uplink. It essentially follows the same outline as the corresponding downlink description provided in the previous chapter with detailed descriptions regarding transport-channel processing (Section 11.1), the reference-signal structure (Section 11.2), multi-antenna transmission (Section 11.3), and the uplink L1/L2 control-channel structure (Section 11.4). Physical aspects directly related to some specific uplink functions and procedures such as random access will be provided in later chapters. The chapter ends with a discussion on the uplink timing-alignment procedure in Section 11.5.

11.1 TRANSPORT-CHANNEL PROCESSING

This section describes the physical-layer processing applied to the uplink shared channel (UL-SCH) and the subsequent mapping to the uplink physical resource in the form of the basic OFDM time-frequency grid. The UL-SCH is the only uplink transport-channel type in LTE¹ and is used for the transmission of all uplink higher-layer information (both data and control).

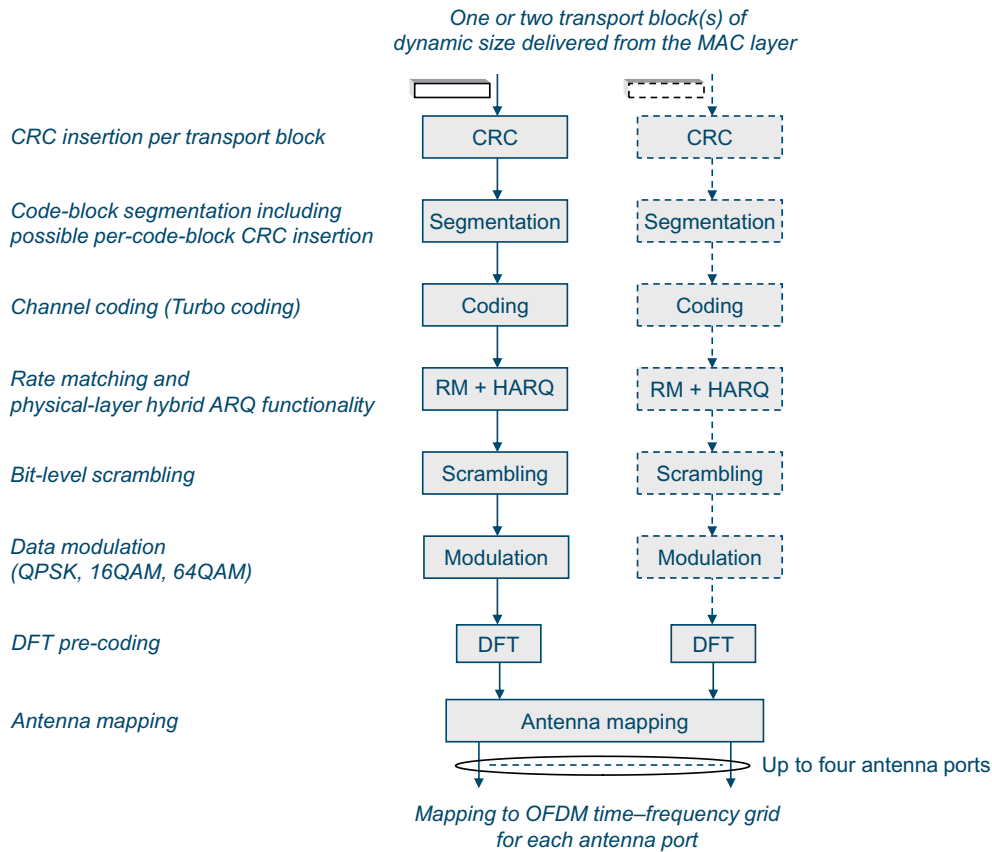
11.1.1 Processing Steps

Figure 11.1 outlines the different steps of the UL-SCH physical-layer processing. Similar to the downlink, in the case of uplink carrier aggregation the different component carriers correspond to separate transport channels with separate physical-layer processing. Also, most of the uplink processing steps outlined in Figure 11.1 are identical to the corresponding steps for the downlink transport-channel processing outlined in Section 10.1.

The different steps of the uplink transport-channel processing are summarized below. For a more detailed overview of the different steps, the reader is referred to the corresponding downlink description in Section 10.1.

- *CRC insertion per transport block.* A 24-bit CRC is calculated for and appended to each uplink transport block.
- *Code-block segmentation and per-code-block CRC insertion.* In the same way as for the downlink, code-block segmentation, including per-code-block CRC insertion, is applied for transport blocks larger than 6144 bits.

¹Strictly speaking, the LTE Random-Access Channel is also defined as a transport-channel type, see Chapter 8. However, RACH only includes a layer-1 preamble and carries no actual transport-channel data.

**FIGURE 11.1**

Physical-layer processing for UL-SCH.

- *Channel coding.* Rate-1/3 Turbo coding with QPP-based inner interleaving is also used for the uplink.
- *Rate matching and physical-layer hybrid-ARQ functionality.* The physical-layer part of the uplink rate-matching and hybrid-ARQ functionality is essentially the same as the corresponding downlink functionality, with sub-block interleaving and insertion into a circular buffer followed by bit selection with four redundancy versions. Note that there are some important differences between the downlink and uplink hybrid-ARQ protocols, such as asynchronous versus synchronous operation as described in Chapter 12. However, these differences are not visible in terms of the physical-layer processing.
- *Bit-level scrambling.* The aim of uplink scrambling is the same as for the downlink – that is, to randomize the interference and thus ensure that the processing gain provided by the channel code can be fully utilized.

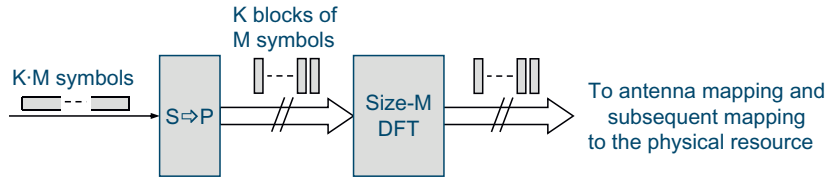


FIGURE 11.2

DFT precoding K blocks of M modulation symbols.

- *Data modulation.* Similar to the downlink, QPSK, 16QAM, and 64QAM modulation can also be used for uplink transport-channel transmission.
- *DFT precoding.* As illustrated in Figure 11.2, the modulation symbols, in blocks of M symbols, are fed through a size- M DFT, where M corresponds to the number of subcarriers assigned for the transmission. The reason for the precoding is to reduce the cubic metric for the transmitted signal. From an implementation-complexity point of view, the DFT size should preferably be constrained to a power of 2. However, such a constraint would limit the scheduler flexibility in terms of the amount of resources that can be assigned for an uplink transmission and, from a flexibility point of view, all possible DFT sizes should preferably be allowed. For LTE, a middle way has been adopted where the DFT size, and thus also the size of the resource allocation, is limited to products of the integers 2, 3, and 5. Thus, for example, DFT sizes of 60, 72, and 96 are allowed but a DFT size of 84 is not allowed.² In this way, the DFT can be implemented as a combination of relatively low-complex radix-2, radix-3, and radix-5 FFT processing.
- *Antenna mapping.* The antenna mapping maps the output of the DFT precoder to antenna ports for subsequent mapping to the physical resource (the OFDM time–frequency grid). In the first releases of LTE (release 8 and 9), only single-antenna transmission was supported for the uplink.³ However, as part of LTE release 10, multi-antenna transmission, more specifically antenna precoding with up to four antennas ports, is supported in the uplink. More details about LTE uplink multi-antenna transmission are provided in Section 11.3.

11.1.2 Mapping to the Physical Resource

The scheduler assigns a set of resource-block pairs to be used for the uplink transmission, more specifically for transmission of the PUSCH physical channel that carries the UL-SCH transport channel. Each such resource-block pair spans 14 OFDM symbols (one subframe) in time.⁴ However, as will be described in Section 11.2.1, two of these symbols are used for *uplink demodulation reference signals* and are thus not available for PUSCH transmission. Furthermore, one additional symbol may be

²As uplink resource assignments are always done in terms of resource blocks of size 12 subcarriers, the DFT size is always a multiple of 12.

³Actually, uplink *antenna selection* has been part of the LTE specification since release 8. However, it is an optional terminal feature that has, as of today, not been implemented in any commercial terminals.

⁴Twelve symbols in the case of extended cyclic prefix.

reserved for the transmission of *sounding reference signals* (Section 11.2.2). Thus, 11 or 12 OFDM symbols are available for PUSCH transmission within each resource-block pair.

Figure 11.3 illustrates how $K \cdot M$ DFT-precoded symbols at the output of the antenna mapping are mapped to the basic OFDM time–frequency grid, where K is the number of available OFDM symbols within a subframe (11 or 12 according to above) and M is the assigned bandwidth in number of subcarriers. As there are 12 subcarriers within a resource-block pair, $M = N \cdot 12$, where N is the number of assigned resource-block pairs.

Figure 11.3 assumes that the set of assigned resource-block pairs are contiguous in the frequency domain. This is the typical assumption for DFTS-OFDM and was strictly the case for LTE release 8 and 9. Mapping of the DFT-precoded signal to frequency-contiguous resource elements is preferred in order to retain good cubic-metric properties of the uplink transmission. At the same time, such a restriction implies an additional constraint for the uplink scheduler, something which may not always be desirable. Therefore, LTE release 10 introduced the possibility to also assign partly frequency-separated resources for PUSCH transmission. More specifically, in release 10 the assigned uplink resource may consist of a maximum of two frequency-separated *clusters*, as illustrated in Figure 11.4, where each cluster consists of a number of resource-block pairs ($N1$ and $N2$ resource-block pairs respectively). In the case of such *multi-cluster transmission*, a single DFT precoding spans the overall assigned resource in the frequency domain – that is, both clusters. This means that the total assigned bandwidth in number of subcarriers ($M = M1 + M2$) should be aligned with the restrictions on available DFT sizes described above.

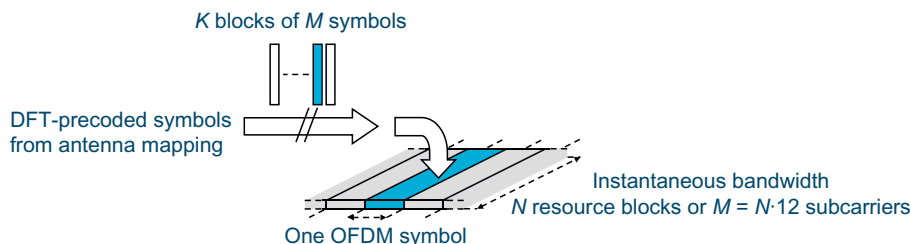


FIGURE 11.3

Mapping to the uplink physical resource.

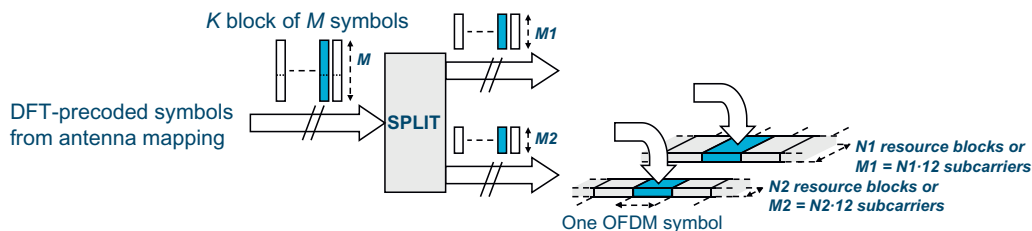


FIGURE 11.4

Uplink multi-cluster transmission.

11.1.3 PUSCH Frequency Hopping

In Chapter 10 it was described how the notion of *virtual resource blocks* (VRBs) in combination with the mapping from VRBs to *physical resource blocks* (PRBs) allowed for downlink distributed transmission – that is, the spreading of a downlink transmission in the frequency domain. As described, downlink distributed transmission consists of two separate steps: (1) a mapping from VRB pairs provided as part of the uplink scheduling grant to PRB pairs such that frequency-consecutive VRB pairs are not mapped to frequency-consecutive PRB pairs and (2) a split of each resource-block pair such that the two resource blocks of the resource-block pair are transmitted with a certain frequency gap in between. This second step can be seen as frequency hopping on a slot basis. In the absence of uplink frequency hopping, each VRB is directly mapped to a PRB.

The notion of VRBs can also be used for the LTE uplink, allowing for frequency-domain-distributed transmission for the uplink. However, in the uplink, where transmission from a terminal should always be over a set of consecutive subcarriers in the absence of multi-cluster transmission, distributing resource-block pairs in the frequency domain, as in the first step of downlink distributed transmission, is not possible. Rather, uplink distributed transmission is similar to the second step of downlink distributed transmission – that is, a frequency separation of the transmissions in the first and second slots of a subframe. Uplink distributed transmission for PUSCH can thus more directly be referred to as *uplink frequency hopping*.

There are two types of uplink frequency hopping defined for PUSCH:

- Sub-band-based hopping according to cell-specific hopping/mirroring patterns.
- Hopping based on explicit hopping information in the scheduling grant.

Uplink frequency hopping is not supported for multi-cluster transmission as, in that case, sufficient diversity can be obtained by the proper location of the two clusters.

11.1.3.1 Hopping Based on Cell-Specific Hopping/Mirroring Patterns

To support sub-band-based hopping according to cell-specific hopping/mirroring patterns, a set of consecutive sub-bands of a certain size is defined from the overall uplink frequency band, as illustrated in Figure 11.5. It should be noted that the sub-bands do not cover the total uplink frequency

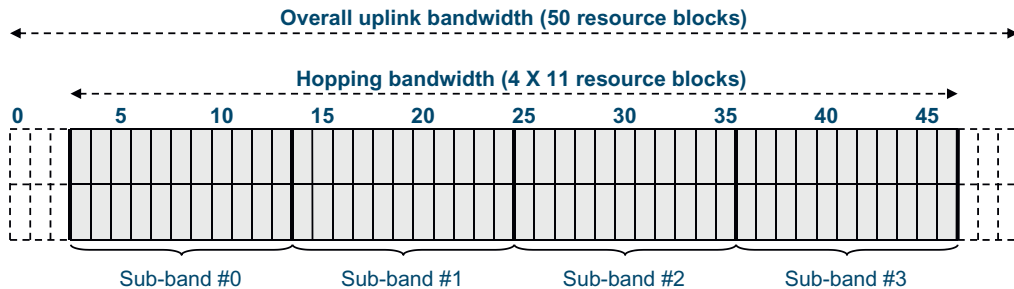


FIGURE 11.5

Definition of sub-bands for PUSCH hopping. The figure assumes a total of four sub-bands, each consisting of 11 resource blocks.

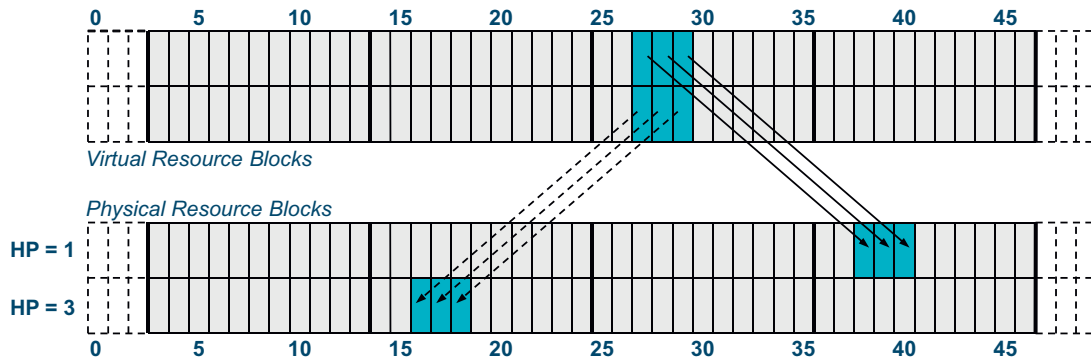


FIGURE 11.6

Hopping according to predefined hopping pattern.

band, mainly due to the fact that a number of resource blocks at the edges of the uplink frequency band are used for transmission of L1/L2 control signaling on the PUCCH. For example, in Figure 11.5 the overall uplink bandwidth corresponds to 50 resource blocks and there are a total of four sub-bands, each consisting of 11 resource blocks. Six resource blocks are not included in the hopping bandwidth and could, for example, be used for PUCCH transmission.

