CHAPTER

DOWNLINK PHYSICAL-LAYER PROCESSING

6

In Chapter 4, the LTE radio-interface architecture was discussed with an overview of the functions and characteristics of the different protocol layers. Chapter 5 then gave an overview of the time-frequency structure of LTE transmissions including the structure of the basic OFDM time-frequency grid being the fundamental physical resource on both uplink and downlink. It also discussed the concept of antenna ports, especially relevant for the understanding of multi-antenna and multi-point transmissions within LTE.

This chapter provides a more detailed description of the downlink physical-layer functionality including the transport-channel processing (Section 6.1), reference signals (Section 6.2), multi-antenna transmission (Section 6.3), and L1/L2 control signaling (Section 6.4). Chapter 7 provides a corresponding description for the *uplink* transmission direction. The later chapters go further into the details of some specific uplink and downlink functions and procedures.

6.1 TRANSPORT-CHANNEL PROCESSING

As described in Chapter 4, the physical layer provides services to the MAC layer in the form of transport channels. 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 detailed description of the physical-layer processing applied to the DL-SCH, including the mapping to the physical resource—that is, to the resource elements of the OFDM time−frequency grid of the set of antenna ports to be used for the transmission. DL-SCH is the transport-channel type in LTE used for transmission of downlink user-specific higher-layer information, both user data and dedicated control information, as well as the main part of the downlink system information (see Chapter 11). 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, as mentioned in Chapter 4, the physical-layer processing, and the structure in general, for BCH transmission is quite

4G, LTE-Advanced Pro and The Road to 5G. http://dx.doi.org/10.1016/B978-0-12-804575-6.00006-6

different. BCH transmission is described in Chapter 11 as part of the discussion on LTE system information.

6.1.1 PROCESSING STEPS

The different steps of the DL-SCH physical-layer processing are outlined in Figure 6.1. In the case of carrier aggregation—that is transmission on multiple component carriers in parallel to the same device—the transmissions on the different carriers correspond to separate transport channels with separate and essentially independent physical-layer processing. The transport-channel processing outlined in Figure 6.1 and the discussion as follows is thus valid also in the case of carrier aggregation.

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

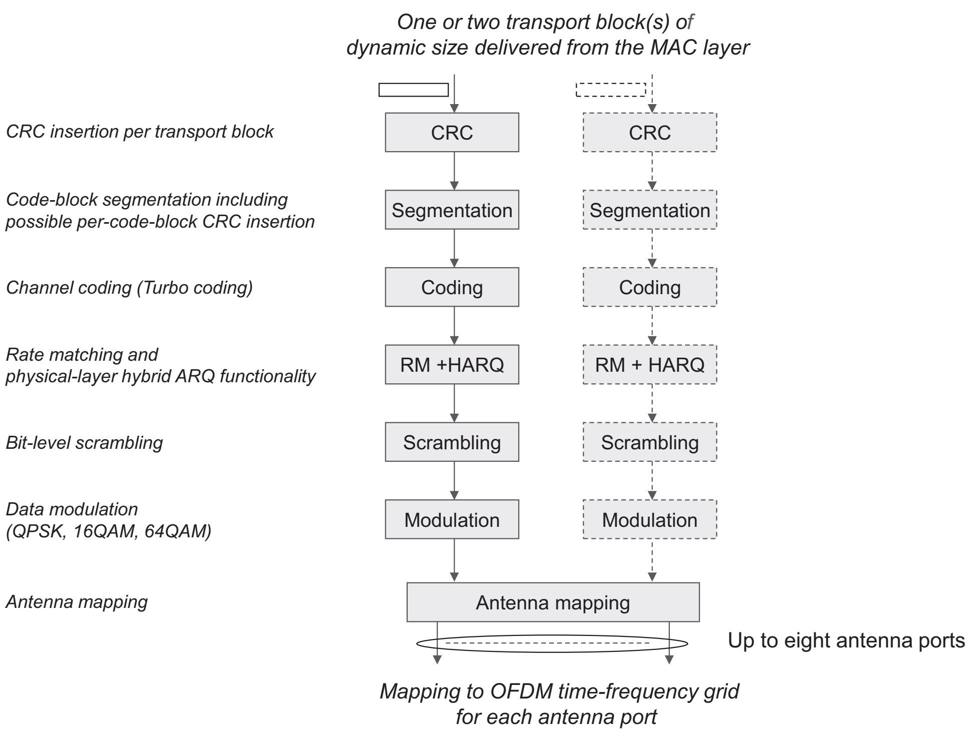

FIGURE 6.1

Physical-layer processing for the downlink shared channel (DL-SCH).

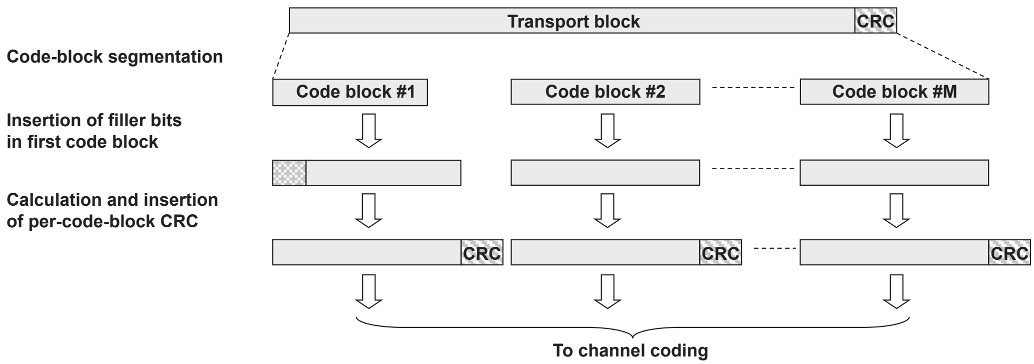

FIGURE 6.2

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

blocks transmitted within a TTI depends on the configuration of the multi-antenna transmission scheme (see Section 6.3):

- 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 device, there are two transport blocks within a TTI.[1]

6.1.1.1 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.

6.1.1.2 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* is applied before the Turbo coding as illustrated in Figure 6.2. 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 LTE 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.

[1]This 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 6.3.

As can be seen in Figure 6.2, code-block segmentation also implies that an additional CRC (also of length 24 bits but different compared to the transport-block CRC described earlier) 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 device 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 device as part of the scheduling assignment transmitted on the physical downlink control channel/enhanced physical downlink control channel (PDCCH/EPDCCH), as described in Section 6.4. Based on this information, the device can determine the code-block size and number of code blocks. The device receiver can thus, based on the information provided in the scheduling assignment, straightforwardly undo the code-block segmentation and recover the decoded transport blocks.

6.1.1.3 Channel Coding

Channel coding for DL-SCH (as well as for PCH and MCH) is based on Turbo coding [17], with encoding according to Figure 6.3. The encoding consists of two rate-1/2, eight-state

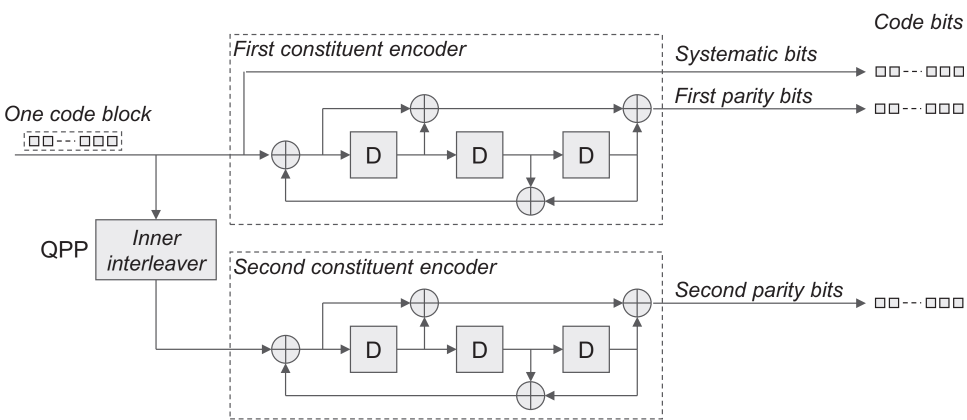

FIGURE 6.3

LTE Turbo encoder.

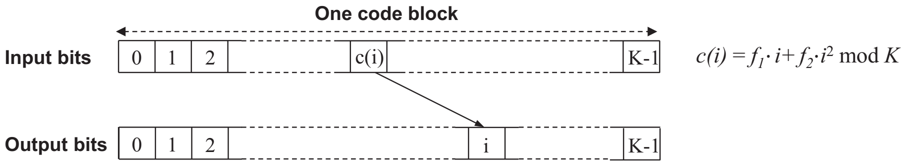

FIGURE 6.4

Principles of QPP-based interleaving.

constituent encoders, implying an overall code rate of 1/3, in combination with QPP-based[2] interleaving [32]. As illustrated in Figure 6.4, 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* [33], implying that the decoding can be parallelized without the risk for contention when the different parallel processes are accessing the 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.

6.1.1.4 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 6.5, 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 RV. The transmitter/scheduler

[2] *QPP*, Quadrature permutation polynomial.

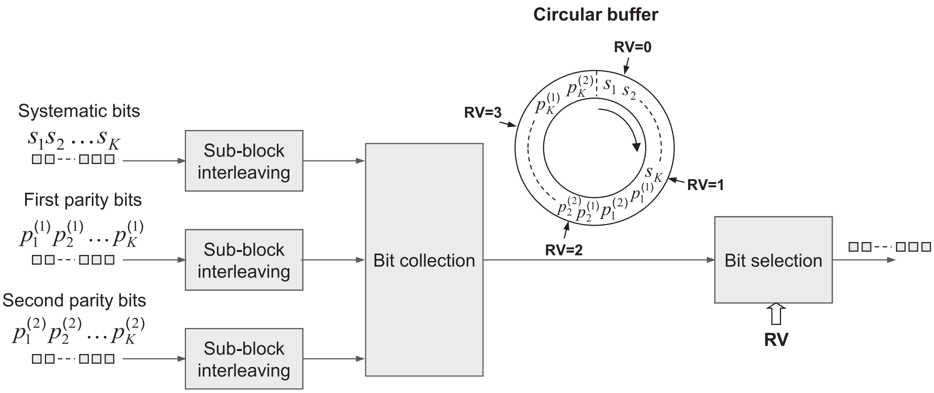

FIGURE 6.5

Rate-matching and hybrid-ARQ functionality.

selects the RV and provides information about the selection as part of the scheduling assignment (see Section 6.4.4).

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.

6.1.1.5 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 device 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 the scrambling sequences differ between neighboring cells (*cell-specific scrambling*). The scrambling also depends on the identity of the device to which the transmission is intended, assuming that the data is intended for a specific device. 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 19). Thus, in the case of MCH transmission the scrambling depends on the MBSFN area identity.

6.1.1.6 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. Optional support for 256QAM, corresponding to eight bits per symbol, was added in release 12, primarily intended for small-cell environments where the achievable SNR can be relatively high.[3]

6.1.1.7 Antenna Mapping

The antenna mapping jointly processes the modulation symbols corresponding to the one or two transport blocks and maps the result to the set of antenna ports to be used for the transmission. 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 6.1, LTE supports simultaneous 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 6.3.

6.1.1.8 Resource-Block Mapping

The resource-block mapping takes the symbols to be transmitted on each antenna port and maps them to the set of available resource elements in the set of resource blocks assigned by the MAC scheduler for the transmission. As described in Chapter 5, each resource block consists of 84 resource elements (twelve subcarriers during seven OFDM symbols).[4] 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 6.2;
- downlink L1/L2 control signaling (one, two, or three OFDM symbols at the head of each subframe) as described in Section 6.4.[5]

Furthermore, as described in Chapter 11, within some resource blocks additional resource elements are reserved for the transmission of *synchronization signals* as well as for the PBCH physical channel which carries the BCH transport channel.

It should also be pointed out that for the so-called Transmission Mode 10, the possibility for more dynamic control of the PDSCH mapping has been introduced to support multi-point transmission. This is further discussed in Chapter 13 as part of the description of CoMP-related features introduced in LTE release 11. The introduction of license-assisted access in release 13 provides some additional flexibility in the PDSCH mapping as described in Chapter 17.

---

[3]For backward-compatibility reasons, 256QAM is not supported for PMCH when carrying MCCH (see Chapter 19 for a description of the MBMS channels).

[4]72 resource elements in case of extended cyclic prefix.

[5]In MBSFN subframes the control region is limited to a maximum of two OFDM symbols.

6.1.2 LOCALIZED AND DISTRIBUTED RESOURCE MAPPING

As already discussed in Chapter 3, when deciding what set of resource blocks to use for transmission to a specific device, 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 device 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 nonfrequency-contiguous resource blocks for the transmission to a device. LTE allows for such *distributed resource-block allocation* by means of *resource allocation types 0 and 1* (see Section 6.4.6.1). 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 control-signaling overhead for the scheduling assignment, see Section 6.4.6

In contrast, *resource-allocation type 2* (Section 6.4.6.1) always allows for the allocation of a single resource-block pair and is also associated with a relatively small control-signaling overhead. 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 6.6.

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;
- even a single VRB pair is distributed in the frequency domain.

The basic principle of distributed transmission is outlined in Figure 6.7 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 6.7). 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.
- 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 6.7). 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 6.6) or distributed (mapped according to Figure 6.7) is indicated as part of the scheduling assignment in the case

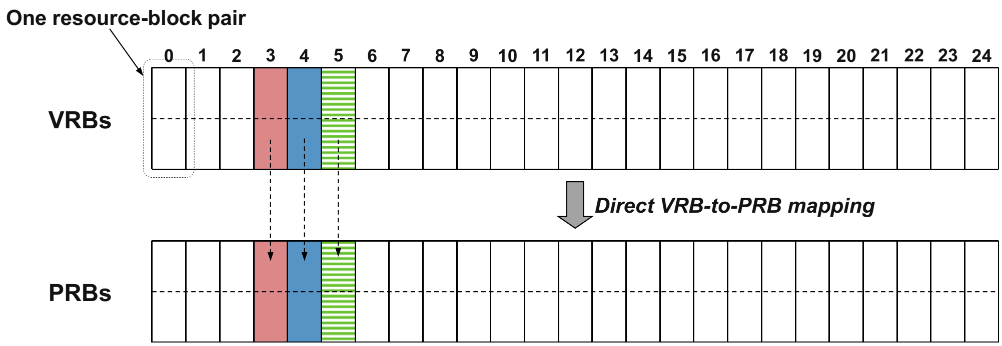

FIGURE 6.6

VRB-to-PRB mapping in case of localized VRBs. Figure assumes a cell bandwidth corresponding to 25 resource blocks.

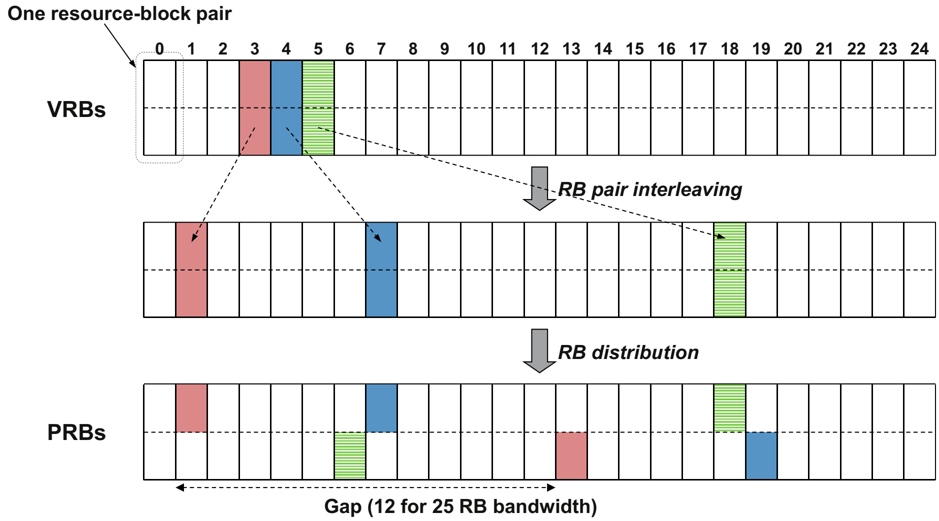

FIGURE 6.7

VRB-to-PRB mapping in case of distributed VRBs. Figure assumes a cell bandwidth corresponding to 25 resource blocks.

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 devices within the same subframe.

The exact size of the frequency gap in Figure 6.7 depends on the overall downlink cell bandwidth according to Table 6.1. 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 6.4.6 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 in the preceding paragraphs and transmissions based on downlink allocation types 0 and 1.

Table 6.1 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
P	1	1	1	2	2	2	3	3	4	4
Gap size	3	4	5	4	8	12	18	27	32	48

Table 6.2 Second Gap Size for Different Cell Bandwidth (Only Applicable for Cell Bandwidths of 50 RBs and Beyond)

Bandwidth	50–63	64–110
Gap size	9	16

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 6.7) and a corresponding gap size equal to 12 according to Table 6.1, the 25th resource-block pair cannot be used for distributed transmission. As another example, for a cell bandwidth corresponding to 50 resource blocks (gap size equal to 27 according to Table 6.1) only 46 resource blocks would be available for distributed transmission.

In addition to the gap size outlined in Table 6.1, 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 6.2). 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 6.1 and the smaller gap according to Table 6.2 is indicated by an additional bit in the resource allocation.

6.2 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 transmitted in different ways and intended to be used for different purposes by a receiving device:

- *Cell-specific reference signals* (CRS) are transmitted in every downlink subframe and in every resource block in the frequency domain. CRS are intended to be used by devices for channel estimation for coherent demodulation of all downlink physical channels except PMCH, PDSCH in case of *transmission modes* 7–10, and the EPDCCH control channel introduced in LTE release 11 (see Section 6.4).[6] CRS are also assumed to be used to acquire *channel-state information* (CSI) by devices configured in transmission modes 1–8. Finally, device measurements on CRS are assumed to be 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 intended to be used by devices for channel estimation for coherent

[6]See Section 6.3.1 for more details on LTE *transmission modes*.

demodulation of PDSCH in case of transmission modes 7–10.[7] DM-RS are also to be used for demodulation of the EPDCCH physical channel. The alternative label "*UE-specific reference signals*" relates to the fact that a specific demodulation reference signal is typically intended to be used for channel estimation by a specific device (UE). The reference signal is then only transmitted within the resource blocks specifically assigned for PDSCH/EPDCCH transmission to that device.

- *CSI reference signals* (CSI-RS) are intended to be used by devices to acquire CSI. More specifically, CSI-RS are intended to be used to acquire CSI by devices configured in transmission modes 9 and 10. CSI-RS have a significantly lower time–frequency density, thus implying less overhead, and a higher degree of flexibility compared to the CRS.
- *MBSFN reference signals* are intended to be used by devices for channel estimation for coherent demodulation in case of MCH transmission using *MBSFN* (see Chapter 19 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 device measurements on multiple LTE cells to estimate the geographical position of the device. 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.

6.2.1 CELL-SPECIFIC REFERENCE SIGNALS

CRS, 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 CRS in a cell, defining one, two, or four corresponding antenna ports, referred to as antenna port 0 to antenna port 3 in the LTE specifications.

6.2.1.1 Structure of a Single Reference Signal

Figure 6.8 illustrates the structure of a single CRS. As can be seen, it consists of *reference symbols* of predefined values inserted within the first and third last[8] 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 CRS 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

[7]In the LTE specifications, these reference signals are actually referred to as *UE-specific reference signals*, although they are still "abbreviated" DM-RS.

[8]This corresponds to the fifth and fourth OFDM symbols of the slot for normal and extended cyclic prefixes, respectively.

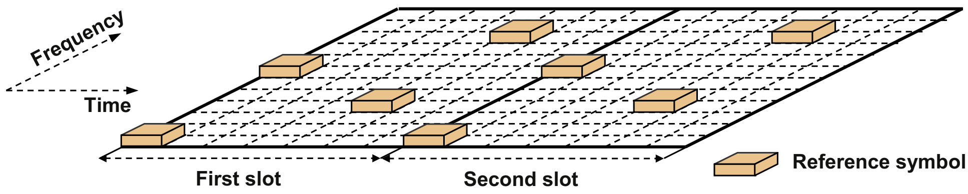

FIGURE 6.8

Structure of CRS within a pair of resource blocks.

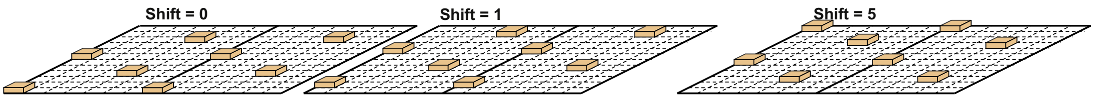

FIGURE 6.9

Different CRS frequency shifts.

maximum possible LTE carrier bandwidth corresponding to 110 resource blocks in the frequency domain. Thus, the basic reference-signal sequence has a length of 8800 symbols.[9] 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 device to estimate the channel corresponding to the center 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 described in more detail in Chapter 11, during the so-called *cell-search procedure* the device detects the physical-layer identity of the cell as well as the cell frame timing. Thus, from the cell-search procedure, the device 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 6.8 is only one of six possible *frequency shifts* for the CRS reference symbols, as illustrated in Figure 6.9. 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

[9]Four reference symbols per resource block, 110 resource blocks per slot, and 20 slots per frame.

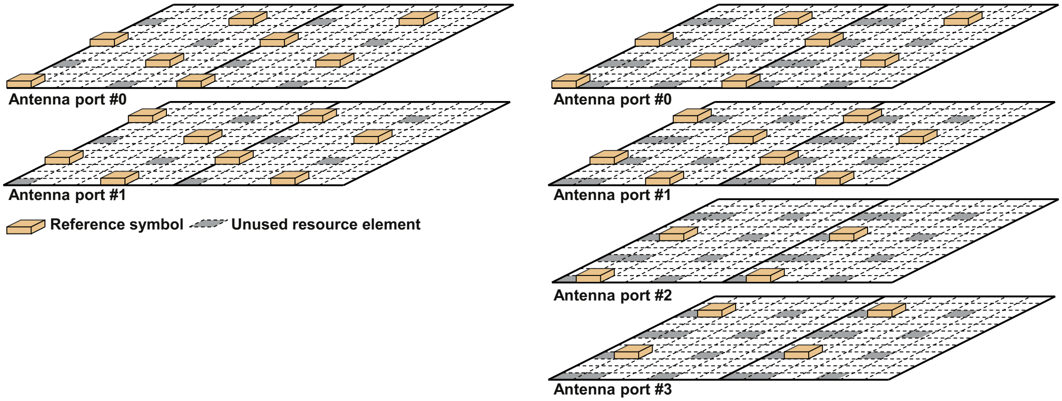

FIGURE 6.10

Structure of CRS in case of multiple reference signals: two reference signals corresponding to two antenna ports (left) and four reference signals corresponding to four antenna ports (right).

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,[10] 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 nonreference symbols of neighboring cells, implying an improved reference-signal SIR in the case of reference-signal boosting.

6.2.1.2 Multiple Reference Signals

Figure 6.10 illustrates the reference-signal structure in the case of multiple, more specifically two and four, CRS, and corresponding multiple antenna ports, within a cell:[11]

- In the case of two reference signals within a cell (left part of Figure 6.10), 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 6.10), 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

[10]This assumes that the cell transmissions are frame-timing aligned.
[11]It is not possible to configure a cell with three CRS.

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 about 5% (4 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 about 2.5%. This obviously has an impact on the possibility for the device 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 CRS 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 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 CRS within the MBSFN part of the MBSFN subframe.

6.2.2 DEMODULATION REFERENCE SIGNALS

In contrast to CRS, a demodulation reference signal (DM-RS) is intended to be used for channel estimation by a specific device and is then only transmitted within the resource blocks assigned for transmission to that device.