In the case of sub-band-based hopping, the set of virtual resource blocks provided in the scheduling grant are mapped to a corresponding set of physical resource blocks according to a cell-specific hopping pattern. The resource to use for transmission, the *PRBs*, is obtained by shifting the *VRBs* provided in the scheduling grant by a number of sub-bands according to the hopping pattern, where the hopping pattern can provide different shifts for each slot. As illustrated in Figure 11.6, a terminal is assigned *VRBs* 27, 28, and 29. In the first slot, the predefined hopping pattern takes the value 1, implying transmission using *PRBs* one sub-band to the right – that is, *PRBs* 38, 39, and 40. In the second slot, the predefined hopping pattern takes the value 3, implying a shift of three sub-bands to the right in the figure and, consequently, transmission using resource blocks 16, 17, and 18. Note that the shifting “wraps-around” – that is, in the case of four sub-bands, a shift of three sub-bands is the same as a negative shift of one sub-band. As the hopping pattern is cell specific – that is, the same for all terminals within a cell – different terminals will transmit on non-overlapping physical resources as long as they are assigned non-overlapping virtual resources.

In addition to the hopping pattern, there is also a cell-specific *mirroring pattern* defined in a cell. The mirroring pattern controls, on a slot basis, whether or not mirroring within each sub-band should be applied to the assigned resource. In essence, mirroring implies that the resource blocks within each sub-band are numbered right to left instead of left to right. Figure 11.7 illustrates mirroring in combination with hopping. Here, the mirroring pattern is such that mirroring is not applied to the first slot while mirroring is applied to the second slot.

Both the hopping pattern and the mirroring pattern depend on the physical-layer cell identity and are thus typically different in neighboring cells. Furthermore, the period of the hopping/mirroring patterns corresponds to one frame.

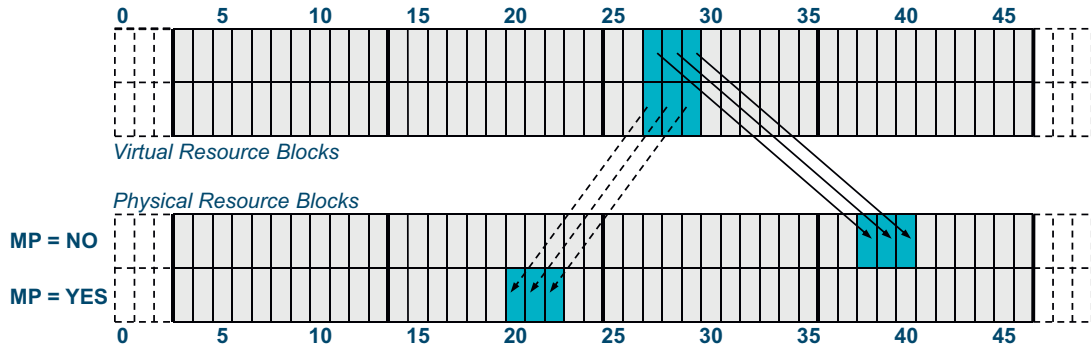


FIGURE 11.7

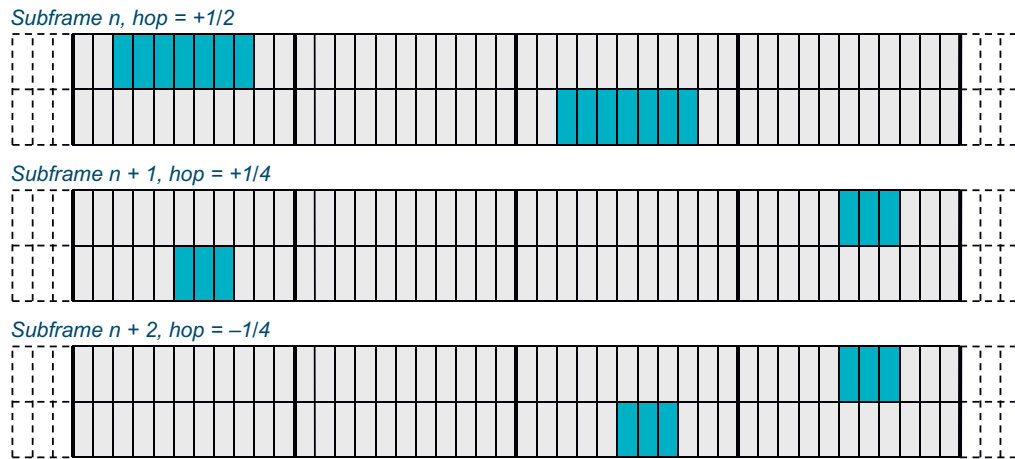
Hopping/mirroring according to predefined hopping/mirroring patterns. Same hopping pattern as in Figure 11.6.

11.1.3.2 Hopping Based on Explicit Hopping Information

As an alternative to hopping/mirroring according to cell-specific hopping/mirroring patterns as described above, uplink slot-based frequency hopping for PUSCH can also be controlled by *explicit hopping information* provided in the scheduling grant. In such a case the scheduling grant includes:

- information about the resource to use for uplink transmission in the first slot, exactly as in the non-hopping case; and
- additional information about the offset of the resource to use for uplink transmission in the second slot, relative to the resource of the first slot.

Selection between hopping according to cell-specific hopping/mirroring patterns as described above or hopping according to explicit information in the scheduling grant can be done dynamically. More specifically, for cell bandwidths less than 50 resource blocks, there is a single bit in the scheduling grant indicating if hopping should be according to the cell-specific hopping/mirroring patterns or should be according to information in the scheduling grant. In the latter case, the hop is always half of the hopping bandwidth. In the case of larger bandwidths (50 resource blocks and beyond), there are two bits in the scheduling grant. One of the combinations indicates that hopping should be according to the cell-specific hopping/mirroring patterns while the three remaining alternatives indicate hopping of $1/2$, $+1/4$, and $-1/4$ of the hopping bandwidth. Hopping according to information in the scheduling grant for the case of a cell bandwidth corresponding to 50 resource blocks is illustrated in Figure 11.8. In the first subframe, the scheduling grant indicates a hop of one-half the hopping bandwidth. In the second subframe, the grant indicates a hop of one-fourth the hopping bandwidth (equivalent to a negative hop of three-fourths of the hopping bandwidth). Finally, in the third subframe, the grant indicates a negative hop of one-fourth the hopping bandwidth.

**FIGURE 11.8**

Frequency hopping according to explicit hopping information.

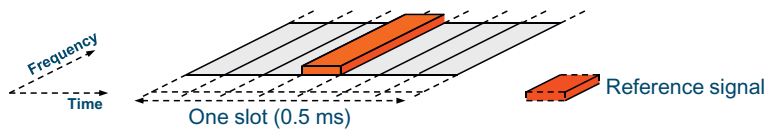
11.2 UPLINK REFERENCE SIGNALS

Similar to the downlink, reference signals are also transmitted on the LTE uplink. There are two types of reference signals defined for the LTE uplink:

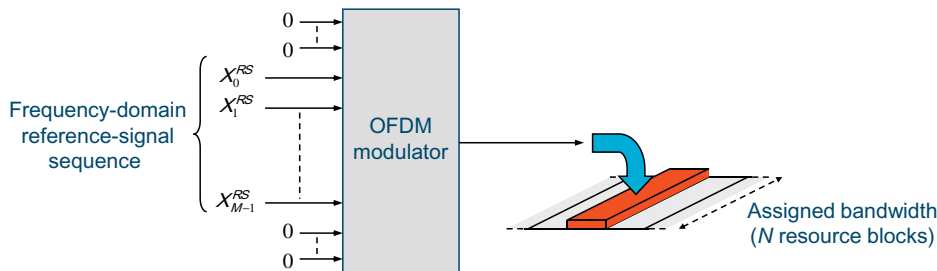
- Uplink *demodulation reference signals* (DM-RS) are intended to be used by the base station for channel estimation for coherent demodulation of the uplink physical channels (PUSCH and PUCCH). Demodulation reference signals are thus only transmitted together with PUSCH or PUCCH and are then transmitted with the same bandwidth as the corresponding physical channel.
- Uplink *sounding reference signals* (SRS) are intended to be used by the base station for channel-state estimation to support uplink channel-dependent scheduling and link adaptation. The SRS can also be used in cases when uplink transmission is needed, although there is no data to transmit. One example is when uplink transmission is needed for the network to be able to control the uplink transmit timing by means of the *uplink-timing-alignment procedure* (Section 11.5). Finally, the uplink SRS can also be used to estimate the downlink channel-state assuming sufficient uplink/downlink reciprocity. This is especially of interest for TDD, where the operation of downlink and uplink on the same carrier frequency typically implies a higher degree of short-term downlink/uplink reciprocity compared to FDD.

11.2.1 Uplink Demodulation Reference Signals

Uplink demodulation reference signals are used for channel estimation for coherent demodulation of the Physical Uplink Shared Channel (PUSCH) to which the UL-SCH transport channel is mapped, as well as for the Physical Uplink Control Channel (PUCCH), which carries different types of L1/L2 control signaling (see Section 11.4). The basic structure for demodulation reference signals is

**FIGURE 11.9**

Transmission of uplink reference signals within a slot in the case of PUSCH transmission.

**FIGURE 11.10**

Generation of uplink reference signal from a frequency-domain reference-signal sequence.

the same for PUSCH and PUCCH transmission, although there are some differences – for example, in terms of the exact set of OFDM symbols in which the reference signals are transmitted.

11.2.1.1 Basic Principles of Uplink DM-RS

Due to the importance of low cubic metric and corresponding high power-amplifier efficiency for uplink transmissions, the principles for uplink reference-signal transmission are different compared to the downlink. In essence, transmitting reference signals in parallel with other uplink transmissions from the same terminal is not suitable for the uplink. Instead, certain OFDM symbols are used exclusively for DM-RS transmission – that is, the reference signals are *time multiplexed* with other uplink transmissions from the same terminal. The structure of the reference signal itself then ensures a low cubic metric in these symbols.

More specifically, in the case of PUSCH transmission a demodulation reference signal is transmitted within the fourth symbol of each uplink slot⁵ (Figure 11.9). Within each subframe, there are thus two reference-signal transmissions, one in each slot.

In the case of PUCCH transmission, the number of OFDM symbols used for reference-signal transmission in a slot, as well as the exact position of these symbols, differs between different *PUCCH formats*, as described in Section 11.4.

Regardless of the kind of uplink transmission (PUSCH or PUCCH), the basic structure of each reference-signal transmission is the same. As illustrated in Figure 11.10, an uplink reference signal

⁵The third symbol in the case of extended cyclic prefix.

can be defined as a *frequency-domain reference-signal sequence* applied to consecutive inputs of the OFDM modulator – that is, to consecutive subcarriers.

In general, there is no reason to estimate the channel outside the frequency band of the corresponding PUSCH/PUCCH transmission that is to be coherently demodulated. The bandwidth of the reference signal, corresponding to the length of the reference-signal sequence in Figure 11.10, should thus be equal to the bandwidth of the corresponding PUSCH/PUCCH transmission measured in number of subcarriers. This means that, for PUSCH transmission, it should be possible to generate reference-signal sequences of different length, corresponding to the possible bandwidths of a PUSCH transmission. Note, however, that the length of the reference-signal sequence will always be a multiple of 12, as uplink resource allocations for PUSCH transmission are always done in terms of resource blocks consisting of 12 subcarriers.

As will be described in Section 11.4, PUCCH transmission from a terminal is always carried out over a single resource block in the frequency domain. In the case of PUCCH transmission, the length of the reference-signal sequence is thus always equal to 12.

Uplink reference signals should preferably have the following properties:

- Limited power variations in the frequency domain to allow for similar channel-estimation quality for all frequencies. Note that this is equivalent to good time-domain cross-correlation properties of the transmitted reference signal.
- Limited power variations in the time domain, leading to low cubic metric of the transmitted signal.

Furthermore, sufficient reference-signal sequences of a given length, corresponding to a certain reference-signal bandwidth, should be available in order to avoid an unreasonable planning effort when assigning reference-signal sequences to cells.

So-called *Zadoff–Chu* sequences [71] have the property of constant power in both the frequency and time domains. The elements of a Zadoff–Chu sequence of length M_{ZC} can be expressed as:

$$X_k^{\text{ZC}} = e^{-j\pi u \frac{k(k+1)}{M_{\text{ZC}}}} \quad 0 \leq k < M_{\text{ZC}}, \quad (11.1)$$

where u is the *index* of the Zadoff–Chu sequence within the set of Zadoff–Chu sequences of length M_{ZC} . From the point of view of small power variations in both the frequency and time domains, Zadoff–Chu sequences would thus be excellent as uplink reference-signal sequences. However, there are two reasons why Zadoff–Chu sequences are not suitable for direct use as uplink reference-signal sequences in LTE:

- The number of available Zadoff–Chu sequences of a certain length, corresponding to the number of possible values of the index u in the expression above, equals the number of integers that are relative prime to the sequence length. This implies that to maximize the number of Zadoff–Chu sequences and thus, in the end, to maximize the number of available uplink reference signals, prime-length Zadoff–Chu sequences would be preferred. At the same time, the length of the uplink reference-signal sequences should be a multiple of 12, which is obviously not a prime number.
- For short sequence lengths, corresponding to narrow uplink transmission bandwidths, relatively few reference-signal sequences would be available even if they were based on prime-length Zadoff–Chu sequences.

Instead, for sequence lengths larger than or equal to 36, corresponding to transmission bandwidths larger than or equal to three resource blocks, the reference-signal sequences are defined as *cyclic extensions* of Zadoff–Chu sequences of length M_{ZC} , where M_{ZC} is the largest prime number smaller than or equal to the desired reference-signal sequence length. For example, the largest prime number less than or equal to 36 is 31, implying that reference-signal sequences of length 36 are defined as cyclic extensions of Zadoff–Chu sequences of length 31. The number of available sequences is then equal to 30 – that is, one less than the length of the Zadoff–Chu sequence. For larger sequence lengths, more sequences are available. For example, for a reference-signal sequence length equal to 72, there are 70 sequences available.⁶

For sequence lengths equal to 12 and 24, corresponding to transmission bandwidths of one and two resource blocks respectively, special QPSK-based sequences have instead been found from computer searches and are explicitly listed in the LTE specifications. For each of the two sequence lengths, 30 sequences are available.

11.2.1.2 Phase-Rotated Reference-Signal Sequences

In the previous section it was described how different reference-signal sequences (a minimum of 30 sequences for each sequence length) can be derived, primarily by cyclically extending *different prime-length Zadoff–Chu sequences*. Additional reference-signal sequences can be derived by applying different linear phase rotations to the same basic reference-signal sequence, as illustrated in Figure 11.11, where the basic reference-signal sequence is defined according to the above – that is, a cyclically extended Zadoff–Chu sequence for sequence lengths larger than or equal to 36 and a “special” sequence for shorter sequence lengths.

Applying a linear phase rotation in the frequency domain is equivalent to applying a cyclic shift in the time domain. Thus, although being *defined* as different *frequency-domain phase rotations*, in

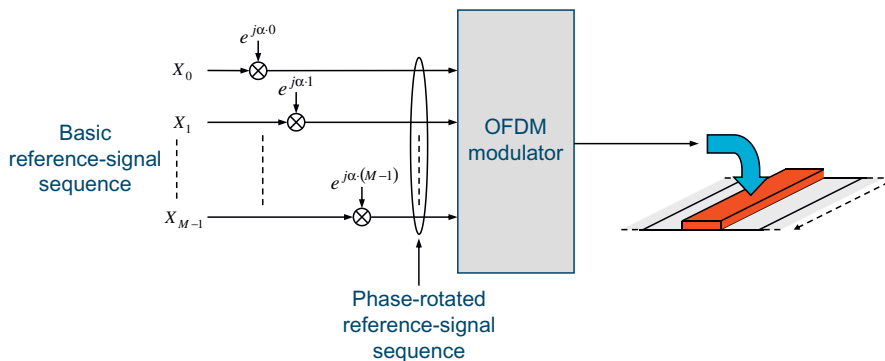


FIGURE 11.11

Generation of uplink reference-signal sequence from linear phase rotation of a basic reference-signal sequence.

⁶The largest prime number smaller than or equal to 72 is 71. The number of sequences is then one less than the length of the Zadoff–Chu sequence, i.e. 70.

the LTE specification this is often referred to as applying different *cyclic shifts*⁷ to the same basic reference-signal sequence. Here the term “phase rotation” will be used. However, it should be borne in mind that what is referred to here as “phase rotation” is often referred to as “cyclic shift” in the 3GPP-related literature.

Demodulation reference signals defined from different reference-signal sequences typically have relatively low but still non-zero mutual correlation. In contrast, reference signals defined from different phase rotations of the same basic reference-signal sequence can be made completely orthogonal, thus causing no interference to each other, assuming the parameter α in Figure 11.11 takes a value $m\pi/6$, where m ranges from 0 to 11.⁸ Up to 12 orthogonal reference signals can thus be defined from each basic reference-signal sequence.

However, to preserve the orthogonality between these reference signals also at the receiver side, the frequency response of the channel should essentially be constant (non-selective) over the span of 12 subcarriers (one resource block). Alternatively expressed, the main part of the channel time dispersion should not extend beyond the length of the cyclic shift mentioned above. If that is not the case, a subset of the available values for α may be used – for example, only the values $\{0, 2\pi/6, 4\pi/6, \dots, 10\pi/6\}$ or perhaps even fewer values. Limiting the set of values for α implies orthogonality over a smaller number of subcarriers and thus less sensitivity to channel frequency selectivity.