DM-RS was supported already in the first release of LTE (release 8). However, the use of DM-RS was then limited to the demodulation of single-layer PDSCH transmission—that is, no spatial multiplexing—corresponding to transmission mode 7. In LTE release 9, transmission based on DM-RS was extended to support dual-layer PDSCH transmission corresponding to transmission mode 8, requiring up to two simultaneous reference signals (one for each layer). Transmission based on DM-RS was then further extended in LTE release 10 to support up to eight-layer PDSCH transmission (transmission mode 9 and, from release 11, also transmission mode 10), corresponding to up to eight reference signals.[12]

Actually, the dual-layer-supporting DM-RS structure introduced in LTE release 9 was not a straightforward extension of the release 8 single-layer-limited DM-RS structure but rather a new structure, supporting both single-layer and dual-layer transmission. Already at the time of finalizing LTE release 9 it was relatively clear that the LTE radio-access technology should

[12]Transmission mode 10 is a release 11 extension of transmission mode 9 introducing improved support for multi-point coordination/transmission (CoMP), see also Section 6.3.1.

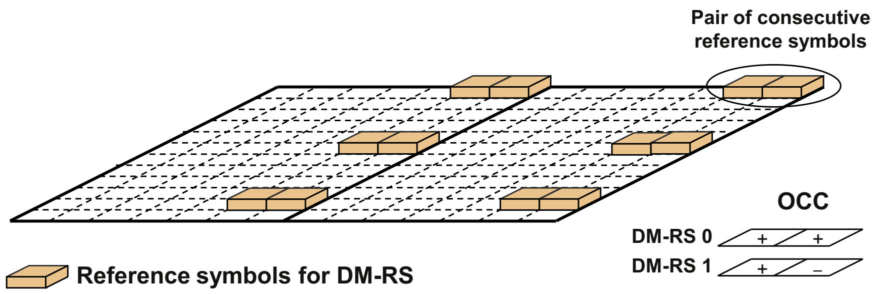

FIGURE 6.11

Structure of DM-RS for the case of one or two reference signals including size-two OCC to separate the two reference signals.

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 instead decided to introduce a new, more future-proof structure already in release 9. Here we focus on the DM-RS structure introduced in LTE release 9 including the release 10 extension to support up to eight simultaneous reference signals.

The structure of the DM-RS for EPDCCH is very similar to that of DM-RS for PDSCH although with some limitations such as support for a maximum of four reference signals.

6.2.2.1 DM-RS for PDSCH

Figure 6.11 illustrates the DM-RS time–frequency structure for the case of one or two reference signals.[13] As can be seen, there are 12 reference symbols within a resource-block pair. In contrast to CRS, for which the reference symbols of one reference signal correspond to unused resource elements for other reference signals (see Figure 6.10), in the case of two DM-RS all 12 reference symbols in Figure 6.11 are transmitted for both reference signals. Interference between the reference signals is instead avoided by applying mutually orthogonal patterns, referred to as *orthogonal cover codes* (OCC), to pairs of consecutive reference symbols as illustrated in the lower right part of the figure.

Figure 6.12 illustrates the extended DM-RS structure introduced in LTE release 10 to support up to eight reference signals. In this case, there are up to 24 reference-symbol positions within a resource-block pair. The reference signals are frequency multiplexed in groups of up to four reference signals while, within each group, the up to four reference signals are separated by means of OCC spanning four reference symbols in the time domain (two pairs of consecutive reference symbols). It should be noted that orthogonality between

[13] In 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.

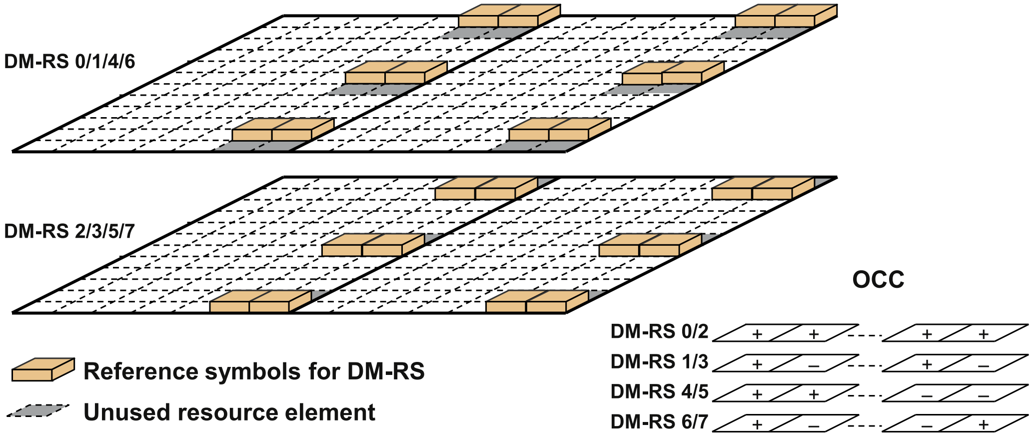

FIGURE 6.12

Demodulation reference signal structure for the case of more than two reference signals including size-four OCC to separate up to four reference signals.

the full set of eight reference signals requires that the channel does not vary over the four reference symbols spanned by the cover codes. As the four reference symbols that the cover codes span are not consecutive in time, this implies somewhat stronger constraints in terms of the amount of channel variations that can be tolerated without seriously impacting the reference-signal orthogonality. However, more than four DM-RS are only transmitted in case of spatial multiplexing with more than four layers, which is typically only applicable to low-mobility scenarios. Also note that the cover codes are defined 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 (Figure 6.11).

The up to eight different PDSCH DM-RS that can be configured for a device correspond to antenna port 7 up to antenna port 14 in the LTE specification, with antenna port 7 and antenna port 8 corresponding to the release 9 DM-RS supporting up to two-layers spatial multiplexing.[14]

The DM-RS *reference-signal sequence*—that is, the values to be taken by the DM-RS reference symbols—repeats itself every frame. Up to and including LTE release 10, the reference-signal sequence was independent of what device the DM-RS transmission was intended for but depended on the physical-layer cell identity. The reference-signal sequence thus differed between cells. Furthermore, there was the possibility to dynamically—that is, on a subframe basis—select between two different reference-signal sequences. Information about the selected sequence was then signaled to the device by means of a one-bit indicator in

[14]The single DM-RS supported already in release 8 corresponds to antenna port 5.

the scheduling assignment (see also Section 6.4.6). The reason for having the possibility to dynamically select between two reference-signal sequences was to be able to use the same resource block for PDSCH transmission to two different devices and rely on spatial separation, for example, by means of beam-forming, to separate the two transmissions. Such spatial separation, in 3GPP referred to as *multi-user multi-input–multi-output* (MU-MIMO), is typically not perfect in the sense that there will remain some interference between the transmissions. By applying different reference-signal sequences for the two spatially separated transmissions, interference randomization is achieved for the channel estimation. Downlink MU-MIMO is discussed in more detail in Section 6.3 together with the different means to separate the DM-RS of different transmissions.

However, in LTE release 11, the possibility for *device-specific* reference-signal sequences was introduced. This was done by introducing the possibility for the network to explicitly configure a device with a parameter that, if configured, should replace that cell identity when deriving the reference-signal sequence to be used by the device. If no device-specific parameter is configured, the device should assume cell-specific reference-signal sequences in line with releases 9/10 as previously discussed.[15]

The reason for introducing the possibility for device-specific reference-signal sequences was to be able to spatially separate significantly more devices within a cell. Especially in the case of so-called *shared-cell* heterogeneous deployments there may be situations with a large number of low-power transmission points, all being part of the same cell. In such a case one typically wants to be able to reuse the same physical resources—that is, the same resource blocks, for simultaneous PDSCH transmission to different devices from several of the transmission points. For robust channel estimation, the reference signals of each of these transmissions should preferably be based on unique reference-signal sequences, thus making device-specific reference signal sequences desirable. Heterogeneous deployments for LTE is extensively discussed in Chapter 14.

When DM-RS 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 (the 12 and 24 resource elements in Figures 6.11 and 6.12, respectively). Although this modified mapping is not "understood" by earlier-release devices not supporting DM-RS, this is not a problem as DM-RS will only be transmitted in resource blocks that are scheduled for PDSCH transmission to devices of later releases supporting DM-RS and thus "understanding" the modified PDSCH mapping.

As the number of transmitted layers may vary dynamically, the number of transmitted DM-RS may also vary. Thus, the transmission may dynamically change between the DM-RS structures outlined in Figures 6.11 and 6.12, respectively. The device 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.

[15]*Cell-specific reference-signal sequences for DM-RS* should not be mixed up with CRS.

6.2.2.2 DM-RS for EPDCCH

As mentioned earlier, as part of LTE release 11, a new L1/L2 control channel structure was introduced based on the so-called *enhanced PDCCH* (EPDCCH). The EPDCCH is extensively described in Section 6.4.4. Here it can just be said that, in contrast to the legacy control-channel structure (PDCCH), the EPDCCH is transmitted within resource blocks in a similar way as PDSCH. Furthermore, in contrast to the PDCCH, EPDCCH demodulation is assumed to be based on DM-RS transmitted together with the EPDCCH, similar to the use of DM-RS for PDSCH.

The structure of DM-RS for EPDCCH is very similar to the PDSCH DM-RS structure described previously. Especially, the time–frequency structure of the EPDCCH DM-RS is the same as that for PDSCH. However, for EPDCCH, there can only be up to four DM-RS, compared to up to eight DM-RS in case of PDSCH transmission. Thus the four orthogonal covers corresponding to DM-RS 4–7 in Figure 6.11 are not supported for EPDCCH transmission. Furthermore, the EPDCCH reference-signal sequence is *always* device specific—that is, the device is explicitly configured with a parameter that is used to derive the reference-signal sequence. It should be noted that this configuration of the DM-RS reference-signal sequence for EPDCCH is done independently of the corresponding configuration for the PDSCH DM-RS.

The antenna ports corresponding to up to four DM-RS for EPDCCH are, in the LTE specifications, referred to as antenna port 107 to antenna port 110. It should be noted that, although up to four different DM-RS and corresponding antenna ports can be defined for EPDCCH, a specific EPDCCH is only transmitted from a single antenna port in case of localized transmission and two antenna ports in case of distributed transmission (see Section 6.4.4). Thus, in some sense it is somewhat misleading for the specification to talk about *up to four antenna ports* for EPDCCH as a device will only see one or two DM-RS-related antenna port(s).

6.2.3 CSI REFERENCE SIGNALS

CSI-RS were introduced in LTE release 10. CSI-RS are specifically intended to be used by devices to acquire CSI—for example, for channel-dependent scheduling, link adaptation, and transmission settings related to multi-antenna transmission. More specifically, CSI-RS were introduced to acquire CSI for devices configured with transmission mode 9 and 10,[16] but in later releases also serve other purposes.

As mentioned early in Section 6.2, the CRS, available since the first release of LTE, can also be used to acquire CSI. The direct reason to introduce CSI-RS was the introduction of support for up to eight-layers spatial multiplexing in LTE release 10 and the corresponding need for devices to be able to acquire CSI for, at least, up to eight antenna ports.

[16]The reason why CSI-RS are not used for transmission modes 7 and 8 despite the fact that these transmission modes assume DM-RS for channel estimation, was simply that these transmission modes were introduced in LTE releases 8 and 9, respectively, while CSI-RS was not introduced until LTE release 10.

However, there was also a more fundamental desire to separate two different functions of downlink reference signals, namely

- the function to acquire detailed channel estimates for coherent demodulation of different downlink transmissions;
- the function to acquire CSI for, for example, downlink link adaptation and scheduling.

For the early releases of LTE, both these functions relied on CRS. As a consequence, CRS has to be transmitted with high density in both time and frequency to support accurate channel estimation and coherent demodulation also for rapidly varying channels. At the same time, in order to allow for devices to acquire CSI at regular intervals, CRS has to be transmitted in every subframe regardless of whether or not there is any data transmission. For the same reason CRS is transmitted over the entire cell area and cannot be beam-formed in the direction of a specific device.

By introducing separate sets of reference signals for channel estimation and for the acquisition of CSI (DM-RS and CSI-RS, respectively) more opportunities for optimization and a higher degree of flexibility are achieved. The high-density DM-RS are only transmitted when there is data to transmit and can, for example, be subject to more or less arbitrary beam-forming. At the same time, CSI-RS provides a very efficient tool for deriving CSI for a more arbitrary number of network nodes and antenna ports. This is especially important for the support for multi-point coordination/transmission and heterogeneous deployments as is further discussed in Chapters 13 and 14, respectively.

6.2.3.1 CSI-RS Structure

The structure of the CSI-RS to be used by a device is given by a *CSI-RS configuration*. The possibility for up to eight CSI-RS in release 10 is directly related to the support for up to eight-layers spatial multiplexing and corresponding up to eight DM-RS. The number of CSI-RSs was increased to 16 in release 13 in order to better support two-dimensional beam-forming, see Chapter 10. It should be noted though that the antenna ports corresponding to CSI-RS *are not the same* as the antenna ports corresponding to DM-RS. Antenna ports corresponding to CSI-RS typically correspond to actual transmit antennas while antenna ports corresponding to DM-RS may include any antenna precoding applied at the transmitter side (see also Section 6.3.4). The antenna ports corresponding to CSI-RS are referred to as antenna port 15 up to antenna port 22 in the LTE specification, a number that was increased to 30 in release 13.

In the time domain, CSI-RS can be configured for transmission with different periodicity, ranging from a period of 5 ms (twice every frame) to 80 ms (every eighth frame). Furthermore, for a given CSI-RS periodicity, the exact subframe in which CSI-RS is transmitted can also be configured by means of a *subframe offset*.[17] In subframes in which CSI-RS is to be

[17]All up to 16 CSI-RS of a CSI-RS configuration are transmitted within the same set of subframes—that is, with the same period and subframe offset.

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

Within a resource-block pair, different resource elements can be used for CSI-RS transmission (illustrated by the 40 different resource elements colored by gray in Figure 6.13; for TDD there are even more possibilities). Exactly what set of resource elements is used for a certain CSI-RS then depends on the exact CSI-RS configuration. More specifically:

- In the case of a CSI-RS configuration consisting of one or two configured CSI-RS, a CSI-RS consists of two consecutive reference symbols, as illustrated in the upper part of Figure 6.13. In the case of two CSI-RS, the CS-RS are then separated by applying size-two OCC to the two reference symbols, similar to DM-RS. Thus, for the case of one or two CSI-RS, there is a possibility for 20 different CSI-RS configurations in a resource-block pair, two of which are illustrated in Figure 6.13.
- In the case of a CSI-RS configuration consisting of four/eight configured CSI-RS, the CSI-RS are pair-wise frequency multiplexed, as illustrated in the middle/lower part of Figure 6.13. For four/eight CSI-RS there is thus the possibility for ten/five different CSI-RS configurations.

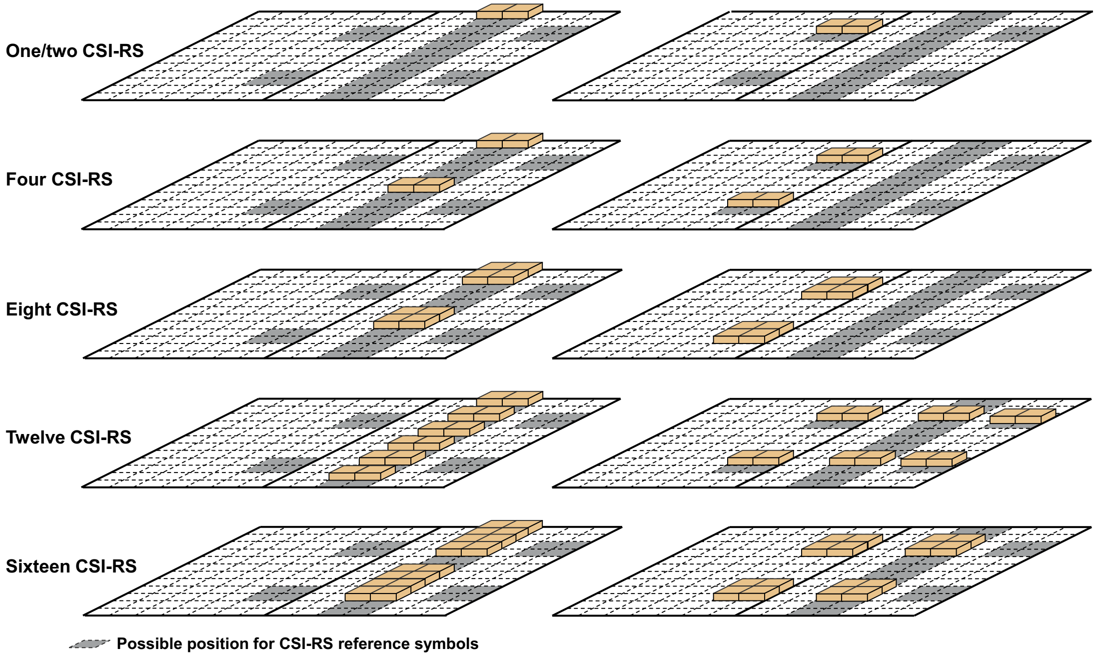

FIGURE 6.13

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

- In case of a CSI-RS configuration consisting of twelve or sixteen configured CSI-RS, a possibility introduced in release 13, aggregation of size-4 or size-8 CSI-RS are used. A configuration consisting of twelve CSI-RS is created by aggregating three size-4 CSI-RS configurations and a configuration consisting of 16 CSI-RS by aggregating two size-8 CSI-RS configurations. In other words, the resource mapping is similar to the four/eight antenna port case but more than one such configuration is used. In addition to size-two orthogonal cover code as done for eight and less configured CSI-RS, there is possibility to configure size-four OCC. The reason for the longer cover codes is to improve the possibilities to "borrow" power between CSI-RS, see Chapter 10.

To summarize, a CSI-RS configuration specifies

- the number of CSI-RS (1, 2, 4, 8, 12, or 16);
- the CSI-RS periodicity (5, 10, 20, 40, or 80 ms);
- the CSI-RS subframe offset within the CSI-RS period;
- the exact CSI-RS configuration within a resource block pair—that is, exactly what resource elements from the 40 possible resource elements (gray resource elements in Figure 6.13) are used for the CSI-RS in a resource block pair; and
- the size of the orthogonal cover code, two or four, in case of more than eight CSI-RS.

CSI-RS configurations are device-specific—meaning that each device is separately provided with a specific CSI-RS configuration that defines the number of CSI-RS to be used by the device and their detailed structure as described previously. Note though that this does not mean that a certain transmitted CSI-RS is only used by one single device. Even if each device is separately provided with its CSI-RS configuration, the configuration will, in practice, be identical for a group of, or even all, devices within a cell implying that the devices will, in practice, use the same set of CSI-RS to acquire CSI. However, the separate configuration of CSI-RS for different devices *allows* for devices within a cell to use different CSI-RS. This is important, for example, in case of shared-cell heterogeneous deployments, see Chapter 14.

6.2.3.2 CSI-RS and PDSCH Mapping

As mentioned in Section 6.2.2.1, when DM-RS 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 resource elements. This "modified" PDSCH mapping, which is not "understood" by earlier-release devices, is possible as DM-RS can be assumed to be transmitted only in resource blocks in which devices supporting such reference signals are scheduled—that is, devices based on LTE release 10 or later.[18] Expressed alternatively, an earlier-release device can be assumed never to be scheduled in a resource block in which DM-RS are transmitted and thus in which the modified PDSCH mapping is used.

[18]Partly also for devices of release 9, but then only for a maximum of two DM-RS.

The situation is different for CSI-RS. As a 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 devices would 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 releases-8/9 device. Instead, in the case of resource blocks scheduled to release 8/9 devices, 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.[19] This will 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 device 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 DM-RS. 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 device being scheduled in the resource block.

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 devices.

6.2.3.3 Zero-Power CSI-RS

As described, release 10 and beyond devices can assume that the PDSCH mapping avoids the resource elements corresponding to the set of CSI-RS configured for the device.

In addition to conventional CSI-RS, there is also the possibility to configure a device with a set of *zero-power CSI-RS* resources, where each zero-power CSI-RS has the same structure as a "conventional" (nonzero-power) CSI-RS:

- A certain periodicity (5, 10, 20, 40, or 80 ms).
- A certain subframe offset within the period.
- A certain configuration within a resource block pair.

The intention with the zero-power CSI-RS is simply to define additional CSI-RS resources to which the device should assume that PDSCH is not mapped. These resources may, for example, correspond to CSI-RS of other devices within the cell or within neighbor cells. They may also correspond to so-called CSI-IM resources as discussed in more detail in Chapter 10.

It should be noted that, despite the name, the zero-power CSI-RS resources may not necessarily be of zero power as they may, for example, correspond to "normal" (nonzero-

[19]In practice, the base station may instead not transmit PDSCH at all or, equivalently, transmit PDSCH with zero energy 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.

power) CSI-RS configured for other devices within the cell. The key point is that a device for which a certain zero-power CSI-RS resource has been configured should assume that PDSCH mapping avoids the corresponding resource elements.

6.2.4 QUASI-COLOCATION RELATIONS

Chapter 4 briefly discussed the concept of quasi-colocated antenna ports. As also mentioned, at least for the downlink an antenna port can be seen as corresponding to a specific reference signal. Thus it is important to understand what assumptions can be made regarding the relations, in terms of quasi-colocation, between downlink antenna ports corresponding to different reference signals.

Downlink antenna ports 0–3, corresponding to up to four CRS, can always be assumed to be jointly quasi-colocated. Similarly, antenna ports 7–14, corresponding to up to eight DM-RS, can also always be assumed to be jointly quasi-located. It should be pointed out though that the quasi-colocation assumption for DM-RS is only valid within a subframe. The reason for this restriction is to be able to switch a PDSCH transmission that relies on DM-RS between different transmission points on a subframe basis implying that quasi-colocation cannot be assumed *between* subframes even for a certain antenna port. Finally, antenna ports 15–30, corresponding to the up to sixteen CSI-RS of a specific CSI-RS configuration, can also always be assumed to be jointly quasi-colocated.

When it comes to quasi-colocation relations between antenna ports corresponding to different *types* of reference signals, for transmission modes 1–9 it can always be assumed that antenna ports 0–3 and 7–30—that is, CRS, DM-RS, and CSI-RS—are all *jointly* quasi-colocated. As a consequence, the only case when quasi-colocation can not necessarily be assumed for different types of reference signals is for the case of transmission mode 10. As also discussed in Section 6.3, transmission mode 10 was specifically introduced in LTE release 10 to support *multi-point* coordination/transmission. It is also in this case that the concept of quasi-colocation and lack thereof becomes relevant and, as indicated in Chapter 4, the concept of quasi-colocation was introduced in LTE release 11 for this specific reason. The specific aspects of quasi-colocation in case of transmission mode 10, and especially the quasi-colocation relation between the CSI-RS configured for a device and the set of DM-RS related to PDSCH transmission for that device, is discussed in Chapter 13 as part of a more detailed discussion on multi-point coordination and transmission.

6.3 MULTI-ANTENNA TRANSMISSION

As illustrated in Figure 6.14, multi-antenna transmission in LTE can, in general, be described as a mapping from the output of the data modulation to a set of antennas ports. The input to the antenna mapping thus consists of the modulation symbols (QPSK, 16QAM, 64QAM, and 256QAM) corresponding to the one or two transport blocks of a TTI.

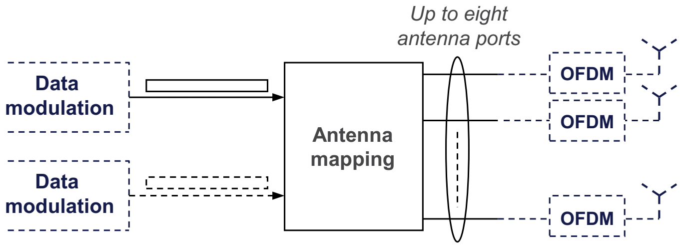

FIGURE 6.14

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

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

6.3.1 TRANSMISSION MODES

The different multi-antenna transmission schemes correspond to different so-called *transmission modes*. There are currently ten different transmission modes defined for LTE. They differ in terms of the specific structure of the antenna mapping of Figure 6.14 but also in terms of what reference signals are assumed to be used for demodulation (CRS or DM-RS, respectively) and how CSI is acquired by the device and fed back to the network. 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*. Furthermore, there are two approaches to downlink antenna precoding—*codebook-based precoding* and *non-codebook-based precoding*. The reason for these specific names is further clarified below.