Another prerequisite for orthogonality between reference signals defined from different phase rotations of the same basic reference-signal sequence is that the reference signals are received well aligned in time. Any timing misalignment between the reference signals will, in the frequency domain, appear as a phase rotation that may counteract the phase rotation applied to separate the reference-signal sequences and cause interference between them.

In general for LTE, uplink transmissions from different terminals within the same cell are typically well time aligned anyway, as this is a prerequisite for retaining the orthogonality between different frequency-multiplexed transmissions. Thus, one possible use of reference signals defined from different phase rotations of the same basic reference-signal sequence is for the case when multiple terminals within the same cell simultaneously transmit on the uplink *using the same frequency resource*. As will be seen in Section 11.4, this is, for example, the case for uplink PUCCH transmissions.

Multiple reference signals are also needed for PUSCH transmission in the case of spatial multiplexing (Section 11.3.1) and multi-user MIMO (Section 11.3.2). As discussed in Section 11.2.1.5, using different phase rotations is one approach by which multiple reference signals can be provided in these cases.

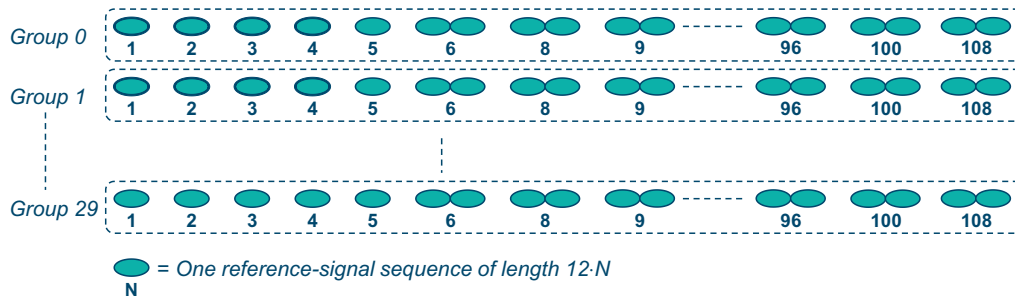
11.2.1.3 Reference-Signal Assignment to Cells

As described above, the design of the LTE reference-signal sequences allows for at least 30 sequences of each sequence length. To assign reference-signal sequences to cells, these sequences are first grouped into 30 *sequence groups* where each group consists of:

- One reference-signal sequence for each sequence length less than or equal to 60, corresponding to a transmission bandwidth of five resource blocks or less.

⁷Not to be confused with the cyclic extension of Zadoff–Chu sequences to generate the basic reference-signal sequences. The cyclic shift discussed here would be a time-domain cyclic shift, equivalent to a frequency-domain phase rotation.

⁸The orthogonality is due to the fact that, for $\alpha = m\pi/6$, there will be an integer number of full-circle rotations over 12 subcarriers – that is, over one resource block.

**FIGURE 11.12**

Grouping of reference-signal sequences into sequence groups. The number indicates the corresponding bandwidth in number of resource blocks.

- Two reference-signal sequences for each sequence length larger than or equal to 72,⁹ corresponding to a transmission bandwidth of six resource blocks or more.

The grouping of reference-signal sequences into sequence groups is illustrated in Figure 11.12. It can be noted that the sequence groups do not include any reference-signal sequences corresponding to seven resource blocks. As described in Section 11.1.1, the uplink transmission bandwidths, measured in number of resource blocks, should always be a product of the integers 2, 3, and 5.¹⁰ A resource assignment of size seven resource blocks is thus not a valid resource assignment and there is no need to define reference-signal sequences of the corresponding length. For the same reason, the maximum length of the reference-signal sequences corresponds to 108 resource blocks, rather than 110 resource blocks. Similarly, the second largest sequence length corresponds to 100 resource blocks.

In a given time slot, the uplink reference-signal sequences to use within a cell are taken from one specific sequence group. What group to use within a cell may then be the same for all slots (fixed assignment). Alternatively, what group to use within a cell may vary between slots, also referred to as *group hopping*. The system information indicates if there is a fixed group assignment or if group hopping is to be used.

Regardless of whether a fixed group assignment or group hopping is used, there are certain differences between PUCCH and PUSCH transmission.

In the case of a fixed group assignment, the sequence group to use for PUCCH transmission is given by the physical-layer cell identity (see Chapter 14) modulo 30, where the cell identity ranges from 0 to 503. Thus, cell identities 0, 30, 60, ..., 480 correspond to sequence group 0, cell identities 1, 31, 61, ..., 481 correspond to sequence group 1, and so on.

In contrast, what sequence group to use for PUSCH transmission is explicitly signaled as part of the cell system information. The reason for explicitly indicating what sequence group to use for

⁹For sequence lengths larger than or equal to 72, there are more than 60 sequences of each length allowing for two sequences per group.

¹⁰Strictly speaking, the number of subcarriers should be a product of the integers 2, 3, and 5. However, as there are $12 = 3 \cdot 2^2$ subcarriers within one resource block, this is equivalent to a constraint that the number of resource blocks should be a product of the integers 2, 3, and 5.

PUSCH transmission is that it should be possible to use the same sequence group for PUSCH transmission in neighboring cells, despite the fact that such cells typically have different cell identities, resulting in reference-signal orthogonality also *between* cells and a corresponding decrease in interference. In this case, the reference signals for PUSCH transmission would instead be distinguished by different phase rotations, as discussed in Section 11.2.1.2.

In the case of group hopping, an additional cell-specific group-hopping pattern is added to the sequence group. Hence, in the case of reference signals for PUCCH transmission, the sequence group to use in a given slot is given by the value of the group-hopping pattern for that slot plus the cell identity modulo 30. The group-hopping pattern is also given by the cell identity and identical group-hopping patterns are used for PUSCH and PUCCH within a cell.

Thus, to have the same sequence group for PUSCH in two cells:

- The cell identities of the two cells should be different (to have different sequence groups for PUCCH, which should always be the case).
- If group hopping is enabled, the cell identities of the two cells should be such that the same group-hopping pattern is used in the two cells.
- The sequence group for PUSCH (indicated as part of the cell system information) should be the same for the two cells.

11.2.1.4 Sequence Hopping

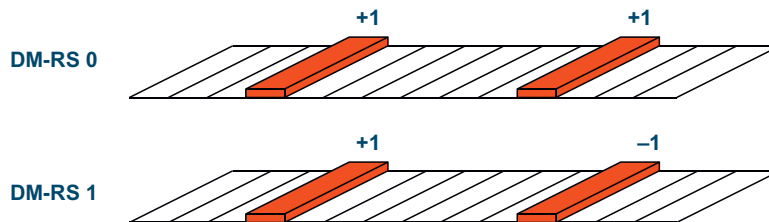
As can be seen from Figure 11.12, for sequence lengths corresponding to six resource blocks and above, there are two reference-signal sequences in each sequence group. What sequence to use may then either be fixed (always the first sequence within the group) or vary between slots, also referred to as *sequence hopping* (not to be mixed with group hopping as described in Section 11.2.1.3). The purpose of sequence hopping is to randomize the inter-cell interference between reference signals. Sequence hopping can only be enabled if group hopping is not enabled – that is, if there is a fixed assignment of groups to cells.

11.2.1.5 Multiple Demodulation Reference Signals

Support for uplink multi-antenna transmission, more specifically multi-antenna precoding including spatial multiplexing, was introduced in LTE release 10. As will be seen in Section 11.3.1, in the case of spatial multiplexing there is a need for one demodulation reference signal *per layer*. As LTE release 10 supports the transmission of up to four spatially multiplexed layers in parallel, there is thus a need to be able to transmit up to four demodulation reference signals from the same terminal. Such multiple reference signals can be generated by different means:

- Multiple mutually orthogonal reference signals can be generated by using different phase rotations (“cyclic shifts”), as already described in Section 11.2.1.2.
- Two different reference signals can be created from the same reference signal sequence by applying mutually orthogonal patterns, similar to the downlink referred to as *orthogonal cover codes* (OCC), to the two reference-signal transmissions within a subframe, as illustrated in Figure 11.13.

Multiple orthogonal reference signals, as described above, are also used in the case of uplink multi-user MIMO (Section 11.3.2).

**FIGURE 11.13**

Generation of multiple demodulation reference signals from orthogonal cover codes.

11.2.2 Uplink Sounding Reference Signals

The demodulation reference signals discussed in Section 11.2.1 are used for channel estimation to allow for coherent demodulation of uplink physical channels (PUSCH or PUCCH). A demodulation reference signal is always transmitted together with and covering the same frequency band as the corresponding physical channel.

In contrast, *sounding reference signals* (SRS) are transmitted on the uplink to allow for the base station to estimate the uplink *channel state* at different frequencies. The channel-state estimates can then, for example, be used by the network scheduler to assign resource blocks of instantaneously good quality for uplink PUSCH transmission (uplink channel-dependent scheduling), as well as to select different transmission parameters such as the instantaneous data rate and different parameters related to uplink multi-antenna transmission. As mentioned earlier, SRS transmission can also be used for uplink timing estimation as well as to estimate downlink channel conditions assuming downlink/uplink channel reciprocity. Thus, an SRS is not necessarily transmitted together with any physical channel and if transmitted together with, for example, PUSCH, the SRS may cover a different, typically larger, frequency span.

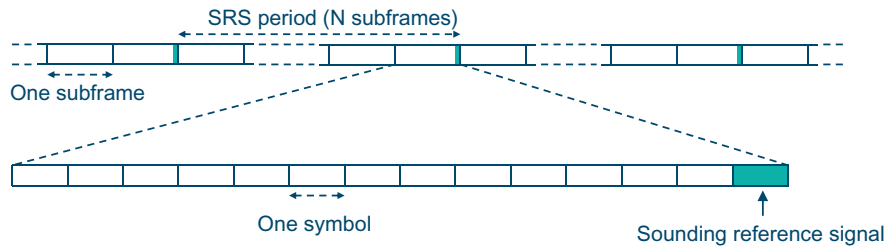
There are two types of SRS transmission defined for the LTE uplink: *periodic* SRS transmission, which has been available from the first release of LTE (release 8); and *aperiodic* SRS transmission, introduced in LTE release 10.

11.2.2.1 Periodic SRS Transmission

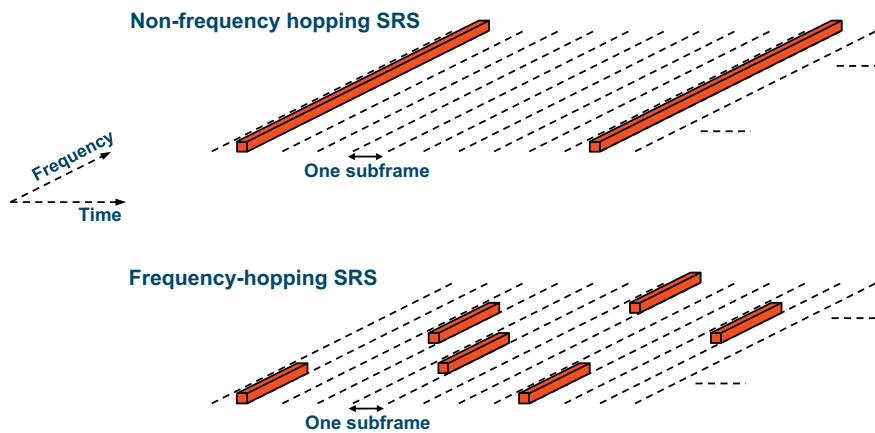
Periodic SRS transmission from a terminal occurs at regular time intervals, from as often as once every 2 ms (every second subframe) to as infrequently as once every 160 ms (every 16th frame). When SRS is transmitted in a subframe, it occupies the last symbol of the subframe, as illustrated in [Figure 11.14](#). As an alternative, in the case of TDD operation, SRS can also be transmitted within the UpPTS.

In the frequency domain, SRS transmissions should cover the frequency band that is of interest for the scheduler. This can be achieved in two ways:

- By means of a sufficiently wideband SRS transmission that allows for sounding of the entire frequency band of interest with a single SRS transmission, as illustrated in the upper part of [Figure 11.15](#).

**FIGURE 11.14**

Transmission of SRS.

**FIGURE 11.15**

Non-frequency-hopping (wideband) versus frequency-hopping SRS.

- By means of more narrowband SRS transmission – that is, hopping in the frequency domain – in such a way that a sequence of SRS transmissions jointly covers the frequency band of interest, as illustrated in the lower part of [Figure 11.15](#).

The main benefit of wideband (non-hopping) SRS transmission according to the upper part of [Figure 11.15](#) is that the entire frequency band of interest can be sounded with a single SRS transmission – that is, within a single OFDM symbol. As described below, the entire OFDM symbol in which SRS is transmitted will be unavailable for data transmission in the cell. A single wideband SRS transmission is thus more efficient from a resource-utilization point of view as less OFDM symbols need to be used to sound a given overall bandwidth. However, in the case of a high uplink path loss, wideband SRS transmission may lead to relatively low received power density, which may degrade the channel-state estimation. In such a case it may be preferable to use a more narrowband SRS

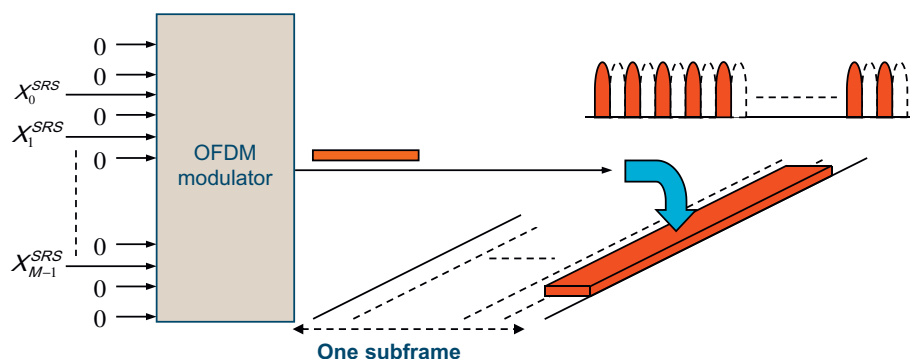


FIGURE 11.16

Generation of SRS from a frequency-domain reference-signal sequence.

transmission, thereby focusing the available transmit power in a more narrow frequency range, and then hop over the total band to be sounded.

In general, different bandwidths of the SRS transmission can be available within a cell. A narrow SRS bandwidth, corresponding to four resource blocks, is always available in all cells, regardless of the uplink cell bandwidth. Up to three additional, more wideband SRS bandwidths may also be configured within the cell. The SRS bandwidths are then always a multiple of four resource blocks. A terminal is then explicitly configured to use one of the SRS bandwidths available in the cell.

If a terminal is transmitting SRS in a certain subframe, the SRS transmission may very well overlap, in the frequency domain, with PUSCH transmissions from other terminals within the cell. To avoid collision between SRS and PUSCH transmissions from different terminals, terminals should in general avoid PUSCH transmission in the OFDM symbols in which SRS transmission may occur. To achieve this, all terminals within a cell are aware of the set of subframes within which SRS *may* be transmitted by *any* terminal within the cell. All terminals should then avoid PUSCH transmission in the last OFDM symbol of those subframes.¹¹ Information about the set of subframes in which SRS *may* be transmitted within a cell is provided as part of the cell system information.

On a more detailed level, the structure for sounding reference signals (SRS) is similar to that of uplink demodulation reference signals described in Section 11.2.1. More specifically, a sounding reference signal is also defined as a frequency-domain reference-signal sequence derived as a cyclic extension of prime-length Zadoff–Chu sequences. However, in the case of SRS, the reference-signal sequence is mapped to *every second subcarrier*, creating a “comb”-like spectrum, as illustrated in Figure 11.16. Taking into account that the bandwidth of the SRS transmission is always a multiple of four resource blocks, the lengths of the reference-signal sequences for SRS are thus always a multiple of 24. The reference-signal sequence to use for SRS transmission within the cell is taken from the same sequence group as the demodulation reference signals used for channel estimation for PUCCH.

¹¹SRS transmission in UpPTS can obviously not collide with PUSCH transmission as there is no PUSCH transmission in UpPTS.

Similar to demodulation reference signals, different phase rotations (also, for SRS, typically referred to as “cyclic shifts”) can be used to generate different SRS that are orthogonal to each other. By assigning different phase rotations to different terminals, multiple SRS can thus be transmitted in parallel in the same subframe, as illustrated by terminals #1 and #2 in the upper part of Figure 11.17. However, it is then required that the reference signals span the same frequency band.