Transmission mode 10 is somewhat of a special case. As mentioned earlier, transmission mode 10 was introduced in LTE release 11 to support different means of dynamic multi-point coordination and transmission (see Chapter 13). From a device point-of-view, the downlink transmission in case of transmission mode 10 is identical to that of transmission mode 9—that is, the device will see an up-to-eight-layers PDSCH transmission and rely on DM-RS for channel estimation. One important difference between transmission mode 9 and transmission mode 10 lies in the acquisition and feedback of CSI where transmission mode 10 allows for more elaborate multi-point measurements and feedback based on CSI processes as is further discussed in Chapter 13. Another important difference lies in what a device can assume in terms of quasi-colocation relations between different types of antenna ports as mentioned in Section 6.2.4 and further discussed in Chapter 13.

It should be pointed out that transmission modes are only relevant 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 when a device is configured in a certain transmission mode. 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.[20] However, this does not mean that the corresponding transmission mode is applied to such transmissions.

The following list 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 subsequent sections:

- *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*: MU-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*: Non-codebook-based precoding supporting single-layer PDSCH transmission
- *Transmission mode 8*: Non-codebook-based precoding supporting up to two layers (introduced in LTE release 9)
- *Transmission mode 9*: Non-codebook-based precoding supporting up to eight layers (extension of transmission mode 8, introduced in LTE release 10)
- *Transmission mode 10*: Extension of transmission mode 9 for enhanced support of different means of downlink multi-point coordination and transmission, also referred to as CoMP (introduced in LTE release 11)

In case of transmission modes 1–6, transmission is carried out from antenna ports 0–3. Thus the CRS are to be used for channel estimation. Transmission mode 7 corresponds to transmission on antenna port 5 while transmission modes 8–10 correspond to transmission on antenna ports 7–14 (in case of transmission mode 8 limited to antenna ports 7 and 8). Thus, for transmission modes 7–10, DM-RS are to be used for channel estimation.

In practice, devices configured for Transmission modes 1–8 can be assumed to rely on CRS to acquire CSI while, for transmission modes 9 and 10, CSI-RS should be used.

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 3–10 there is a possibility for dynamic fallback to transmit diversity without implying that

[20]Actually, only single-antenna transmission and transmit diversity are specified for BCH and L1/L2 control signaling, although the EPDCCH can use non-codebook-based precoding.

the configured transmission mode is changed. One reason for this is to enable the use of smaller DCI formats when the full set of multi-antenna features associated with a certain transmission mode is not used. Another reason is to handle ambiguities about the transmission mode applied by the device during transmission mode reconfiguration as discussed in Section 6.4.5.

6.3.2 TRANSMIT DIVERSITY

Transmit diversity can be applied to any downlink physical channel. However, it is especially applicable to transmissions that cannot be adapted to time-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 "fallback mode" for DL-SCH transmission when the device is configured in transmission mode 3 and higher. More specifically, a scheduling assignment using DCI format 1A (see Section 6.4.6) implies the use of transmit diversity regardless of the configured transmission mode.

Transmit diversity assumes the use of CRS for channel estimation. Thus, a transmit-diversity signal is always transmitted on the same antenna ports as the CRS (antenna ports 0−3). Actually, if a cell is configured with two CRS, transmit diversity for two antenna ports *must be used* for BCH and PCH, as well as for the L1/L2 control signaling on PDCCH.[21] Similarly, if four CRS 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 device 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 CRS configured for a cell.[22]

6.3.2.1 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 can be seen from Figure 6.15, 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 6.15 also indicates how the antenna ports on which a transmit-diversity signal is being transmitted correspond to the CRS, more specifically CRS 0 and CRS 1 in the case of two antenna ports. Note that one should not interpret this such that the CRS is specifically

[21]Note that this is not true for the EPDCCH control channel.

[22]Actually, the situation is partly the opposite—that is, the device blindly detects the number of antenna ports used for BCH transmission and, from that, decides on the number of CRS configured within the cell.

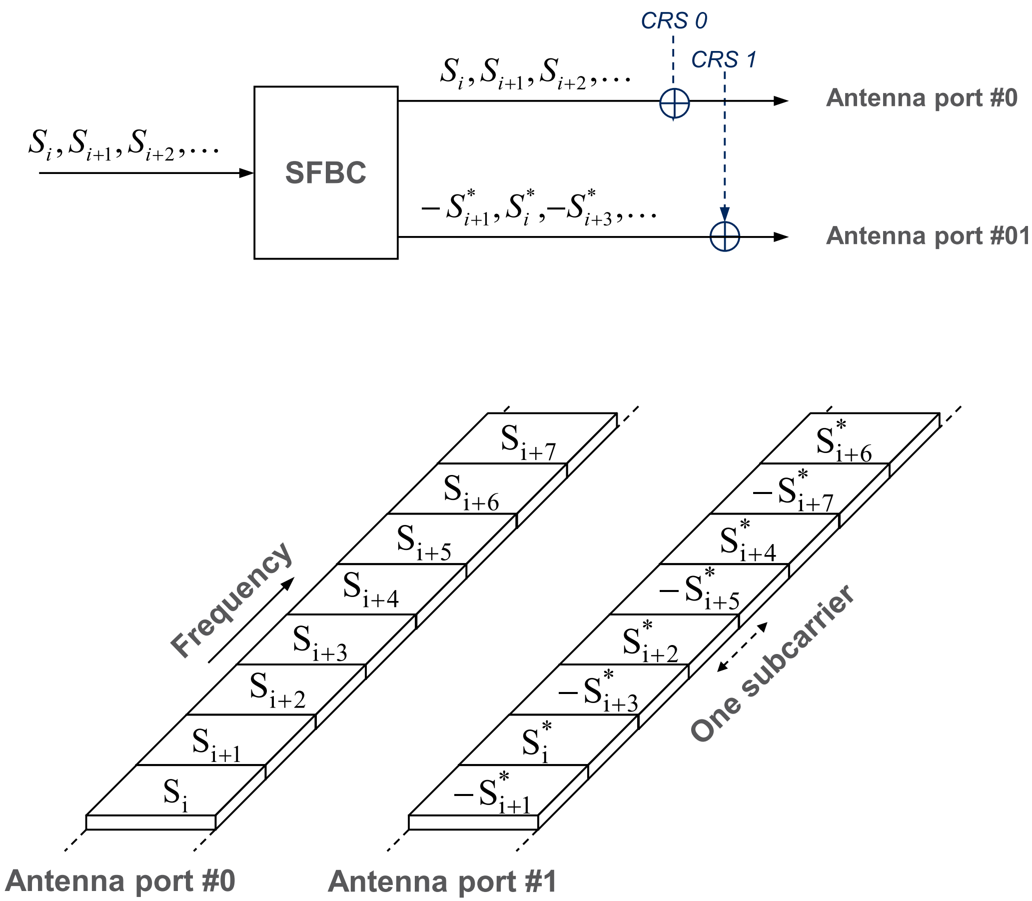

FIGURE 6.15

Transmit diversity for two antenna ports—SFBC.

transmitted for this transmit-diversity signal. There are, in practice, multiple transmissions on antenna ports 0 and 1, all of which rely on the corresponding CRS for channel estimation.

6.3.2.2 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 6.16, 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

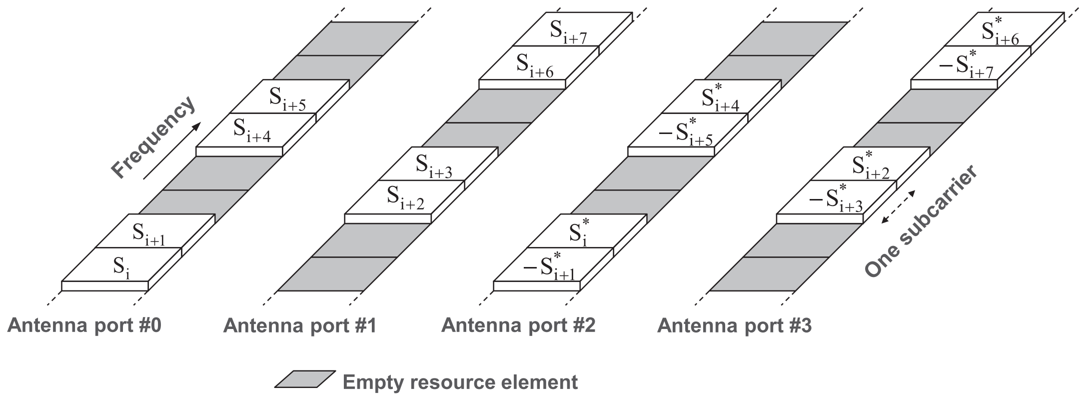

FIGURE 6.16

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

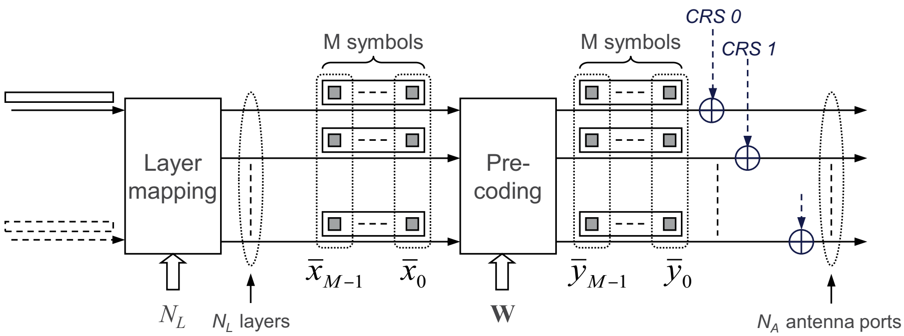

FIGURE 6.17

The basic structure of LTE codebook-based antenna precoding. The figure also indicates how CRS are applied after pre-coding.

and corresponding groups of four frequency-consecutive resource elements on each antenna port.

6.3.3 CODEBOOK-BASED PRECODING

The basic processing for codebook-based precoding is illustrated in Figure 6.17. 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.[23] The layers are then mapped to the antenna ports

[23]In practice, the number of layers is also limited by, and should not exceed, the number of receive antennas available at the device.

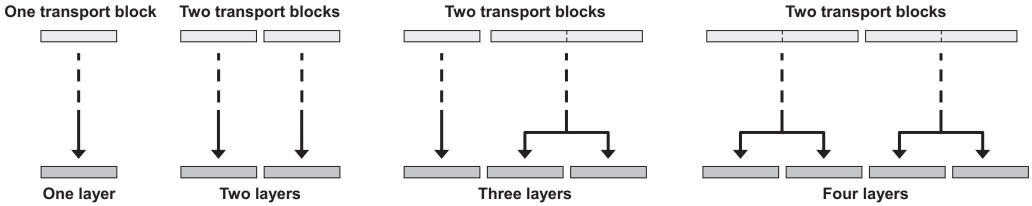

FIGURE 6.18

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

by means of the *precoder*. As codebook-based precoding relies on the CRS for channel estimation, and there are at most four CRS 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 6.18 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 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 will 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*.[24] 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 which, for

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

codebook-based precoding equal two or four. 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 6.17 also indicates how the CRS are applied after antenna precoding. Channel estimation based on the CRS will thus reflect the channel for each antenna port *not including the precoding*. As a consequence, the device receiver must have 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 CRS 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 device.

6.3.3.1 Closed-Loop Precoding

In case of closed-loop precoding it is assumed that the network selects the precoder matrix based on feedback from the device. As already mentioned, closed-loop precoding is associated with transmission mode 4.

Based on measurements on the CRS, the device 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 Indicator* (RI) and a *Precoder-Matrix Indicator* (PMI), as described in Chapter 10. 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 device when selecting the actual transmission rank and precoder matrix to be used for transmission to the device. When not following the device recommendation, the network must explicitly inform the device what precoder matrix is used for the downlink transmission. On the other hand, if the network uses the precoder matrix recommended by the device, 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 device (when reporting PMI) and the network (when selecting the actual precoder matrix to use for the subsequent downlink transmission to the device) should select a precoder matrix from the corresponding codebook. Thus, for device PMI reporting, as well as when the network informs the device about the actual precoder matrix used for the downlink transmission, only the index of the selected matrix needs to be signaled.

As LTE supports codebook-based precoding for 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 6.3. 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 6.3 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 device, the network may, for different reasons, decide to use a lower rank for the transmission, so-called *rank 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 devices with low SINR for which multi-layer transmission would not apply in order to harvest the beam-forming gain.

6.3.3.2 Open-Loop Precoding

Open-loop precoding does not rely on any detailed precoder recommendation being reported by the device and does not require any explicit network signaling of the actual precoder used

Table 6.3 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	$\frac{1}{\sqrt{2}}\begin{bmatrix}+1 & 0\\0 & +1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}+1 & +1\\+1 & -1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}+1 & +1\\+j & -j\end{bmatrix}$	

for the downlink transmission. Instead, the precoder matrix is selected in a predefined and deterministic way known to the device 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. As already mentioned, open-loop precoding is associated with transmission mode 3.

The basic transmission structure for open-loop precoding is aligned with the general codebook-based precoding outlined in Figure 6.17 and only differs from closed-loop precoding in the structure of the precoding matrix **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:

$$\mathbf{W} = \mathbf{W}' \cdot \mathbf{P} \tag{6.1}$$

In the case of two antenna ports, the matrix $\mathbf{W}'$ is the normalized 2×2 identity matrix:[25]

$$\mathbf{W}' = \frac{1}{\sqrt{2}} \begin{bmatrix} +1 & 0 \\ 0 & +1 \end{bmatrix} \tag{6.2}$$

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 $\mathbf{P} = \mathbf{D}_i \neq \mathbf{U}$, 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:

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & e^{-j2\pi/2} \end{bmatrix} \quad \mathbf{D}_i = \begin{bmatrix} 1 & 0 \\ 0 & e^{-j2\pi i/2} \end{bmatrix} \tag{6.3}$$

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 6.3.2 is used—that is, SFBC for two antenna ports and combined SFBC/FSTD for four antenna ports.

6.3.4 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, later extended to transmission mode 10).

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

There is also a release 8 non-codebook-based precoding defined, associated with transmission mode 7. Transmission mode 7 relies on the release 8 DM-RS mentioned but not described in detail in Section 6.2.2 and only supports single-layer transmission. In this description we will focus on non-codebook-based precoding corresponding to transmission modes 8–10.

The basic principles for non-codebook-based precoding can be explained based on Figure 6.19 (where the precoder is intentionally shaded; see later). As can be seen, this figure is very similar to the corresponding figure illustrating codebook-based precoding (Figure 6.17), 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 6.18) 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 6.19 compared to Figure 6.17 (codebook-based precoding) is the presence of 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 DM-RS 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 no need to signal any precoder-matrix information to the device, 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 device feedback, as described below.

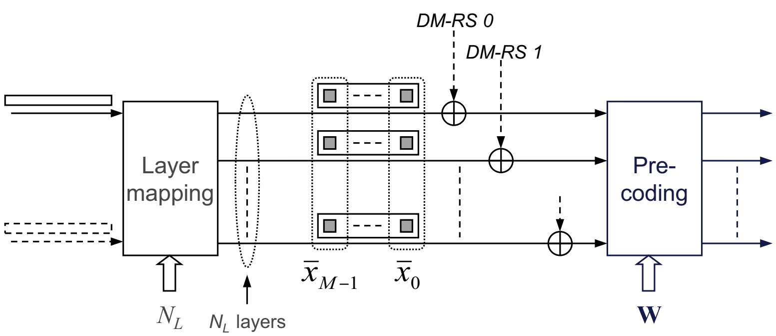

FIGURE 6.19

Basic principles for non-codebook-based antenna precoding.

The possibility to select an arbitrary precoder matrix for the transmission is also the reason why the precoder is shaded in Figure 6.19. The precoder part of Figure 6.19 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 6.14 consists of only the layer mapping. This also means that the antenna ports defined in Figure 6.14 correspond to the different layers in Figure 6.19, 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 device 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 device feedback for precoder-matrix selection. For transmission modes 8−10, this feedback is actually very similar to the corresponding feedback for closed-loop codebook-based precoding, see Chapter 10. Furthermore, for transmission mode 9, the device measurements should be based on CSI-RS, as described in Section 6.2.3, rather than the CRS.

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 device CSI reporting and not for the actual downlink transmission.

6.3.5 DOWNLINK MU-MIMO

Spatial multiplexing implies the transmission of multiple layers—that is, multiple parallel transmissions on the same time−frequency resource—to the same device. 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 has often been referred to as 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 be more clear below.[26]

[26]The term MIMO is then rather used to refer to any transmission using multiple transmit antennas and multiple receive antennas—that is, not only limited to spatial multiplexing but also, for example, single-layer precoding to get beamforming gains.

The term MU-MIMO is, in 3GPP, used to denote transmission to *different* devices 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 devices. The set of devices would demodulate and decode the full set of layers in the same way as for SU-MIMO. The data on a layer not intended for a specific device would then just be discarded by that device after demodulation/decoding.

However, such an approach would imply that all devices 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—that is, transmission on the same set of resource blocks, for all devices involved in the MU-MIMO transmission. All devices 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 device 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 device.[27] There are two approaches to MU-MIMO with explicit support in the LTE specifications, one being an integrated part of transmission modes 8, 9, and 10 corresponding to non-codebook-based precoding, and one being based on codebook-based precoding but associated with a special transmission mode, *transmission mode 5*.

6.3.5.1 MU-MIMO within Transmission Modes 8/9

In principle, MU-MIMO based on transmission modes 8–10 is straightforward. Based on feedback of CSI from devices within the cell, the base station selects two or more devices to transmit to, using the same set of time-frequency resources. Non-codebook-based precoding for one, or in some cases even multiple layers, is then applied to each transmission in such way that they are spatially separated at the receiver (device) sides.

The spatial separation at the device side will typically not be perfect. To enhance channel estimation at the device side it is therefore preferred to use different DM-RS for the different transmissions in order to improve channel estimation.[28]

As discussed in Section 6.2.2.1 there are two methods by which reference-signal sequences can be assigned to downlink DM-RS:

- Cell-specific assignment supported from LTE release 9, that is, the release in which transmission mode 8 and DM-RS were introduced.
- Fully device-specific assignment supported from LTE release 11.

[27]Note, though, that the LTE performance requirements in general assume the presence of at least two receive antennas at the mobile device.

[28]As discussed in Section 6.3.1, transmission mode 8 and 9 are assumed to rely on DM-RS for channel estimation.

As also described in Section 6.2.2.1, in case of cell-specific assignment there is the possibility to dynamically select between two different cell-specific reference-signal sequences.

With pre-release 11 (cell-specific) DM-RS sequences there is the possibility for up to four different DM-RS supporting MU-MIMO between up to four different devices:

- In case of single-layer transmission, the network can explicitly signal on which of antenna ports 7 and 8, corresponding to the two OCC, a transmission is carried out. This, in combination with the possibility to dynamically select between two reference-signal sequences, allows for up to four different DM-RS and a corresponding possibility for single-layer MU-MIMO transmission to up to four devices in parallel.
- In case of dual-layer transmission (on antenna ports 7 and 8), the possibility to dynamically select between two reference-signal sequences allows for MU-MIMO transmission to up to two devices in parallel.
- In case of more than two layers, there is no signaling support for selecting reference-signal sequence and thus no pre-release 11 support for MU-MIMO.

Note that one can also do MU-MIMO for single-layer and dual-layer transmission in parallel. There could, for example, be one two-layer transmission using one of the two reference-signal sequences and up to two one-layer transmissions using the other reference-signal sequence, separated by means of the two different OCCs.

However, with the release 11 introduction of device-specific assignment of reference-signal sequences as described in Section 6.2.2.1, MU-MIMO can, at least in principle, be applied to an arbitrary number of devices regardless of the number of layers. In a normal, for example, macro deployment, the number of devices for which MU-MIMO can jointly be carried out is limited by the number of transmitter-side antennas and, in practice, the possibility for MU-MIMO for up to four devices in parallel as supported prior to release 11 is typically sufficient. However, the possibility for MU-MIMO between significantly more devices is important in specific scenarios, especially in so-called shared-cell heterogeneous deployments, see further Chapter 14.

6.3.5.2 MU-MIMO Based on CRS

The MU-MIMO transmission described in the preceding section is part of transmission modes 8, 9, and 10 and thus became available in LTE release 9, with further extension in subsequent releases. 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 CRS.

In general, for transmission modes relying on CRS (as well as when relying on DM-RS) for channel estimation the device will use the reference signal as a phase reference and 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 device needs to know the power offset between the CRS and the PDSCH.

The device 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 device 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 devices, with less PDSCH power being available for each transmission. However, devices are not aware of the presence of parallel transmissions to other devices 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 device 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 power-offset signaling is not needed for MU-MIMO based on DM-RS 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.

6.4 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 device to be able to properly receive, demodulate, and decode the DL-SCH[29] on a component carrier, uplink scheduling grants informing the device about the resources and transport format to use for uplink (UL-SCH) transmission, and hybrid-ARQ acknowledgments 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.

The basic time−frequency structure for transmission of L1/L2 control signaling is illustrated in Figure 6.20 with control signaling being located at the beginning of each subframe

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

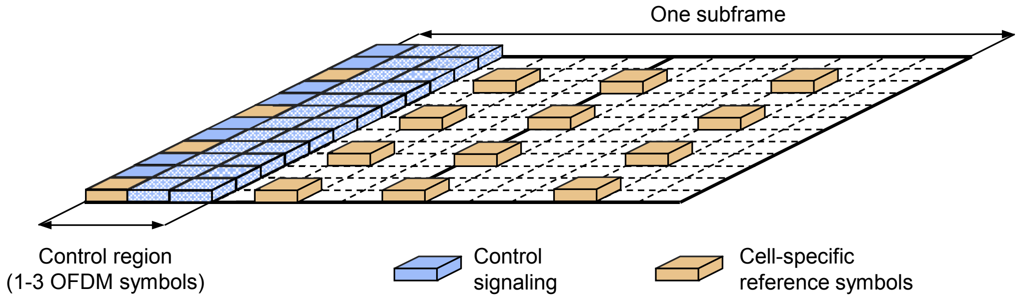

FIGURE 6.20

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

and spanning the full downlink carrier bandwidth. Each subframe can therefore 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. Starting from release 11, there is also a possibility to locate parts of the L1/L2 control signaling in the data region as described later. However, the split of a subframe into a control region and a data region still applies.

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).

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. 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 maximum size of the control region is normally three OFDM symbols (four in the case of narrow cell bandwidths), as mentioned in the preceding paragraphs. However, there are a few exceptions to this rule. When operating in TDD mode, the control region in subframes one and six is restricted to at most two OFDM symbols since, for TDD, the primary synchronization signal (see Chapter 11) occupies the third OFDM symbol in those subframes. Similarly, for MBSFN subframes (see Chapter 5), the control region is restricted to a maximum of two OFDM symbols.

The reason for transmitting the control signaling at the beginning of the subframe is to allow for devices 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—devices that are not scheduled in the subframe may power down the receiver circuitry for a part of the subframe, allowing for reduced device power consumption.