Another way to allow for SRS to be simultaneously transmitted from different terminals is to rely on the fact that each SRS only occupies every second subcarrier. Thus, SRS transmissions from two terminals can be *frequency multiplexed* by assigning them to different frequency shifts or “combs”, as illustrated by terminal #3 in the lower part of Figure 11.17. In contrast to the multiplexing of SRS transmission by means of different “cyclic shifts”, frequency multiplexing of SRS transmissions does not require the transmissions to cover identical frequency bands.

To summarize, the following set of parameters defines the characteristics of an SRS transmission:

- SRS transmission bandwidth – that is, the bandwidth covered by a single SRS transmission.
- Hopping bandwidth – that is, the frequency band over which the SRS transmission is frequency hopping.
- Frequency-domain position – that is, the starting point of the SRS transmission in the frequency domain.
- Transmission comb, as illustrated in Figure 11.17.
- Phase rotation (or equivalently cyclic shift) of the reference-signal sequence.
- SRS transmission time-domain period (from 2 to 160 ms) and subframe offset.

A terminal that is to transmit SRS is configured with these parameters by means of higher layer (RRC) signaling. In addition, all terminals within a cell should be informed in what subframes SRS may be transmitted within the cell as, within these subframes, the “SRS symbol” should not be used for PUSCH transmission.

11.2.2.2 Aperiodic SRS Transmission

In contrast to periodic SRS, aperiodic SRS are *one-shot* transmissions, triggered by signaling on PDCCH as part of the scheduling grant. The frequency-domain structure of an aperiodic SRS

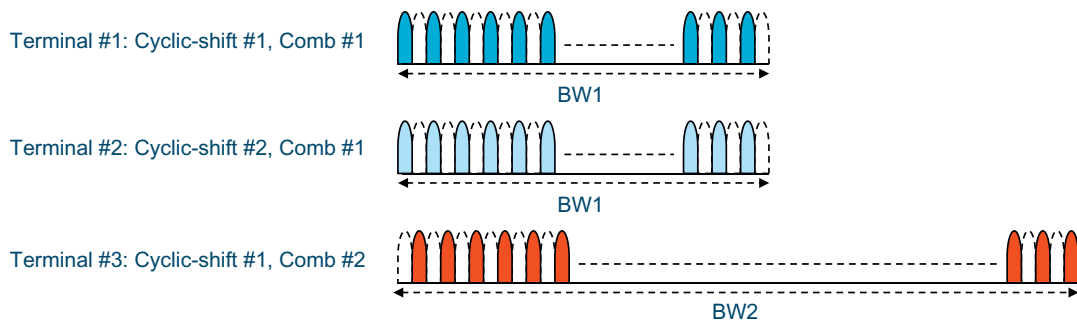


FIGURE 11.17

Multiplexing of SRS transmissions from different terminals.

transmission is identical to that of periodic SRS. Also, in the same way as for periodic SRS transmission, aperiodic SRS are transmitted within the last symbol of a subframe. Furthermore, the time instants when aperiodic SRS may be transmitted are configured per terminal using higher-layer signaling.

The frequency-domain parameters for aperiodic SRS (bandwidth, odd or even “comb”, etc.) are configured by higher-layer (RRC) signaling. However, no SRS transmission will actually be carried out until the terminal is explicitly triggered to do so by an explicit *SRS trigger* on PDCCH. When such a trigger is received, a single SRS is transmitted in the next available aperiodic SRS instant configured for the terminal using the configured frequency-domain parameters. Additional SRS transmissions can then be carried out if additional triggers are received.

Three different parameter sets can be configured for aperiodic SRS, for example differing in the frequency position of the SRS transmission and/or the transmission comb. Information on what parameters to use when the SRS is actually transmitted is included in the PDCH information, which consists of two bits, three combinations of which indicate the specific SRS parameter set. The fourth combination simply indicates that no SRS should be transmitted.

11.3 UPLINK MULTI-ANTENNA TRANSMISSION

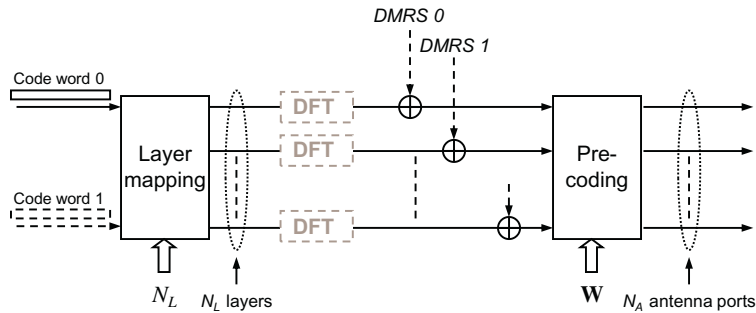
Downlink multi-antenna transmission was supported by the LTE specification from its first release (release 8). With LTE release 10, support for uplink multi-antenna transmission – that is, uplink transmission relying on multiple transmit antennas at the terminal side – was also introduced for LTE. Uplink multi-antenna transmission can be used to improve the uplink link and/or system performance in different ways:

- To improve the achievable data rates and spectral efficiency for uplink data transmission by allowing for antenna precoding supporting spatial multiplexing with up to four layers for the uplink physical data channel PUSCH.
- To improve the uplink control-channel performance by allowing for transmit diversity for the uplink physical control channel PUCCH.

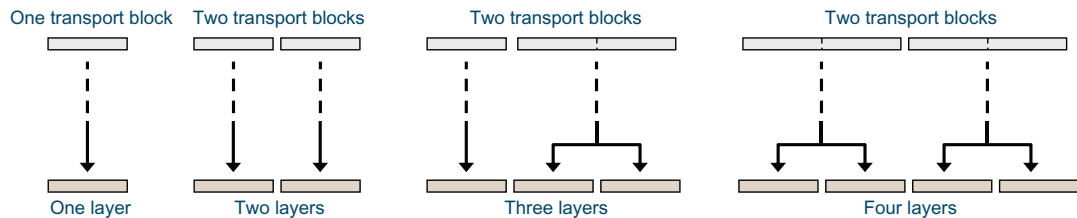
11.3.1 Precoder-Based Multi-Antenna Transmission for PUSCH

In the case of PUSCH transmission, multi-antenna transmission is based on antenna precoding, as illustrated in [Figure 11.18](#). The structure of the uplink antenna precoding is very similar to that of downlink antenna precoding (Section 10.3), including the presence of precoded demodulation reference signals (one per layer) similar to downlink non-codebook-based precoding ([Figure 10.19](#) in Section 10.3.3). Uplink antenna precoding supports transmission using up to four antenna ports, allowing for spatial multiplexing with up to four layers.

The principles for mapping of the modulation symbols to layers are also the same as for the downlink. For an initial transmission, there is one transport block in the case of a single layer and two transport blocks for more than one layer, as illustrated in [Figure 11.19](#). Similar to the downlink, in the case of a hybrid-ARQ retransmission, a single transport block may also be transmitted on multiple layers in some cases.

**FIGURE 11.18**

Precoder-based multi-antenna transmission for LTE uplink.

**FIGURE 11.19**

Uplink transport-channel-to-layer mapping (initial transmission).

As can be seen in Figure 11.18, the DFT precoding is actually taking place after layer mapping – that is, each layer is separately DFT precoded. To simplify the description this was not really visible in Figure 11.1, outlining the overall physical-layer transport-channel processing.

It can also be noted that, in contrast to Figure 10.19 in Section 10.3.3, the precoder in Figure 11.18 is not shaded. As discussed in Section 10.3.3, for downlink non-codebook-based precoding, the precoder part of the antenna mapping is not visible in the specification and the network can, in essence, apply an arbitrary precoding for the downlink transmission. Due to the use of precoded demodulation reference signals, the terminal can recover the different layers without knowledge of exactly what precoding has been applied at the transmitted side.

The same is also true for the uplink – that is, the presence of precoded demodulation reference signals would allow for the base station to demodulate the uplink multi-antenna transmission and recover the different layers without knowledge of the precoding taking place at the transmitter side. However, for LTE the uplink precoder matrix is selected by the network and conveyed to the terminal as part of the scheduling grant. The terminal should then follow the precoder matrix selected by the network. Thus, in the uplink, the precoder is visible in the specification and, in order to limit the downlink signaling, there is a limited set of precoder matrices specified for each transmission rank.

More specifically, for each combination of transmission rank N_L and number of antennas ports N_A , a set of precoder matrices of size $N_A \times N_L$ is defined, as illustrated in Tables 11.1 and 11.2 for

Table 11.1 Uplink Precoder Matrices for Two Antenna Ports

Transmission Rank	Codebook Index					
	0	1	2	3	4	5
1	$\begin{bmatrix} +1 \\ +1 \end{bmatrix}$	$\begin{bmatrix} +1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} +1 \\ +1 \end{bmatrix}$	$\begin{bmatrix} +1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} +1 \\ +1 \end{bmatrix}$	$\begin{bmatrix} +1 \\ -1 \end{bmatrix}$

Full-rank (2×2) identity matrix is not shown in the table.

Table 11.2 Subset of Uplink Precoder Matrices for Four Antenna Ports and Different Transmission Ranks

Transmission Rank	Codebook Index				
	0	1	2	3	...
1	$\begin{bmatrix} +1 \\ +1 \\ +1 \\ -1 \end{bmatrix}$	$\begin{bmatrix} +1 \\ +1 \\ +j \\ +j \end{bmatrix}$	$\begin{bmatrix} +1 \\ +1 \\ -1 \\ +1 \end{bmatrix}$	$\begin{bmatrix} +1 \\ +1 \\ -j \\ -j \end{bmatrix}$	...
2	$\begin{bmatrix} +1 & 0 \\ +1 & 0 \\ 0 & +1 \\ 0 & -j \end{bmatrix}$	$\begin{bmatrix} +1 & 0 \\ +1 & 0 \\ 0 & +1 \\ 0 & +j \end{bmatrix}$	$\begin{bmatrix} +1 & 0 \\ -j & 0 \\ 0 & +1 \\ 0 & +1 \end{bmatrix}$	$\begin{bmatrix} +1 & 0 \\ -j & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}$	
3	$\begin{bmatrix} +1 & 0 & 0 \\ +1 & 0 & 0 \\ 0 & +1 & 0 \\ 0 & 0 & +1 \end{bmatrix}$	$\begin{bmatrix} +1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & +1 & 0 \\ 0 & 0 & +1 \end{bmatrix}$	$\begin{bmatrix} +1 & 0 & 0 \\ 0 & +1 & 0 \\ +1 & 0 & 0 \\ 0 & 0 & +1 \end{bmatrix}$	$\begin{bmatrix} +1 & 0 & 0 \\ 0 & +1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & +1 \end{bmatrix}$	

Full-rank (4×4) identity matrix is not shown in the table.

two and four antenna ports respectively. For full-rank transmission – that is, when the transmission rank or number of layers equals the number of transmit antennas – only a single precoder matrix is defined, namely the identity matrix of size $N_A \times N_A$ (not shown in the tables). Note that, for the case of four antenna ports, only a subset of the defined matrices is shown. In total there are 24 rank-1 matrices, 16 rank-2 matrices, and 12 rank-3 matrices defined for four antenna ports, in addition to the single rank-4 matrix.

As can be seen, all the precoder matrices in Table 11.1 contain one and only one non-zero element in each row, and this is generally true for all precoder matrices defined for the uplink. As a consequence, the signal transmitted on a certain antenna port (corresponding to a certain row of the

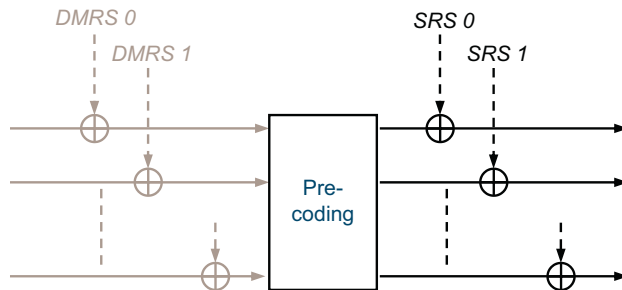
**FIGURE 11.20**

Illustration of SRS transmitted *after* uplink antenna precoding.

precoder matrix) always depends on one and only one specific layer (corresponding to a specific column of the precoder matrix). Expressed alternatively, the precoder matrix maps the layers to the antenna ports *with at most one layer being mapped to each antenna port*. Due to this, the good cubic-metric properties of the transmitted signal are also preserved for each antenna port when antenna precoding is applied. The precoder matrices of [Tables 11.1](#) and [11.2](#) are therefore also referred to as *cubic-metric-preserving precoder matrices*.

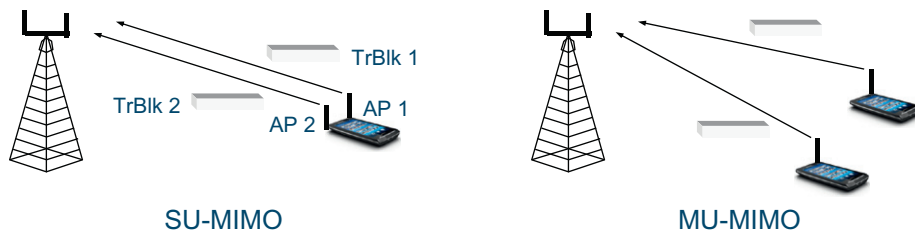
In order to select a suitable precoder, the network needs information about the uplink channel. Such information can, for example, be based on measurements on the uplink sounding reference signals (Section 11.2.2). As indicated in [Figure 11.20](#), sounding reference signals (SRS) are transmitted non-precoded – that is, directly on the different antenna ports. The received SRS thus reflect the channel of each antenna port, not including any precoding. Based on the received SRS, the network can thus decide on a suitable uplink transmission rank and corresponding uplink precoder matrix and provide information about the selected rank and precoder matrix as part of the scheduling grant.

The previous paragraph assumed the same number of antenna ports for PUSCH as for SRS. Clearly, this is a relevant situation and the SRS is, in this case, used to aid the selection of the precoding matrix, as discussed above. However, there are also situations when SRS and PUSCH use *different* numbers of antenna ports. One example is uplink transmission of two layers (two antenna ports), where the eNodeB would like to use SRS to probe the channel for potential four-layer transmission. In this case the SRS is transmitted on a *different* set of antenna ports than the PUSCH to aid the eNodeB in assessing the benefits, if any, of switching to four-layer transmission.

11.3.2 Uplink Multi-User MIMO

As described in Section 10.3.5, downlink multi-user MIMO (MU-MIMO) implies downlink transmission to different terminals using *the same time–frequency resource* and relying on the availability of multiple antennas, at least on the network side, to suppress interference between the transmissions. The term MU-MIMO originated from the resemblance to SU-MIMO (spatial multiplexing).

Uplink MU-MIMO is essentially the same thing but for the uplink transmission direction – that is, uplink MU-MIMO implies uplink transmissions from multiple terminals using *the same uplink time–frequency resource* and relying on the availability of multiple receive antennas at the base station

**FIGURE 11.21**

Relation between uplink SU-MIMO (spatial multiplexing) and MU-MIMO (SDMA).

to separate the two or more transmissions. Thus, MU-MIMO is really just another term for uplink *Space-Division Multiple Access* (SDMA).

Actually, on the uplink, the relation between MU-MIMO and SU-MIMO (spatial multiplexing) is even closer. Uplink spatial multiplexing, for example with two antenna ports and two layers, implies that the terminal transmits two transport blocks with one transport block transmitted on each layer and thus on each antenna port,¹² as illustrated in the left part of Figure 11.21. As illustrated in the right part of the figure, MU-MIMO is essentially equivalent to separating the two antennas into two different terminals and transmitting one transport block from each terminal. The base-station processing to separate the two transmissions could essentially be identical to the processing used to separate the two layers in the case of spatial multiplexing. It should be noted though that the separation of the two transmissions at the receiver side could be simplified, or at least the possible means to achieve this separation are extended, if the two terminals are well separated in space, something which is, for obvious reasons, not the case for two antennas attached to the same terminal. As an example, for sufficiently separated terminals, classical beam-forming relying on correlated receiver antennas can be used to separate the uplink transmissions. Alternatively, uncorrelated receiver antennas can be used, and the separation means are then essentially the same as for SU-MIMO.

One important benefit of uplink MU-MIMO is that one can get similar gains in *system throughput* as SU-MIMO (spatial multiplexing) without the need for multiple transmit antennas at the terminal side, allowing for less complex terminal implementation. It should be noted though that spatial multiplexing could still provide substantial gains in terms of *user throughput* and peak data rates that can be provided from a single terminal. The potential system gains of uplink MU-MIMO also rely on more than one terminal being actually available for transmission. The process of “pairing” terminals that should share the time–frequency resources is non-trivial and requires suitable radio-channel conditions.