The downlink L1/L2 control signaling consists of six different physical-channel types, all located in the control region with two exceptions located in the data region of a subframe:

- The *physical control format indicator channel* (PCFICH), informing the device 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 hybrid-ARQ indicator channel* (PHICH), used to signal hybrid-ARQ acknowledgments in response to uplink UL-SCH transmissions. Multiple PHICHs can exist in each cell.
- The *physical downlink control channel* (PDCCH), used to signal downlink scheduling assignments, uplink scheduling grants, or power-control commands. Each PDCCH typically carries signaling for a single device, but can also be used to address a group of devices. Multiple PDCCHs can exist in each cell.
- The *enhanced physical downlink control channel* (EPDCCH), introduced in release 11 to support DM-RS-based control signaling reception and carrying similar types of information as the PDCCH. However, in contrast to the PDCCH, the EPDCCH is located in the data region. Also, the EPDCCH can be subject to non-codebook-based precoding.
- The *MTC physical downlink control channel* (MPDCCH), introduced in release 13 as part of the improved MTC support, see Chapter 20. In essence it is a variant of the EPDCCH.
- The *relay physical downlink control channel* (R-PDCCH), introduced in release 10 to support relaying. A detailed discussion can be found in Chapter 18 in conjunction with the overall description of relays; at this stage it suffices to note that the R-PDCCH is transmitted in the data region.

In the following sections, the PCFICH, PHICH, PDCCH, and EPDCCH are described in detail while the MPDCCH and the R-PDCCH are described in Chapters 20 and 18, respectively.

6.4.1 PHYSICAL CONTROL FORMAT INDICATOR CHANNEL

The PCFICH indicates the instantaneous 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 device will neither know how to process the control channels nor where the data region starts for the corresponding subframe.[30] The PCFICH consists of two bits of information,

[30]Theoretically, the device 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.

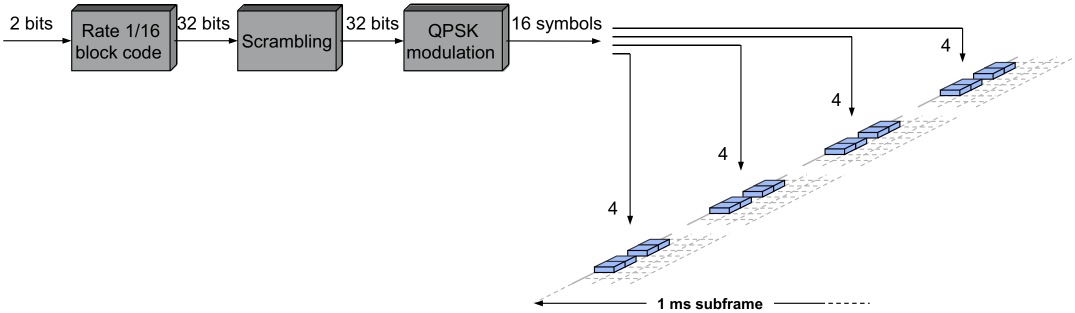

FIGURE 6.21

Overview of the PCFICH processing.

corresponding to the three[31] 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 6.21.

To describe the mapping of the PCFICH, and L1/L2 control signaling in general, to resource elements, some terminology is required. As previously mentioned, the mapping is specified in terms of groups of four resource elements, so-called *resource-element groups*. To

[31]The fourth combination is reserved for future use.

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 6.3, transmit diversity with up to four antenna ports is specified for L1/L2 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 6.22, 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 nonused resource elements corresponding to reference symbols on the other antenna port). As also illustrated in Figure 6.22, 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 6.22 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 6.23, 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 6.2.1, are also seen in the figure.

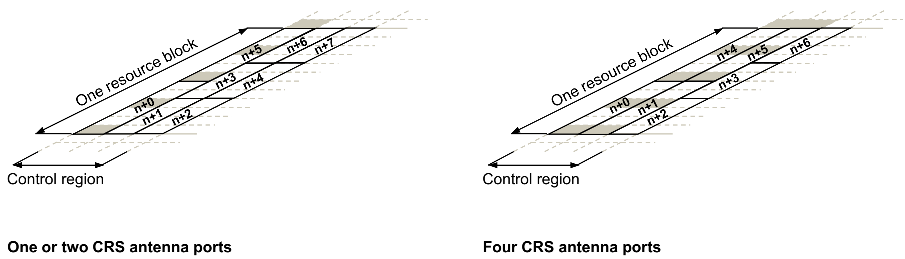

FIGURE 6.22

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

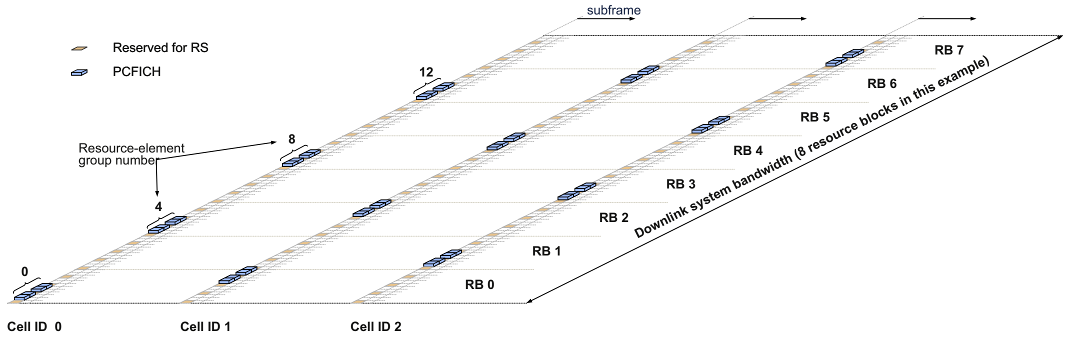

FIGURE 6.23

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

6.4.2 PHYSICAL HYBRID-ARQ INDICATOR CHANNEL

The PHICH is used for transmission of hybrid-ARQ acknowledgments in response to UL-SCH transmission. In essence the PHICH is a one-bit scheduling grant commanding a retransmission on the UL-SCH. 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 and to avoid spurious uplink transmissions, 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 device 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 acknowledgment (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 highly time-dispersive environments, the radio channel may not be flat over the frequency spanned by a length-four sequence. A nonflat 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 6.24.

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 about one-third of the downlink cell bandwidth. In the first OFDM 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 later, the PDCCHs are mapped to the remaining resource elements.

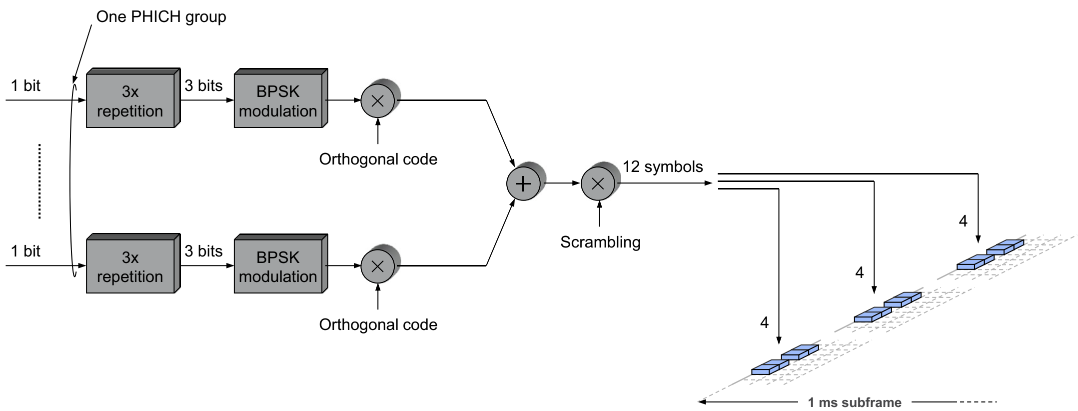

FIGURE 6.24

PHICH structure.

Typically, the PHICH is transmitted in the first OFDM symbol only, which allows the device 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. 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.

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 MU-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 device 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 device therefore has to blindly process the PDCCHs under different PHICH configuration hypotheses.

In order to minimize the overhead and not introduce any additional signaling in the uplink grants, the PHICH that the device will expect the hybrid-ARQ acknowledgment upon is derived from the number of the first resource block upon which the corresponding uplink PUSCH transmission occurred. This principle is also compatible with semi-persistently scheduled transmission (see Chapter 9) as well as retransmissions. In addition, the resources used for a particular PHICH further depend on the reference-signal phase rotation signaled as part of the uplink grant (see Section 6.4.7). In this way, multiple devices scheduled on the same set of resources using MU-MIMO will use different PHICH resources as their reference signals are assigned different phase rotations through the corresponding field in the uplink grant. For spatial multiplexing, where two PHICH resources are needed, the second PHICH uses the same principle as the first, but to ensure that different PHICHs are used for the two transport blocks, the resource for the second PHICH is derived not from the first but from the second resource block upon which the PUSCH was transmitted.[32]

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

6.4.3 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 can include:

- 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 acknowledgments 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 devices as a complement to the commands included in the scheduling assignments/grants.
- Control information related to sidelink operation as described in Chapter 21.
- Control information to support eMTC devices as described in Chapter 20.

The different types of control information, both between the groups and within the groups, correspond to different DCI message sizes. For example, supporting spatial multiplexing with noncontiguous 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 6.4, including the size for an example of 20 MHz FDD operation with 2 Tx antennas at the base station and no carrier aggregation. 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 CRS 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 later; at this stage it suffices to note that formats 0, 1A, 3, and 3A have the same message size.[33]

One PDCCH carries one DCI message with one of the formats mentioned earlier. To support different radio-channel conditions, link adaptation can be used, where the code rate (and transmission power) of the PDCCH is selected to match the radio-channel conditions. As multiple devices 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. A device may also

[33]The 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.

Table 6.4 DCI Formats

	DCI Format	Example Size (Bits)	Usage
Uplink	0	45	Uplink scheduling grant
	4	53	Uplink scheduling grant with spatial multiplexing
	6-0A, 6-0B	46, 36	Uplink scheduling grant for eMTC devices (see Chapter 20)
Downlink	1C	31	Special purpose compact assignment
	1A	45	Contiguous allocations only
	1B	46	Codebook-based beam-forming using CRS
	1D	46	MU-MIMO using CRS
	1	55	Flexible allocations
	2A	64	Open-loop spatial multiplexing using CRS
	2B	64	Dual-layer transmission using DM-RS (TM8)
	2C	66	Multi-layer transmission using DM-RS (TM9)
	2D	68	Multi-layer transmission using DM-RS (TM10)
	2	67	Closed-loop spatial multiplexing using CRS
	6-1A, 6-1B	46, 36	Downlink scheduling grants for eMTC devices (see Chapter 20)
Special	3, 3A	45	Power control commands
	5		Sidelink operation (see Chapter 21)
	6-2		Paging/direct indication for eMTC devices (see Chapter 20)

receive *multiple* DCI messages in the same subframe (on different PDCCHs), for example if it is scheduled simultaneously in uplink and downlink.

For carrier aggregation, scheduling assignments/grants are transmitted individually per component carrier as further elaborated upon in Chapter 12.

Having introduced the concept of DCI formats, the transmission of the DCI message on a PDCCH can be described. The processing of downlink control signaling is illustrated in Figure 6.25. A CRC is attached to each DCI message payload. The identity of the device (or devices) addressed—that is, the *Radio Network Temporary Identifier* (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 device-specific C-RNTI is used.

Upon reception of DCI, the device 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 device. Thus, the identity of the device 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 device point of view, there is no difference between a corrupt message whose CRC will not check and a message intended for another device.