Essentially, support for uplink MU-MIMO only requires the possibility to explicitly assign a specific orthogonal reference signal for the uplink transmission, thereby ensuring orthogonality between reference-signal transmissions from the different terminals involved in the MU-MIMO transmission. The orthogonal reference signals used for MU-MIMO are the same as for spatial multiplexing described in the previous section – that is, based on a combination of OCC (orthogonal cover codes)

¹²Note that the 2×2 precoder matrix is the identity matrix (see Table 11.1).

and different phase rotations. To support uplink MU-MIMO, each scheduling assignment includes a three-bit reference-signal indicator denoting a certain combination of OCC and phase rotation.

Also note that, similar to downlink MU-MIMO, nothing prevents the combination of spatial multiplexing and MU-MIMO – that is, multiplexing several terminals by means of MU-MIMO where one or several of the terminals are transmitting more than one layer. The eventual limitation is the number of orthogonal reference signals that can be assigned to the different users.

11.3.3 PUCCH Transmit Diversity

Precoder-based multi-layer transmission is only used for the uplink data transmission on PUSCH. However, in the case of a terminal with multiple transmit antennas, one obviously wants to use the full set of terminal antennas and corresponding terminal power amplifiers also for the L1/L2 control signaling on PUCCH in order to be able to utilize the full power resource and achieve maximum diversity. To achieve additional diversity, LTE release 10 also introduced the possibility for two-antenna *transmit diversity* for PUCCH. More specifically, the transmit diversity supported for PUCCH is referred to as Spatial Orthogonal-Resource Transmit Diversity (SORTD).

The basic principle of SORTD is simply to transmit the uplink control signaling using different resources (time, frequency, and/or code) on the different antennas. In essence, the PUCCH transmissions from the two antennas will be identical to PUCCH transmissions from two different terminals using different resources. Thus, SORTD creates additional diversity but achieves this by using twice as many PUCCH resources, compared to non-SORTD transmission.

For four physical antennas at the terminal, implementation-specific *antenna virtualization* is used. In essence, a transparent scheme is used to map the two-antenna-port signal to four physical antennas.

11.4 UPLINK L1/L2 CONTROL SIGNALING

Similar to the LTE downlink, there is also a need for uplink L1/L2 control signaling to support the transmission of downlink and uplink transport channels. Uplink L1/L2 control signaling consists of:

- hybrid-ARQ acknowledgements for received DL-SCH transport blocks;
- channel-state reports related to the downlink channel conditions, used to assist downlink scheduling; and
- scheduling requests, indicating that a terminal needs uplink resources for UL-SCH transmissions.

There is no information indicating the UL-SCH transport format signaled on the uplink. As mentioned in Chapter 8, the eNodeB is in complete control of the uplink UL-SCH transmissions and the terminal always follows the scheduling grants received from the network, including the UL-SCH transport format specified in those grants. Thus, the network knows the transport format used for the UL-SCH transmission in advance and there is no need for any explicit transport-format signaling on the uplink.

Uplink L1/L2 control signaling needs to be transmitted on the uplink regardless of whether or not the terminal has any uplink transport-channel (UL-SCH) data to transmit and thus regardless of whether or not the terminal has been assigned any uplink resources for UL-SCH transmission. Hence, two different methods are supported for the transmission of the uplink L1/L2 control signaling,

depending on whether or not the terminal has been assigned an uplink resource for UL-SCH transmission:

- *No simultaneous transmission of UL-SCH and L1/L2 control.* If the terminal does not have a valid scheduling grant – that is, no resources have been assigned for the UL-SCH in the current subframe – a separate physical channel, the *Physical Uplink Control Channel* (PUCCH), is used for transmission of uplink L1/L2 control signaling. In the case of uplink carrier aggregation, the PUCCH can be transmitted on the primary component carrier only. A consequence of this is that transmissions on multiple component carriers in the downlink may need to be acknowledged on a single uplink component carrier.
- *Simultaneous transmission of UL-SCH and L1/L2 control.* If the terminal has a valid scheduling grant – that is, resources have been assigned for the UL-SCH in the current subframe – the uplink L1/L2 control signaling is time multiplexed with the coded UL-SCH on to the PUSCH prior to DFT precoding and OFDM modulation. Obviously, as the terminal has been assigned UL-SCH resources, there is no need to support transmission of the scheduling request in this case. Instead, scheduling information can be included in the MAC headers, as described in Chapter 12. In the case of carrier aggregation, all control signaling is transmitted on one of the uplink component carriers – that is, uplink control signaling cannot be spread across multiple UL-SCHs.

The reason to differentiate between the two cases above is to minimize the cubic metric for the uplink power amplifier in order to maximize coverage. However, in situations when there is sufficient power available in the terminal, simultaneous transmission of PUSCH and PUCCH can be used with no impact on the coverage. The possibility for simultaneous PUSCH and PUCCH transmission was therefore introduced in release 10 as one part of several features¹³ adding flexibility at the cost of a somewhat higher cubic metric. In situations where this cost is not acceptable, simultaneous PUSCH and PUCCH can always be avoided by using the basic mechanism introduced in the first version of LTE.

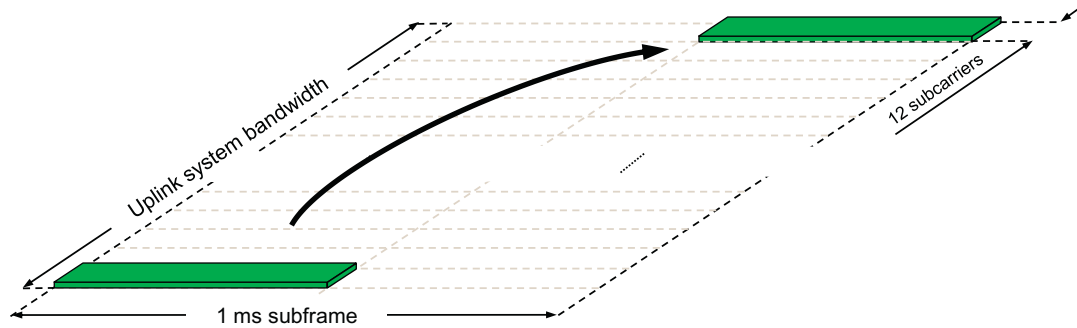
These two cases, control on PUCCH and PUSCH respectively, are described in more detail in the following.

11.4.1 Uplink L1/L2 Control Signaling on PUCCH

If the terminal has not been assigned an uplink resource for UL-SCH transmission, the L1/L2 control information (channel-state reports, hybrid-ARQ acknowledgements, and scheduling requests) is transmitted on uplink resources (resource blocks) specifically assigned for uplink L1/L2 control on PUCCH. As mentioned in the previous section, the PUCCH can be transmitted on the primary component carrier only. However, as the primary component carrier is specified on a per-terminal basis, there may, from a network perspective, be PUCCH resources used on multiple component carriers.

Figure 11.22 illustrates the location of the PUCCH resources on one component carrier at the edges of the bandwidth allocated to that component carrier. Each such resource consists of 12 subcarriers (one resource block) within each of the two slots of an uplink subframe. To provide frequency diversity, the frequency resource is frequency hopping on the slot boundary – that is, one “frequency resource” consists of 12 subcarriers at the upper part of the spectrum within the first slot of a subframe

¹³Other examples of such features are simultaneous transmission on multiple uplink component carriers and uplink multi-cluster transmission.

**FIGURE 11.22**

Uplink L1/L2 control signaling transmission on PUCCH.

and an equally sized resource at the lower part of the spectrum during the second slot of the subframe or vice versa. If more resources are needed for the uplink L1/L2 control signaling, for example in the case of very large overall transmission bandwidth supporting a large number of users, then additional resource blocks can be assigned next to the previously assigned resource blocks.

The reasons for locating the PUCCH resources at the edges of the overall available spectrum are twofold:

- Together with the frequency hopping described above, this maximizes the frequency diversity experienced by the control signaling.
- Assigning uplink resources for the PUCCH at other positions within the spectrum – that is, not at the edges – would have fragmented the uplink spectrum, making it impossible to assign very wide transmission bandwidths to a single terminal and still preserve the low-cubic-metric properties of the uplink transmission.

The bandwidth of one resource block during one subframe is too large for the control signaling needs of a single terminal. Therefore, to efficiently exploit the resources set aside for control signaling, multiple terminals can share the same resource block. This is done by assigning the different terminals to different orthogonal phase rotations of a cell-specific length-12 frequency-domain sequence, where the sequence is identical to a length-12 reference-signal sequence. Furthermore, as described in conjunction with the reference signals in Section 11.2, a linear phase rotation in the frequency domain is equivalent to applying a cyclic shift in the time domain. Thus, although the term “phase rotation” is used herein, the term cyclic shift is sometimes used with an implicit reference to the time domain.

The resource used by a PUCCH is therefore not only specified in the time–frequency domain by the resource-block pair, but also by the phase rotation applied. Similarly to the case of reference signals, there are up to 12 different phase rotations specified, providing up to 12 different orthogonal sequences from each cell-specific sequence. However, in the case of frequency-selective channels, not all the 12 phase rotations can be used if orthogonality is to be retained. Typically, up to six rotations are considered usable in a cell from a radio-propagation perspective, although inter-cell interference may result in a smaller number being useful from an overall system perspective. Higher-layer signaling is used to configure the number of rotations that is used in a cell.

As mentioned earlier, uplink L1/L2 control signaling includes hybrid-ARQ acknowledgements, channel-state reports, and scheduling requests. Different combinations of these types of messages are possible as described later, but to explain the structure for these cases it is beneficial to discuss separate transmission of each of the types first, starting with the hybrid-ARQ acknowledgement and the scheduling request. There are three formats defined for PUCCH, formats 1, 2, and 3, each capable of carrying a different number of bits.

11.4.1.1 PUCCH Format 1

Hybrid-ARQ acknowledgements are used to acknowledge the reception of one (or two in the case of spatial multiplexing) transport blocks on the DL-SCH. The hybrid-ARQ acknowledgement is only transmitted when the terminal correctly received control signaling related to DL-SCH transmission intended for this terminal on one of the PDCCHs. In the case of downlink carrier aggregation, there can be multiple simultaneous DL-SCHs for a single terminal, one per downlink component carrier and, consequently, multiple acknowledgement bits need to be conveyed in the uplink. This is described further below once the non-carrier-aggregated case has been treated.

If no valid DL-SCH-related control signaling is detected, then nothing is transmitted on the PUCCH (i.e. DTX). Apart from not unnecessarily occupying PUCCH resources that can be used for other purposes, this allows the eNodeB to perform three-state detection, ACK, NAK, or DTX, on the PUCCH received. Three-state detection is useful as NAK and DTX may need to be treated differently. In the case of NAK, retransmission of additional parity bits is useful for incremental redundancy, while for DTX the terminal has most likely missed the initial transmission of systematic bits and a better alternative than transmitting additional parity bits is to retransmit the systematic bits.

Scheduling requests are used to request resources for uplink data transmission. Obviously, a scheduling request should only be transmitted when the terminal is requesting resources, otherwise the terminal should be silent to save battery resources and not create unnecessary interference. Hence, unlike hybrid-ARQ acknowledgements, no explicit information bit is transmitted by the scheduling request; instead the information is conveyed by the presence (or absence) of energy on the corresponding PUCCH. However, the scheduling request, although used for a completely different purpose, shares the same PUCCH format as the hybrid-ARQ acknowledgement. This format is referred to as *PUCCH format 1* in the specifications.¹⁴

PUCCH format 1 uses the same structure in the two slots of a subframe, as illustrated in [Figure 11.23](#). For transmission of a hybrid-ARQ acknowledgement, the single hybrid-ARQ acknowledgement bit related to one downlink component carrier is used to generate a BPSK symbol (in the case of downlink spatial multiplexing the two acknowledgement bits are used to generate a QPSK symbol). For a scheduling request, the same constellation point as for a negative acknowledgement is used. The modulation symbol is then used to generate the signal to be transmitted in each of the two PUCCH slots.

There are seven OFDM symbols per slot for a normal cyclic prefix (six in the case of an extended cyclic prefix). In each of those seven OFDM symbols, a length-12 sequence, obtained by phase rotation of the cell-specific sequence as described earlier, is transmitted. Three of the symbols are used

¹⁴There are actually three variants in the LTE specifications, formats 1, 1A, and 1B, used for transmission of scheduling requests and one or two hybrid-ARQ acknowledgements respectively. However, for simplicity, they are all referred to as format 1 herein.

as reference signals to enable channel estimation by the eNodeB and the remaining four¹⁵ are modulated by the BPSK/QPSK symbols described earlier. In principle, the BPSK/QPSK modulation symbol could directly modulate the rotated length-12 sequence used to differentiate terminals transmitting on the same time–frequency resource. However, this would result in unnecessarily low capacity on the PUCCH. Therefore, the BPSK/QPSK symbol is multiplied by a length-4 orthogonal cover sequence.¹⁶ Multiple terminals may transmit on the same time–frequency resource using the same phase-rotated sequence and be separated through different orthogonal covers. To be able to estimate the channels for the respective terminals, the reference signals also employ an orthogonal cover sequence, with the only difference being the length of the sequence – three for the case of a normal cyclic prefix. Thus, since each cell-specific sequence can be used for up to $3 \cdot 12 = 36$ different terminals (assuming all 12 rotations are available; typically at most six of them are used), there is a threefold improvement in the PUCCH capacity compared to the case of no cover sequence. The cover sequences are three Walsh sequences of length 4 for the data part and three DFT sequences of length 3 for the reference signals (for an extended cyclic prefix, the reference-signal cover sequence is of length 2).

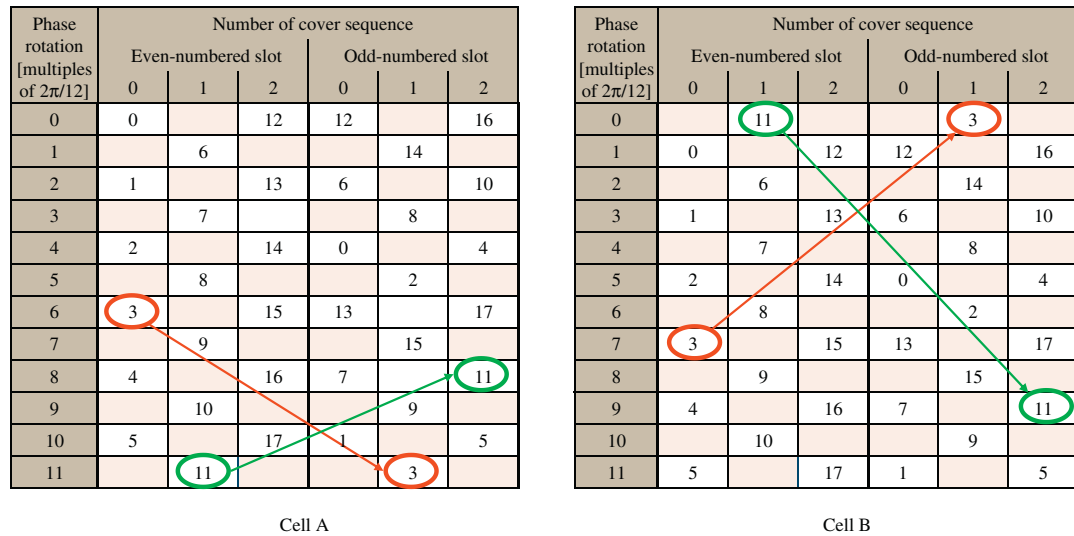
A PUCCH format 1 resource, used for either a hybrid-ARQ acknowledgement or a scheduling request, is represented by a single scalar resource index. From the index, the phase rotation and the orthogonal cover sequence are derived.

The use of a phase rotation of a cell-specific sequence together with orthogonal sequences as described earlier provides orthogonality between different terminals in the same cell transmitting PUCCH on the same set of resource blocks. Hence, in the ideal case, there will be no intra-cell interference, which helps improve the performance. However, there will typically be inter-cell interference for the PUCCH as the different sequences used in neighboring cells are non-orthogonal. To randomize the inter-cell interference, the phase rotation of the sequence used in a cell varies on a symbol-by-symbol basis in a slot according to a hopping pattern derived from the physical-layer cell identity. On top of this, slot-level hopping is applied to the orthogonal cover and phase rotation to further randomize the interference. This is exemplified in Figure 11.24 assuming normal cyclic prefix and six of 12 rotations used for each cover sequence. To the phase rotation given by the cell-specific hopping a slot-specific offset is added. In cell A, a terminal is transmitting on PUCCH resource number 3, which in this example corresponds to using the (phase rotation, cover sequence) combination (6, 0) in the first slot and (11, 1) in the second slot of this particular subframe. PUCCH resource number 11, used by another terminal in cell A transmitting in the same subframe, corresponds to (11, 1) and (8, 2) in the first and second slots respectively of the subframe. In another cell the PUCCH resource numbers are mapped to different sets (rotation, cover sequence) in the slots. This helps to randomize the inter-cell interference.

For an extended cyclic prefix, the same structure as in Figure 11.23 is used, with the difference being the number of reference symbols in each slot. In this case, the six OFDM symbols in each slot are divided such that the two middle symbols are used for reference signals and the remaining four symbols used for the information. Thus, the length of the orthogonal sequence used to spread the

¹⁵The number of symbols used for reference signals and the acknowledgement is a trade-off between channel-estimation accuracy and energy in the information part; three symbols for reference symbols and four symbols for the acknowledgement have been found to be a good compromise.

¹⁶In the case of simultaneous SRS and PUCCH transmissions in the same subframe, a length-3 sequence is used, thereby making the last OFDM symbol in the subframe available for the sounding reference signal.

**FIGURE 11.24**

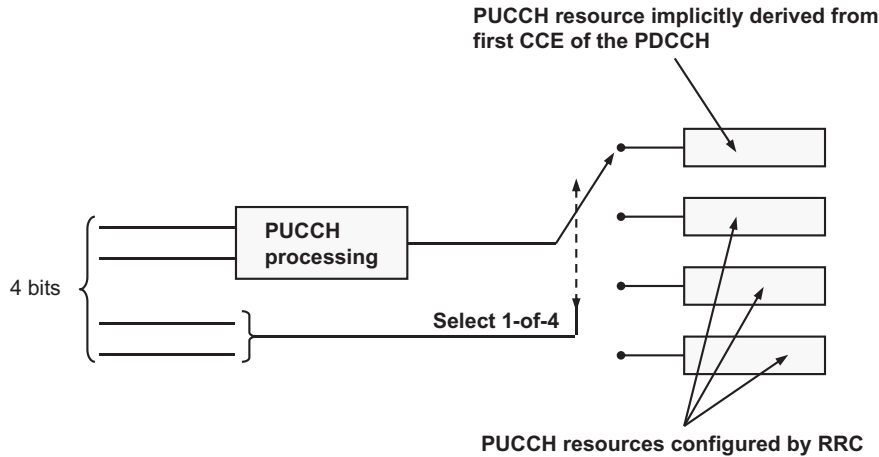
Example of phase rotation and cover hopping for two PUCCH resource indices in two different cells.

reference symbols is reduced from 3 to 2 and the multiplexing capacity is lower. However, the general principles described above still apply.

As mentioned earlier, a PUCCH resource can be represented by an index. For a hybrid-ARQ acknowledgement, the resource index to use is given as a function of the first CCE in the PDCCH used to schedule the downlink transmission to the terminal. In this way, there is no need to explicitly include information about the PUCCH resources in the downlink scheduling assignment, which of course reduces overhead. Furthermore, as described in Chapter 12, hybrid-ARQ acknowledgements are transmitted a fixed time after the reception of a DL-SCH transport block and when to expect a hybrid ARQ on the PUCCH is therefore known to the eNodeB.

In addition to dynamic scheduling by using the PDCCH, there is also, as described in Chapter 13, the possibility to semi-persistently schedule a terminal according to a specific pattern. In this case there is no PDCCH to derive the PUCCH resource index from. Instead, the configuration of the semi-persistent scheduling pattern includes information on the PUCCH index to use for the hybrid-ARQ acknowledgement. In either of these two cases, a terminal is using PUCCH resources only when it has been scheduled in the downlink. Thus, the amount of PUCCH resources required for hybrid-ARQ acknowledgements does not necessarily increase with an increasing number of terminals in the cell, but, for dynamic scheduling, is rather related to the number of CCEs in the downlink control signaling.

The description above was mainly concerned with the case of downlink carrier aggregation not being used. For downlink carrier aggregation, there can be multiple simultaneous DL-SCHs scheduled for a single terminal, one per downlink component carrier and, consequently, multiple acknowledgement bits need to be conveyed in the uplink (one, or two in the case of spatial multiplexing, per

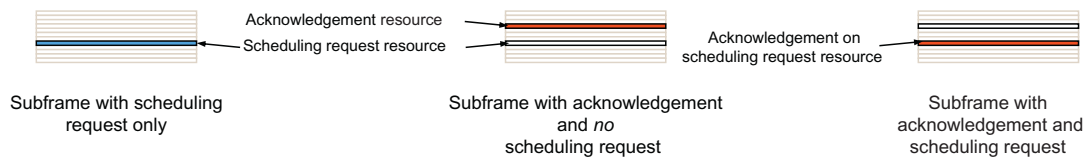
**FIGURE 11.25**

Resource selection for carrier aggregation.

downlink component carrier). PUCCH format 1 can be used to support more than two bits in the uplink by using resource selection. Assume four bits are to be transmitted in the uplink. With resource selection, two bits indicate which PUCCH resource to use while the remaining two bits are transmitted using the normal PUCCH structure but on the resource pointed to by the first two bits. This is illustrated in Figure 11.25. In total, four PUCCH resources are needed. One resource is derived from the first CCE using the same rule as in the absence of carrier aggregation (assuming that the scheduling assignment is transmitted on, and relating to, the primary component carrier). The remaining resources are semi-statically configured by RRC signaling.

For more than four bits, resource selection is less efficient and PUCCH format 3, described in a subsequent section, was therefore introduced in LTE release 10 in conjunction with carrier aggregation.

Unlike the hybrid-ARQ acknowledgements, whose occurrence is known to the eNodeB from the downlink scheduling decisions, the need for uplink resources for a certain terminal is in principle unpredictable by the eNodeB. One way to handle this would be to have a contention-based mechanism for requesting uplink resources. The random-access mechanism is based on this principle and can, to some extent, also be used for scheduling requests, as discussed in Chapter 13. Contention-based mechanisms typically work well for low intensities, but for higher scheduling-request intensities, the collision rate between different terminals simultaneously requesting resources becomes too large. Therefore, LTE provides a contention-free scheduling-request mechanism on the PUCCH, where each terminal in the cell is given a reserved resource on which it can transmit a request for uplink resources. The contention-free resource is represented by a PUCCH resource index as described earlier, occurring at every n th subframe. The more frequently these time instants occur, the lower the scheduling-request delay at the cost of higher PUCCH resource consumption. As the eNodeB configures all the terminals in the cell, when and on which resources a terminal can request resources is known to the eNodeB. A single scheduling request resource is also sufficient for the case

**FIGURE 11.26**

Multiplexing of scheduling request and hybrid-ARQ acknowledgement from a single terminal.

of carrier aggregation, as it only represents a request for uplink resources, which is independent of whether carrier aggregation is used or not.

The discussion earlier has considered transmission of *either* a hybrid-ARQ acknowledgement *or* a scheduling request. However, there are situations when the terminal needs to transmit *both* of them. In such a situation, the hybrid-ARQ acknowledgement is transmitted on the scheduling-request resource (see Figure 11.26). This is possible as the same PUCCH structure is used for both of them and the scheduling request carries no explicit information. By comparing the amount of energy detected on the acknowledgement resource and the scheduling-request resource for a specific terminal, the eNodeB can determine whether the terminal is requesting resources or not. Once the resource used for transmission of the acknowledgement is detected, the hybrid-ARQ acknowledgement can be decoded. Other, more advanced methods jointly decoding hybrid ARQ and scheduling request can also be envisioned.

11.4.1.2 PUCCH Format 2

Channel-state reports, the contents of which are discussed in Chapter 13, are used to provide the eNodeB with an estimate of the channel properties as seen from the terminal to aid channel-dependent scheduling. A channel-state report consists of multiple bits transmitted in one subframe. PUCCH format 1, which is capable of at most two bits of information per subframe, obviously cannot be used for this purpose. Transmission of channel-state reports on the PUCCH is instead handled by PUCCH format 2, which is capable of multiple information bits per subframe.¹⁷

PUCCH format 2, illustrated for a normal cyclic prefix in Figure 11.27, is based on a phase rotation of the same cell-specific sequence as format 1. After block coding using a punctured Reed–Müller code and QPSK modulation of the channel-state information, there are 10 QPSK symbols to transmit in the subframe: the first five symbols are transmitted in the first slot and the remaining five in the last slot.

Assuming a normal cyclic prefix, there are seven OFDM symbols per slot. Of the seven OFDM symbols in each slot, two¹⁸ are used for reference-signal transmission to allow coherent demodulation at the eNodeB. In the remaining five, the respective QPSK symbol to be transmitted is multiplied

¹⁷There are actually three variants in the LTE specifications, formats 2, 2A, and 2B, where the last two formats are used for simultaneous transmission of channel-state reports and hybrid-ARQ acknowledgements, as discussed later in this section. However, for simplicity, they are all referred to as format 2 herein.

¹⁸Similarly to format 1, the number of symbols used for reference signals and the coded channel-quality information is a trade-off between channel-estimation accuracy and energy in the information part. Two symbols for reference symbols and five symbols for the coded information part in each slot were found to be the best compromise.

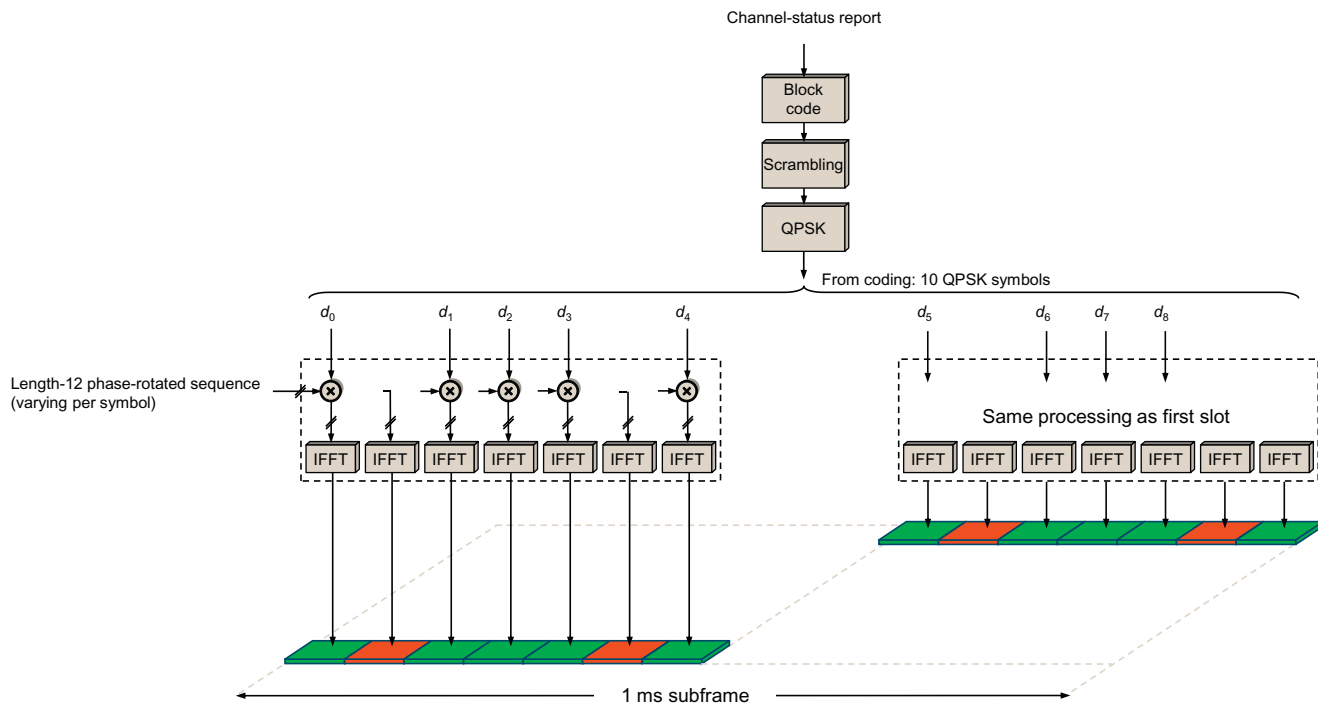


FIGURE 11.27

PUCCH format 2 (normal cyclic prefix).

by a phase-rotated length-12 cell-specific sequence and the result is transmitted in the corresponding OFDM symbol. For an extended cyclic prefix, where there are six OFDM symbols per slot, the same structure is used but with one reference-signal symbol per slot instead of two.

Basing the format 2 structure on phase rotations of the same cell-specific sequence as format 1 is beneficial as it allows the two formats to be transmitted in the same resource block. As phase-rotated sequences are orthogonal, one rotated sequence in the cell can be used either for one PUCCH instance using format 2 or three PUCCH instances using format 1. Thus, the “resource consumption” of one channel-state report is equivalent to three hybrid-ARQ acknowledgements (assuming normal cyclic prefix). Note that no orthogonal cover sequences are used for format 2.

The rotation angles to use in the different symbols for PUCCH format 2 are hopping in a similar way as for format 1, motivated by interference randomization. Resources for PUCCH format 2 can, similar to format 1, be represented by a scalar index, which can be seen as a “channel number” and higher-layer signaling is used to configure each terminal with a resource to transmit its channel-state report on, as well as when those reports should be transmitted. Hence, the eNodeB has full knowledge of when and on which resources each of the terminals will transmit channel-state reports on PUCCH. The channel-state reports on PUCCH are also known as *periodic* reports; as will be discussed later, there are also *aperiodic* reports but those reports can only be transmitted on PUSCH.

Periodic CSI reports in combination with carrier aggregation are provided for by transmitting multiple periodic reports, one per component carrier, offset in time such that reports for two different component carriers do not collide.

11.4.1.3 PUCCH Format 3

For downlink carrier aggregation, multiple hybrid-ARQ acknowledgement bits need to be fed back in the case of simultaneous transmission on multiple component carriers. As discussed above, PUCCH format 1 in combination with resource selection can be used, but for more than four bits this is not an efficient solution from a performance perspective. To enable the possibility of transmitting more than four bits in an efficient way, PUCCH format 3 can be used. A terminal capable of more than two downlink component carriers – that is, capable of more than four bits for hybrid-ARQ acknowledgements – needs to support PUCCH format 3. For such a terminal, PUCCH format 3 can also be used for less than four bits of feedback relating to simultaneous transmission on multiple component carriers if configured by higher-layer signaling not to use PUCCH format 1 with resource selection.

The basis for PUCCH format 3, illustrated in Figure 11.28, is DFT-precoded OFDM – that is, the same transmission scheme as used for UL-SCH. The acknowledgement bits, one or two per downlink component carrier depending on the transmission mode configured for that particular component carrier, are concatenated with a scheduling request bit (if present) into a sequence of bits where bits corresponding to unscheduled transport blocks are set to zero. Block coding is applied,¹⁹ followed by scrambling using a cell-specific scrambling sequence to randomize inter-cell interference. The resulting 48 bits are QPSK-modulated and divided into two groups, one per slot, of 12 QPSK symbols each.

Assuming a normal cyclic prefix, there are seven OFDM symbols per slot. Similarly to PUCCH format 2, two OFDM symbols (one in the case of an extended cyclic prefix) in each slot are used for reference signal transmission, leaving five symbols for data transmission. In each slot, the block of 12 DFT-precoded QPSK symbols is transmitted in the five available DFTS-OFDM symbols. To further

¹⁹A (32, k) Reed–Müller code is used, but for 12 or more bits in TDD two Reed–Müller codes are used in combination.

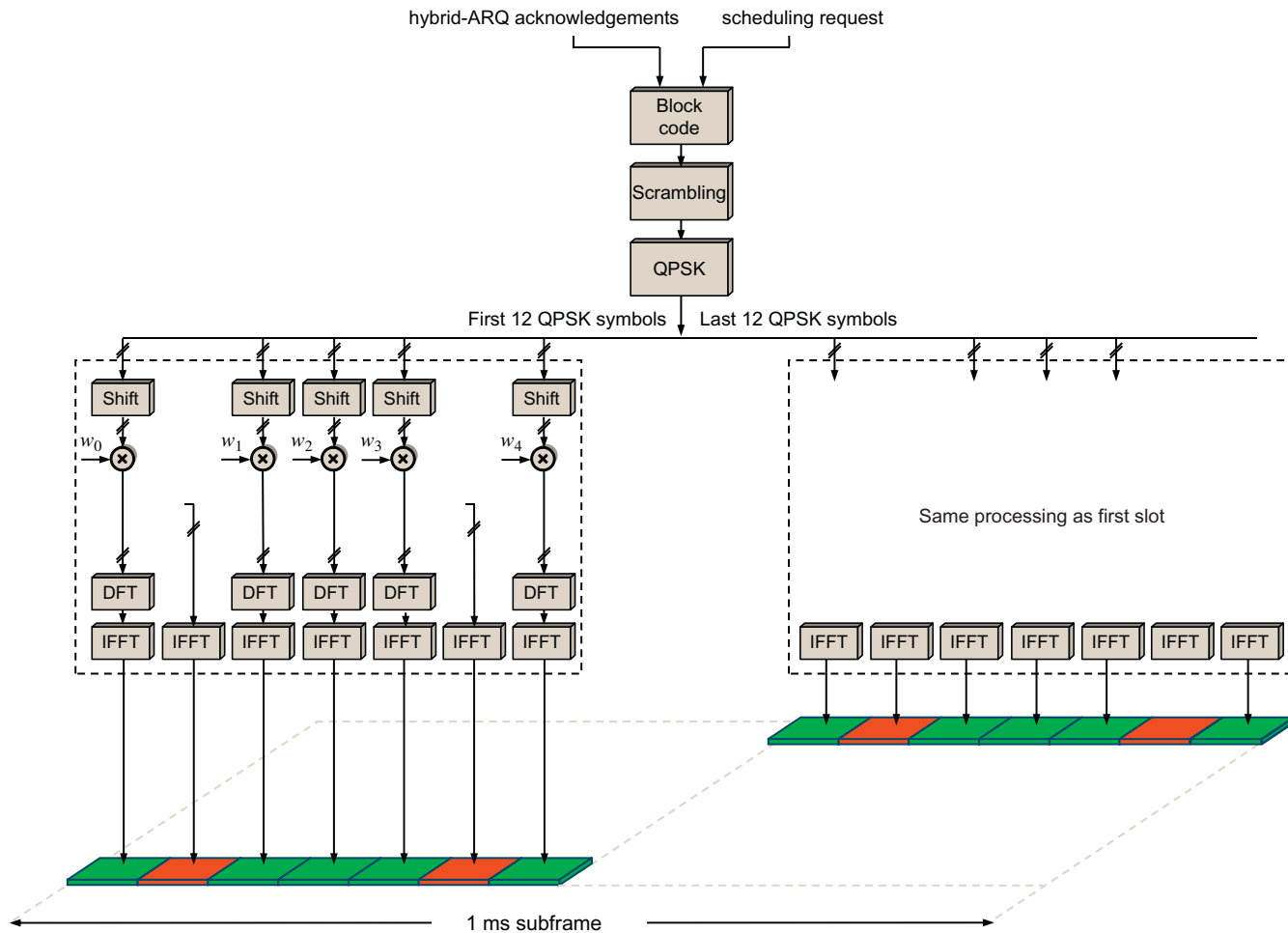


FIGURE 11.28

PUCCH format 3 (normal cyclic prefix).

randomize the inter-cell interference, a cyclic shift of the 12 inputs to the DFT, varying between OFDM symbols in a cell-specific manner, is applied to the block of 12 QPSK symbols prior to DFT precoding.

To increase the multiplexing capacity, a length-5 orthogonal sequence is used with each of the five OFDM symbol carrying data in a slot being multiplied by one element of the sequence. Thus, up to five terminals may share the same resource-block pair for PUCCH format 3. Different length-5 sequences are used in the two slots to improve the performance in high-Doppler scenarios. To facilitate channel estimation for the different transmissions sharing the same resource block, different reference-signal sequences are used.

The length-5 orthogonal cover sequences are obtained as five DFT sequences. There is also the possibility to use a length-4 Walsh sequence for the second slot in order to leave the last OFDM symbol unused for the case when sounding is configured in the subframe.

In the same manner as for the other two PUCCH formats, a resource can be represented by a single index from which the orthogonal sequence and the resource-block number can be derived. A terminal can be configured with four different resources for PUCCH format 3. In the scheduling grant for a secondary carrier, two bits²⁰ inform the terminal which of the four resources to use. In this way, the scheduler can avoid PUCCH collisions between different terminals by assigning them to different resources.

Note that, due to the differences in the underlying structure of PUCCH format 3 compared to the other two formats, resource blocks cannot be shared between format 3 and the other two formats.

11.4.1.4 Simultaneous Transmission of Multiple Feedback Reports

The above description has focused on transmission of hybrid-ARQ acknowledgements or channel-state reports alone. However, there is also a need to handle the case when hybrid ARQ or scheduling requests need to be transmitted in the same subframe. Transmitting multiple PUCCHs simultaneously from the same terminal would increase the cubic metric. Therefore, a single PUCCH structure supporting simultaneous transmission of multiple feedback signals is used.

In principle, the following four situations requiring simultaneous transmission of multiple feedback signals from a single terminal can be envisioned:

1. *Hybrid-ARQ acknowledgement and scheduling request.* For PUCCH format 1, this is supported by transmitting the hybrid-ARQ acknowledgement on the scheduling-request resource as described earlier. For PUCCH format 3, the transmission structure supports inclusion of a scheduling-request bit.
2. *Scheduling request and channel-state report.* The eNodeB is in control of when a terminal may transmit a scheduling request and when the terminal should report the channel state. Hence, this situation can be avoided by proper configuration, but if this is not done the terminal drops the channel-state report and transmits the scheduling request only. Missing a channel-state report is not detrimental and only incurs some degradation in the scheduling and rate-adaptation accuracy, whereas the scheduling request is critical for uplink transmissions.

²⁰The power control bits in an assignment relating to a secondary component carrier are reinterpreted as an *acknowledgement resource indicator* (ARI) – see Chapter 10.

3. *Scheduling request, hybrid-ARQ acknowledgement, and channel-state report.* This is similar to the previous situation; the channel-state report is dropped and multiplexing of hybrid ARQ and scheduling request is handled as described in point 1 above.
4. *Hybrid-ARQ acknowledgement and channel-state report.* Simultaneous transmission of acknowledgement and channel-state report from the same terminal is possible, as described below. There is also the possibility to configure the terminal to drop the channel-state report and only transmit the acknowledgement. In the case of carrier aggregation, the channel-state report is always dropped if it happens to collide with PUCCH format 3.

Transmission of data in the downlink implies transmission of hybrid-ARQ acknowledgements in the uplink. At the same time, since data is transmitted in the downlink, up-to-date channel-state reports are beneficial to optimize the downlink transmissions. Hence, simultaneous transmission of hybrid-ARQ acknowledgements and channel-state reports is supported by LTE. The basis is PUCCH format 2 as described earlier for both normal and extended cyclic prefixes, although the detailed solution differs between the two.

For a normal cyclic prefix, each slot in PUCCH format 2 has two OFDM symbols used for reference signals. When transmitting a hybrid-ARQ acknowledgement at the same time as a channel-state report, the second reference signal in each slot is modulated by the acknowledgement, as illustrated in Figure 11.29a. Either BPSK or QPSK is used, depending on whether one or two acknowledgement bits are to be fed back. The fact that the acknowledgement is superimposed on the reference signal needs to be accounted for at the eNodeB. One possibility is to decode the acknowledgement bit(s) modulated on to the second reference symbol using the first reference symbol for channel estimation. Once the acknowledgement bit(s) have been decoded, the modulation imposed on the second reference symbol can be removed and channel estimation and decoding of the channel-state report can be handled in the same way as in the absence of simultaneous hybrid-ARQ acknowledgement. This two-step approach works well for low to medium Doppler frequencies; for higher Doppler frequencies the acknowledgement and channel-state reports are preferably decoded jointly.

For an extended cyclic prefix, there is only a single reference symbol per slot. Hence, it is not possible to overlay the hybrid-ARQ acknowledgement on the reference symbol. Instead, the acknowledgement bit(s) are jointly coded with the channel-state report prior to transmission using PUCCH format 2, as illustrated in Figure 11.29b.

The time instances when to expect channel-state reports and hybrid-ARQ acknowledgements are known to the eNodeB, which therefore knows whether to expect a hybrid-ARQ acknowledgement along with the channel-state report or not. If the PDCCH assignment is missed by the terminal, then only the channel-state report will be transmitted as the terminal is not aware that it has been scheduled. In the absence of a simultaneous channel-state report, the eNodeB can employ DTX detection to discriminate between a missed assignment and a failed decoding of downlink data. However, one consequence of the structures above is that DTX detection is cumbersome, if not impossible. This implies that incremental redundancy needs to be operated with some care if the eNodeB has scheduled data such that the acknowledgement occurs at the same time as a channel-state report. As the terminal may have missed the original transmission attempt in the downlink, it may be preferable for the eNodeB to select the redundancy version of the retransmission such that systematic bits are also included in the retransmission.

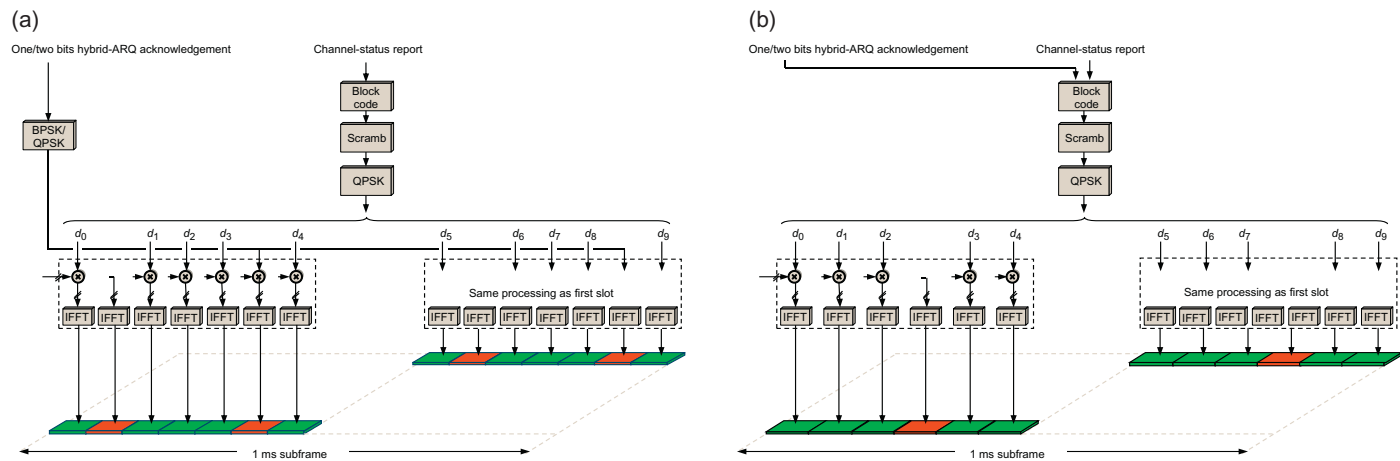


FIGURE 11.29

Simultaneous transmission of channel-state reports and hybrid-ARQ acknowledgements. (a) Normal cyclic prefix. (b) Extended cyclic prefix.

One possibility to circumvent this is to configure the terminal to drop the channel-state report in the case of simultaneous transmission of a hybrid-ARQ acknowledgement. In this case, the eNodeB can detect DTX as the acknowledgement can be transmitted using PUCCH format 1, as described earlier. Obviously, there will be no channel-state report sent in this case, which needs to be taken into account in the scheduling process. Dropping the channel-state report can also be useful for handling coverage-limited situations when the available transmission power is insufficient for a channel-state report in addition to the acknowledgement.

11.4.1.5 Resource-Block Mapping for PUCCH

The signals described above for all three of the PUCCH formats are, as already explained, transmitted on a resource-block pair with one resource block in each slot. The resource-block pair to use is determined from the PUCCH resource index. Multiple resource-block pairs can be used to increase the control-signaling capacity; when one resource-block pair is full, the next PUCCH resource index is mapped to the next resource-block pair in sequence. The mapping is in principle done such that PUCCH format 2 (channel-state reports) is transmitted closest to the edges of the uplink cell bandwidth with PUCCH format 1 (hybrid-ARQ acknowledgements, scheduling requests) next, as illustrated in Figure 11.30. The location of PUCCH format 3 is configurable and it can, for example, be located between formats 1 and 2. A semi-static parameter, provided as part of the system information, controls on which resource-block pair the mapping of PUCCH format 1 starts. Furthermore, the semi-statically configured scheduling requests are located at the outermost parts of the format 1 resources. As the amount of resources necessary for hybrid-ARQ acknowledgements varies dynamically, this maximizes the amount of contiguous spectrum available for PUSCH.

In many scenarios, the configuration of the PUCCH resources can be done such that the three PUCCH formats are transmitted on separate sets of resource blocks. However, for the smallest cell bandwidths, this would result in too high an overhead. Therefore, it is possible to mix the first two PUCCH formats in one of the resource-block pairs – for example, in Figure 11.30 this is the case for the resource with index 2. Although this mixture is primarily motivated by the smaller cell bandwidths, it can equally well be used for the larger cell bandwidths. In the resource-block pair where PUCCH formats 1 and 2 are mixed, the rotation angles are divided between the two formats. Furthermore,

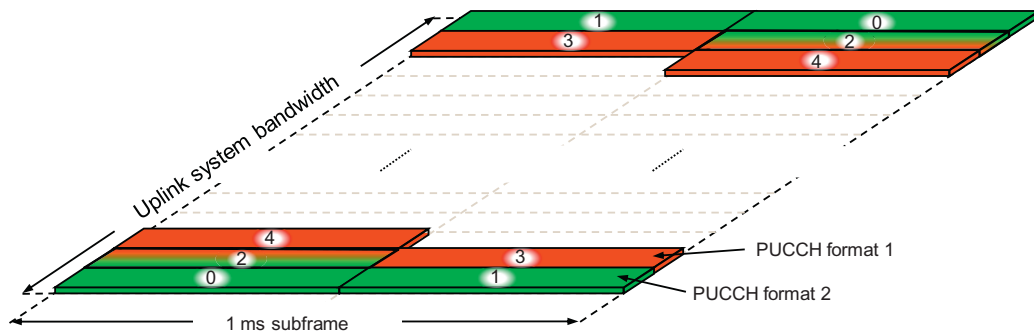


FIGURE 11.30

Allocation of resource blocks for PUCCH.

some of the phase rotations are reserved as “guard”; hence the efficiency of such a mixed resource-block pair is slightly lower than a resource-block pair carrying only one of the first two PUCCH formats.

11.4.2 Uplink L1/L2 Control Signaling on PUSCH

If the terminal is transmitting data on PUSCH – that is, has a valid scheduling grant in the subframe – control signaling is time multiplexed²¹ with data on the PUSCH instead of using the PUCCH (in release 10, simultaneous PUSCH and PUCCH can be used). Only hybrid-ARQ acknowledgements and channel-state reports are transmitted on the PUSCH. Obviously there is no need to request a scheduling grant when the terminal is already scheduled; instead, in-band buffer-status reports are sent as part of the MAC headers, as described in Chapter 12.

Time multiplexing of channel-state reports and hybrid-ARQ acknowledgements is illustrated in Figure 11.31. However, although they both use time multiplexing there are some differences in the details for the two types of uplink L1/L2 control signaling motivated by their different properties.

The hybrid-ARQ acknowledgement is important for proper operation of the downlink. For one and two acknowledgements, robust QPSK modulation is used, regardless of the modulation scheme used for the data, while for a larger number of bits the same modulation scheme as for the data is used. Furthermore, the hybrid-ARQ acknowledgement is transmitted near to the reference symbols as the channel estimates are of better quality close to the reference symbols. This is especially important at high Doppler frequencies, where the channel may vary during a slot. Unlike the data part, the hybrid-ARQ acknowledgement cannot rely on retransmissions and strong channel coding to handle these variations.

In principle, the eNodeB knows when to expect a hybrid-ARQ acknowledgement from the terminal and can therefore perform the appropriate demultiplexing of the acknowledgement and the data part. However, there is a certain probability that the terminal has missed the scheduling assignment on the PDCCH, in which case the eNodeB will expect a hybrid-ARQ acknowledgement while the terminal will not transmit one. If the rate-matching pattern were to depend on whether an acknowledgement is transmitted or not, all the coded bits transmitted in the data part could be affected by a missed PDCCH assignment, which is likely to cause the UL-SCH decoding to fail. To avoid this error, the hybrid-ARQ acknowledgements are therefore punctured into the coded UL-SCH bit stream. Thus, the non-punctured bits are not affected by the presence/absence of hybrid-ARQ acknowledgements and the problem of a mismatch between the rate matching in the terminal and the eNodeB is avoided.

The contents of the channel-state reports is described in Chapter 12; at this stage it suffices to note that a channel-state report consists of *Channel-Quality Indicator* (CQI), *Precoding Matrix Indicator* (PMI), and *Rank Indicator* (RI). The CQI and PMI are time multiplexed with the coded data bits from PUSCH and transmitted using the same modulation as the data part. Channel-state reports are mainly useful for low-to-medium Doppler frequencies for which the radio channel is relatively constant, hence the need for special mapping is less pronounced. The RI, however, is mapped differently than the CQI and PMI; as illustrated in Figure 11.31, the RI is located near to the reference symbols using a similar mapping as the hybrid-ARQ acknowledgements. The more robust mapping of

²¹In the case of spatial multiplexing, the CQI/PMI is time multiplexed with one of the codewords, implying that it is spatially multiplexed with the other codeword.

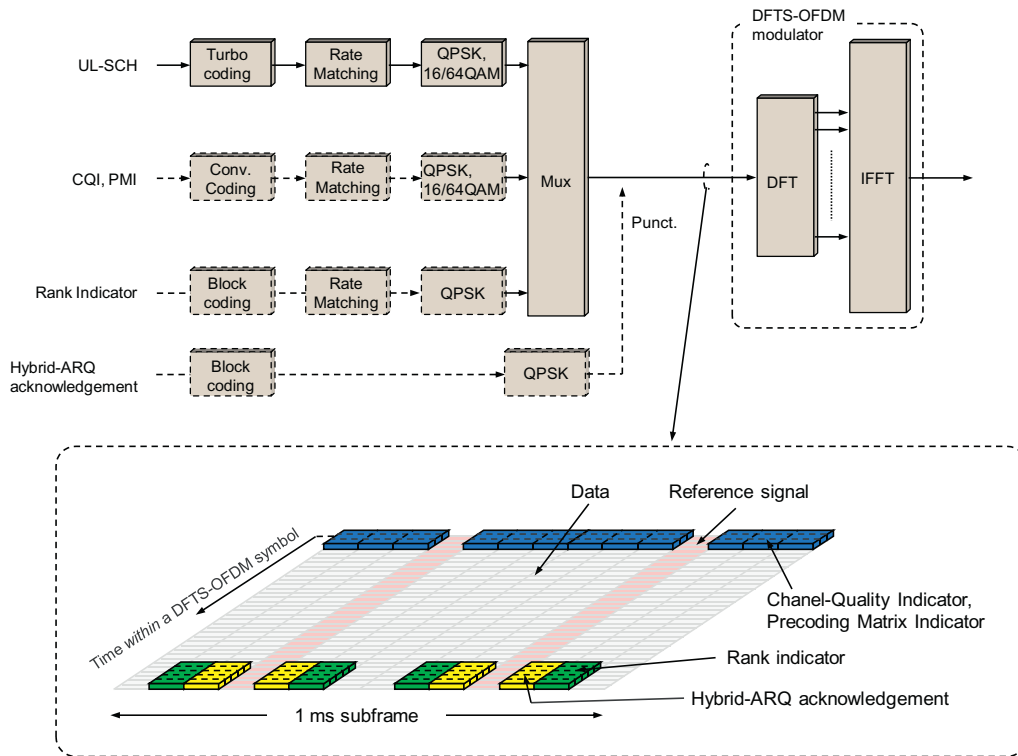


FIGURE 11.31

Multiplexing of control and data on to PUSCH. Note that there are *two* time dimensions and no frequency dimension in the “grid” at the bottom, the reason being that the signal prior to DFT precoding is illustrated. The acknowledgement (and the RI) will thus be located near but not as close as possible to the reference signal as possible as the “grid” is read column by column.

the RI is motivated by the fact that the RI is required in order to correctly interpret the CQI/PMI. The CQI/PMI, on the other hand, is simply mapped across the full subframe duration. Modulation-wise, the RI is handled in the same way as the hybrid-ARQ acknowledgements.

For uplink spatial multiplexing, in which case two transport blocks are transmitted simultaneously on the PUSCH, the CQI and PMI are multiplexed with the coded transport block using the highest modulation-and-coding scheme, followed by applying the multiplexing scheme described above per layer (Figure 11.32). The intention behind this approach is to transmit the CQI and PMI on the (one or two) layers with the best quality.²²

The hybrid-ARQ acknowledgements and the rank indicator are replicated across all transmission layers and multiplexed with the coded data in each layer in the same way as the single layer case

²² Assuming the MCS follows the channel quality, this holds for one, two, and four layers but not necessarily for three layers.

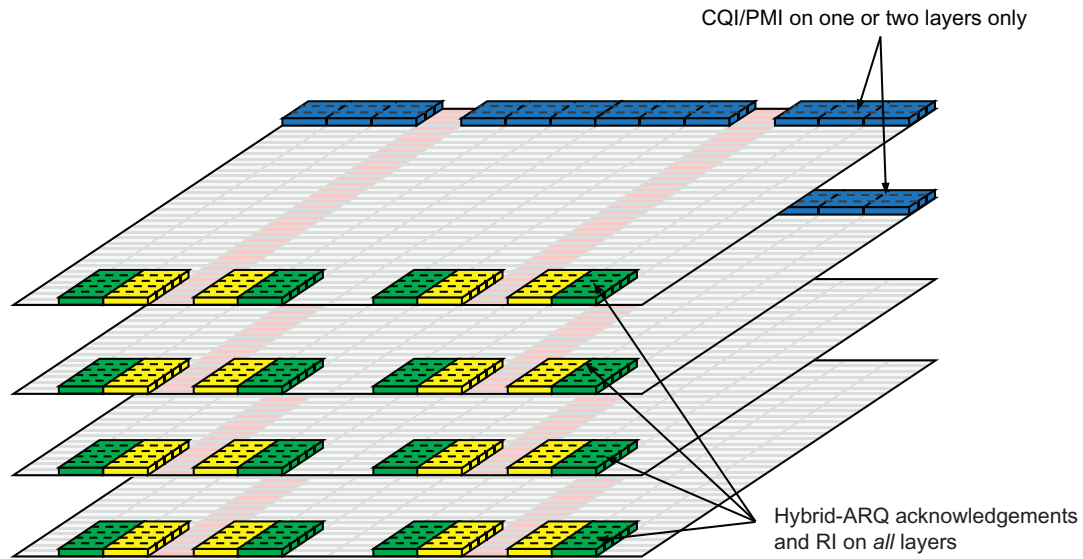


FIGURE 11.32

Multiplexing of CQI/PMI, RI, and hybrid-ARQ acknowledgements in the case of uplink spatial multiplexing.

described above. The bits may, though, have been scrambled differently on the different layers. In essence, as the same information is transmitted on multiple layers with different scrambling, this provides diversity.

The basis for channel-state reports on the PUSCH is *aperiodic reports*, where the eNodeB requests a channel-state report from the terminal by setting the CSI request bit in the scheduling grant, as mentioned in Chapter 10. UL-SCH rate matching takes the presence of the channel-state reports into account; by using a higher code rate a suitable number of resource elements is made available for transmission of the channel-state report. Since the reports are explicitly requested by the eNodeB, their presence is known and the appropriate rate de-matching can be done at the receiver. If one of the configured transmission instances for a periodic report coincides with the terminal being scheduled on the PUSCH, the periodic report is “rerouted” and transmitted on the PUSCH resources. Also, in this case there is no risk of mismatch in rate matching; the transmission instants for periodic reports are configured by robust RRC signaling and the eNodeB knows in which subframes such reports will be transmitted.

The channel coding of the channel-state reports depends on the report size. For the smaller sizes such as a periodic report that otherwise would have been transmitted on the PUCCH, the same block coding as used for the PUCCH reports is used. For the larger reports, a tail-biting convolutional code is used for CQI/PMI, whereas the RI uses a (3, 2) block code for a single component carrier.

Unlike the data part, which relies on rate adaptation to handle different radio conditions, this cannot be used for the L1/L2 control-signaling part. Power control could, in principle, be used as an alternative, but this would imply rapid power variations in the time domain, which negatively impact

the RF properties. Therefore, the transmission power is kept constant over the subframe and the amount of resource elements allocated to L1/L2 control signaling – that is, the code rate of the control signaling – is varied according to the scheduling decision for the data of the data part. High data rates are typically scheduled when the radio conditions are advantageous and hence a smaller amount of resource needs to be used by the L1/L2 control signaling compared to the case of poor radio conditions. To account for different hybrid-ARQ operating points, an offset between the code rate for the control-signaling part and the modulation-and-coding scheme used for the data part can be configured via higher-layer signaling.

For carrier aggregation, control signaling is time multiplexed on one uplink component carrier only – that is, uplink control information cannot be split across multiple uplink component carriers. Apart from the aperiodic CSI reports, which are transmitted upon the component carrier that triggered the report, the primary component carrier is used for uplink control signaling if scheduled in the same subframe, otherwise one of the secondary component carriers is used.

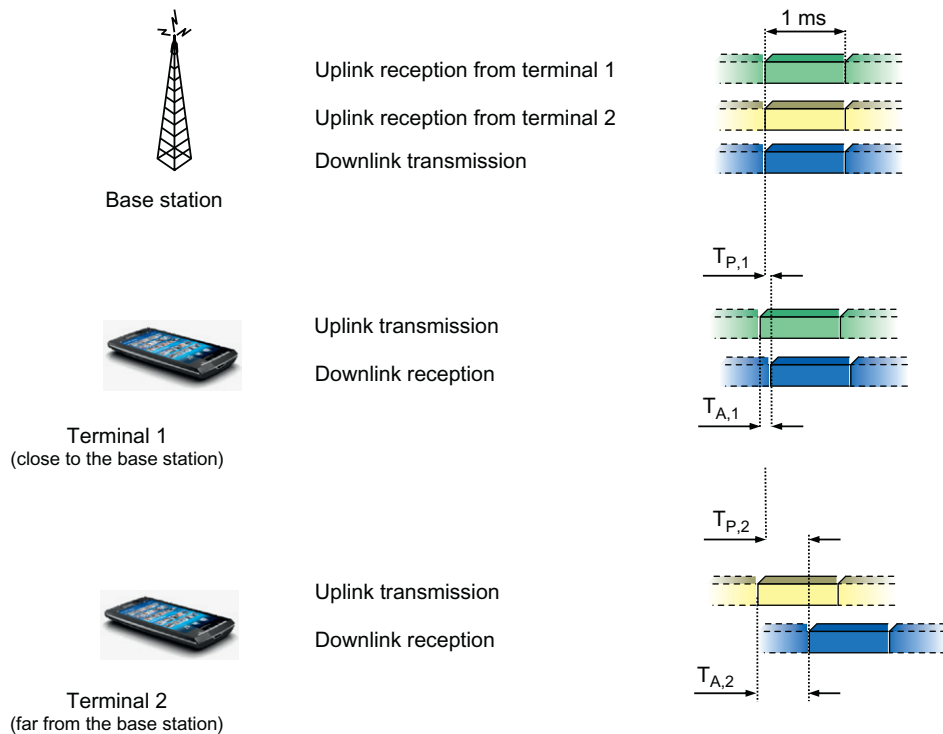
11.5 UPLINK TIMING ALIGNMENT

The LTE uplink allows for uplink intra-cell orthogonality, implying that uplink transmissions received from different terminals within a cell do not cause interference to each other. A requirement for this *uplink orthogonality* to hold is that the signals transmitted from different terminals within the same subframe but within different frequency resources (different resource blocks) arrive approximately time aligned at the base station. More specifically, any timing misalignment between received signals received should fall within the cyclic prefix. To ensure such receiver-side time alignment, LTE includes a mechanism for *transmit-timing advance*.

In essence, timing advance is a negative offset, at the terminal, between the start of a received downlink subframe and a transmitted uplink subframe. By controlling the offset appropriately for each terminal, the network can control the timing of the signals received at the base station from the terminals. Terminals far from the base station encounter a larger propagation delay and therefore need to start their uplink transmissions somewhat in advance, compared to terminals closer to the base station, as illustrated in Figure 11.33. In this specific example, the first terminal is located close to the base station and experiences a small propagation delay, $T_{P,1}$. Thus, for this terminal, a small value of the timing advance offset $T_{A,1}$ is sufficient to compensate for the propagation delay and to ensure the correct timing at the base station. However, a larger value of the timing advance is required for the second terminal, which is located at a larger distance from the base station and thus experiences a larger propagation delay.

The timing-advance value for each terminal is determined by the network based on measurements on the respective uplink transmissions. Hence, as long as a terminal carries out uplink data transmission, this can be used by the receiving base station to estimate the uplink receive timing and thus be a source for the timing-advance commands. Sounding reference signals can be used as a regular signal to measure upon, but in principle the base station can use any signal transmitted from the terminals.

Based on the uplink measurements, the network determines the required timing correction for each terminal. If the timing of a specific terminal needs correction, the network issues a timing-advance command for this specific terminal, instructing it to retard or advance its timing relative

**FIGURE 11.33**

Uplink timing advance.

to the current uplink timing. The user-specific timing-advance command is transmitted as a MAC control element (see Chapter 8 for a description of MAC control elements) on the DL-SCH. The maximum value possible for timing advance is 0.67 ms, corresponding to a terminal-to-base-station distance of slightly more than 100 km. This is also the value assumed when determining the processing time for decoding, as discussed in Section 12.1. Typically, timing-advance commands to a terminal are transmitted relatively infrequently – for example, one or a few times per second.

If the terminal has not received a timing-advance command during a (configurable) period, the terminal assumes it has lost the uplink synchronization. In this case, the terminal must re-establish uplink timing using the random-access procedure prior to any PUSCH or PUCCH transmission in the uplink.

For carrier aggregation, there may be multiple component carriers transmitted from a single terminal. In principle, different timing advance commands for different component carriers could be envisioned. One motivation for this could be inter-band carrier aggregation, where the different component carriers are received at different geographical locations, for example by using remote radio heads for some of the bands but not others. However, such deployments are not common and in the interest of simplicity LTE is using a single timing-advance command valid for all uplink component carriers.

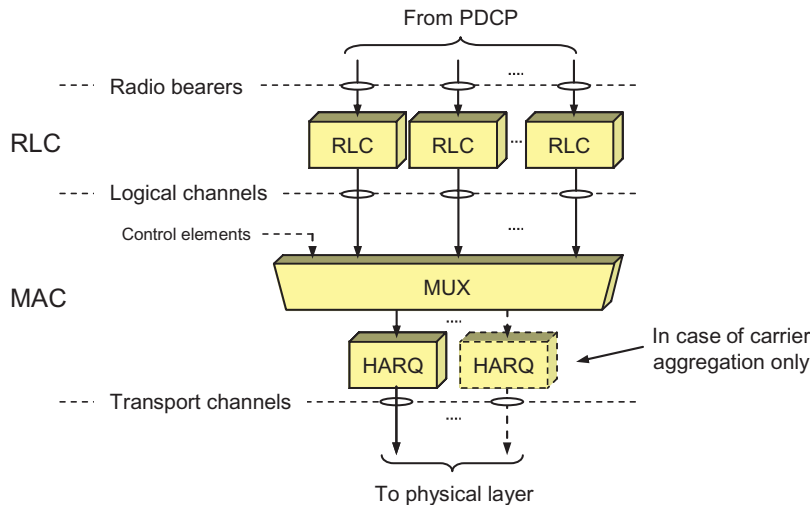
Retransmission Protocols

12

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

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

Compared to the hybrid-ARQ acknowledgements, the RLC status reports are transmitted relatively infrequently and thus the cost of obtaining a reliability of 10^{-5} or lower is relatively small. Hence, the combination of hybrid-ARQ and RLC attains a good combination of small round-trip time and a modest feedback overhead where the two components complement each other – fast retransmissions due to the hybrid-ARQ mechanism and reliable packet delivery due to the RLC. As the MAC and RLC protocol layers are located in the same network node, a tight interaction between the two protocols is possible. Hence, to some extent, the combination of the two can be viewed as *one* retransmission mechanism with *two* feedback channels. However, note that, as discussed in Chapter 8 and illustrated in Figure 12.1, the RLC operates per logical channel, while the hybrid-ARQ operates per transport channel (that is, per component carrier). One hybrid-ARQ entity may therefore retransmit data belonging to multiple logical channels. Similarly, one RLC entity may handle data transmitted across multiple component carriers in the presence of carrier aggregation. Finally, for carrier aggregation, hybrid-ARQ retransmissions must occur on the same component carrier as the original transmission, as each component carrier has its own independent hybrid-ARQ entity. RLC retransmissions, on the other hand, are not tied to a specific component carrier as, in essence, carrier aggregation is invisible above the MAC layer.

**FIGURE 12.1**

RLC and hybrid-ARQ retransmission mechanisms in LTE.

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

12.1 HYBRID ARQ WITH SOFT COMBINING

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

The basis for the LTE hybrid-ARQ mechanism is a structure with multiple stop-and-wait protocols, each operating on a single transport block. In a stop-and-wait protocol, the transmitter stops and waits for an acknowledgement after each transmitted transport block. This is a simple scheme; the only feedback required is a single bit indicating positive or negative acknowledgement of the transport block. However, since the transmitter stops after each transmission, the throughput is also low. LTE therefore applies *multiple* stop-and-wait processes operating in parallel such that, while waiting for acknowledgement from one process, the transmitter can transmit data to another hybrid-ARQ process. This is illustrated in Figure 12.2; while processing the data received in the first hybrid-ARQ process the receiver can continue to receive using the second process and so on. This structure, multiple hybrid-ARQ processes operating in parallel to form one hybrid-ARQ entity, combines the simplicity of a stop-and-wait protocol while still allowing continuous transmission of data.