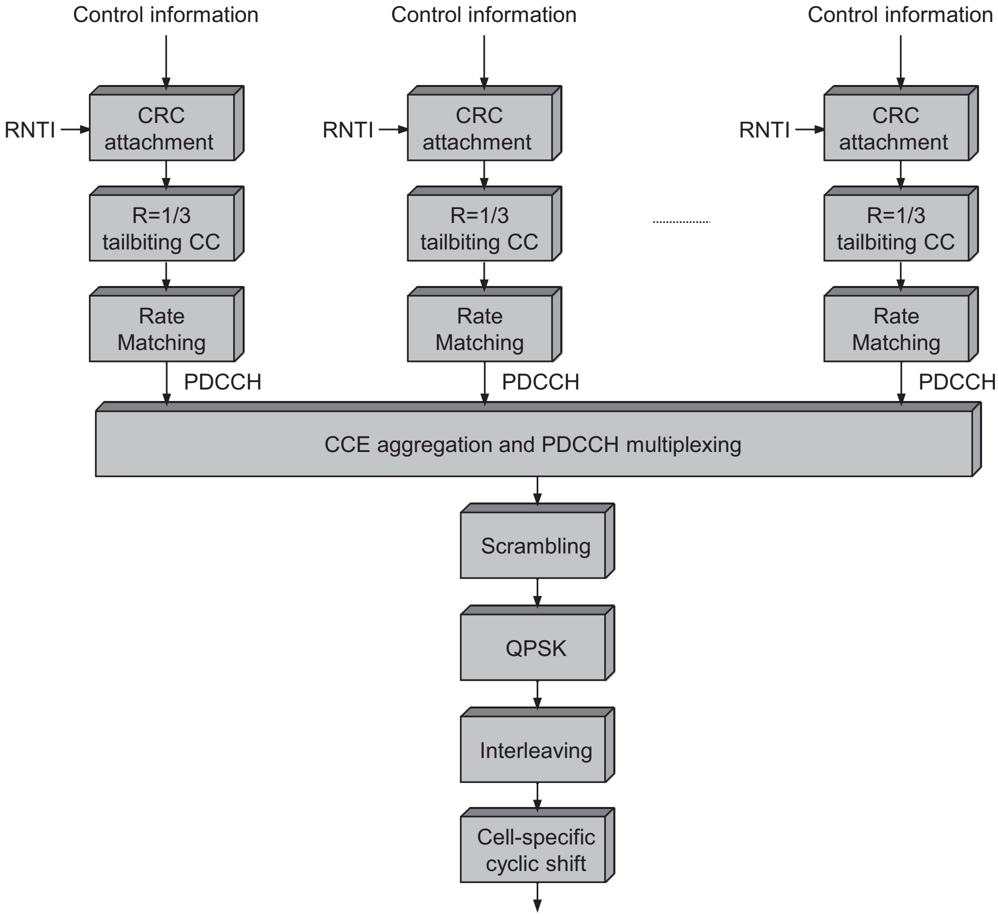

FIGURE 6.25

Processing of L1/L2 control signaling.

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 in the following paragraphs), 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 device, 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 6.4.1). 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 device to which the PDCCH is intended are 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 6.26. 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 device 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 divisible by 8, as illustrated in Figure 6.26. The same principle

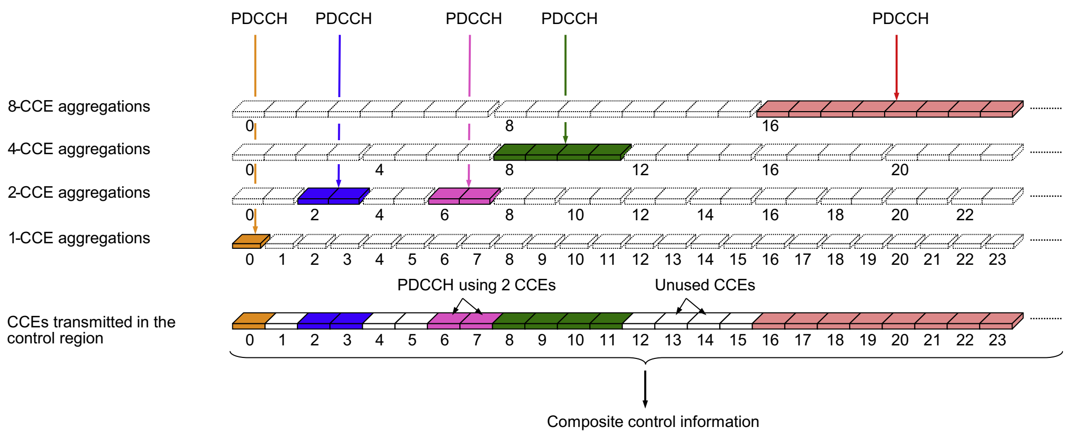

FIGURE 6.26

CCE aggregation and PDCCH multiplexing.

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 not supported.

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 device, 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 PCFICH and PHICH, 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 6.22, skipping resource-element groups used for PCFICH and PHICH. Time-first mapping 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 in the preceding paragraphs, 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 devices 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. Furthermore, lowering the code rate is generally more efficient than increasing the transmission power. 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 6.27. 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 previously, 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 PHICHs. The resource-element groups left after the PCFICH and PHICHs 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 number 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

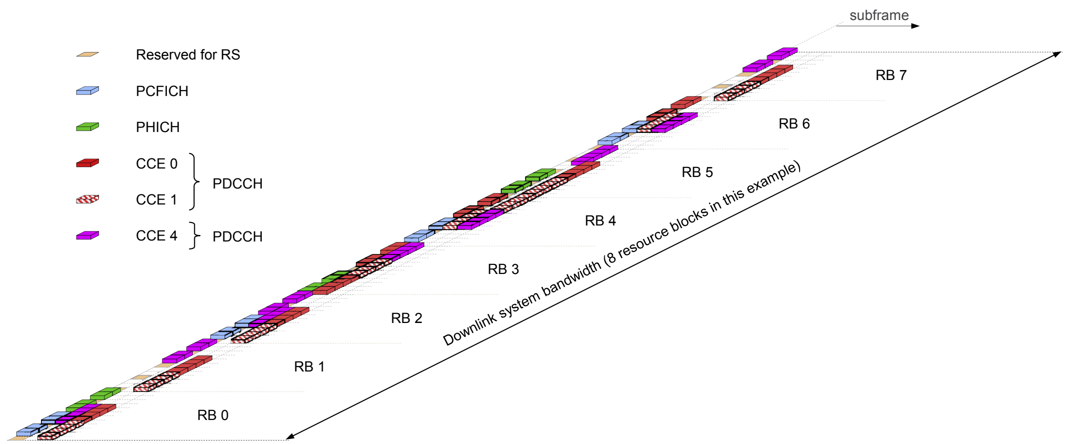

FIGURE 6.27

Example of mapping of PCFICH, PHICH, and PDCCH.

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.

6.4.4 ENHANCED PHYSICAL DOWNLINK CONTROL CHANNEL

In release 11, a complementary control channel was introduced, the *enhanced physical downlink control channel* (EPDCCH), primarily differing from the PDCCH in terms of precoding and mapping to physical resources. The reasons for introducing the EPDCCH are twofold:

- to enable frequency-domain scheduling and interference coordination also for control signaling;
- to enable DM-RS-based reception for the control signaling.

Unlike the PDCCH, which is transmitted in the control region and spans the full system bandwidth, the EPDCCH is transmitted in the data region and normally spans only a small portion of the overall bandwidth, see Figure 6.28. Hence, as it is possible to control in which part of the overall spectrum an EPDCCH is transmitted, it is possible not only to benefit from frequency-selective scheduling for control channels but also to implement various forms of inter-cell interference coordination schemes in the frequency domain. This can be very useful, for example, in heterogeneous deployments (see Chapter 14). Furthermore, as the EPDCCH uses DM-RS for demodulation in contrast to the CRS-based reception of the PDCCH, any precoding operations are therefore transparent to the device. The EPDCCH can therefore be seen as a prerequisite for CoMP (see Chapter 13) as well as more advanced antenna solutions in general for which it is beneficial for the network to have the freedom of changing the precoding without explicitly informing the device. Massive machine-type communication devices also use the EPDCCH for control signaling but with some smaller enhancements as described in Chapter 20. The rest of this section focuses on the EPDCCH processing in general and not the enhancements for machine-type communication.

In contrast to the PDCCH, the EPDCCH decoding result is not available until the end of the subframe, which leaves less processing time for the PDSCH. Hence, for the largest

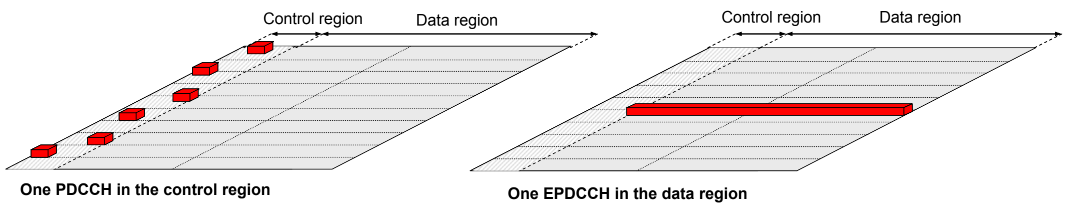

FIGURE 6.28

Illustration of principle for PDCCH mapping (left) and EPDCCH mapping (right).

DL-SCH payloads, the device may need a slightly faster PDSCH decoder compared to the PDCCH case to meet the hybrid ARQ timing.

With a few exceptions, the EPDCCH generally supports the same DCI formats as the PDCCH (see Table 6.4) and can thus be used for scheduling all the different transmission modes in both uplink and downlink. However, DCI formats 3/3A, used for uplink power control of multiple devices, and format 1C, used for scheduling system information to multiple devices, are not supported on the EPDCCH for reasons discussed in conjunction with search spaces in Section 6.4.5.

The EPDCCH processing, illustrated in Figure 6.29, is virtually identical to that of a PDCCH apart from precoding and the resource-element mapping. The identity of the device addressed—that is, the RNTI—is implicitly encoded in the CRC and not explicitly transmitted. 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 available for the EPDCCH in question. The coded and rate-matched bits are scrambled with a device and EPDCCH-set specific

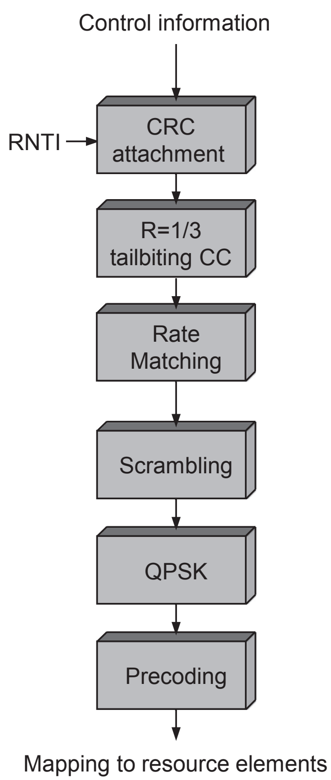

FIGURE 6.29

EPDCCH coding.

scrambling sequence (description of EPDCCH sets follows) and QPSK-modulated before being mapped to the physical resources. Link adaptation is supported by adjusting the number of control channel resources utilized by an encoded EPDCCH message, similar to the PDCCH.

To describe the mapping of EPDCCH to physical resources, it is worthwhile to view this from a device perspective. Each device using EPDCCH is configured with one or two sets of PRBs where EPDCCH transmission to that device may occur. Each set consists of two, four, or eight PRB pairs and the two sets may be of different size. There is full flexibility in the location of the resource-block pairs across the full downlink system bandwidth (the location is configured by higher layers using a combinatorial index) and the resource-block pairs do not have to be contiguous in frequency. It is important to understand that these sets are defined from a *device* perspective and only indicates where a device *may* receive EPDCCH transmissions. A resource-block pair not used for EPDCCH transmission to a particular device in a certain subframe can still be used for data transmission, either to the same device or to another device, despite being part of one of the two EPDCCH sets.

An EPDCCH set is configured as being either *localized* or *distributed*. For a localized set, a single EPDCCH is mapped to one PRB pair (or a few in case of the highest aggregation levels) to allow exploitation of frequency-domain scheduling gains. For a distributed set, on the other hand, a single EPDCCH is distributed over multiple PRB pairs with the intention of providing robustness for situations in which there is limited or no CSI available at the transmitter.

Providing two EPDCCH sets instead of just one is beneficial for several reasons. For example, one set can be configured for localized transmission in order to benefit from frequency-dependent scheduling and the other set can be configured for distributed transmission and act as a fallback in case the channel-state feedback becomes unreliable due to the device moving rapidly. It is also useful for supporting CoMP as discussed further as follows.

Inspired by the PDCCH, the EPDCCH mapping to physical resources is subject to a structure based on *enhanced control-channel elements* (ECCEs) and *enhanced resource-element groups* (EREGs). An EPDCCH is mapped to a number of ECCEs, where the number of ECCEs used for an EPDCCH is known as the *aggregation level* and is used to realize link adaptation. If the radio conditions are worse, or if the EPDCCH payload is large, a larger number of ECCEs is used than for small payloads and/or good radio conditions. An EPDCCH uses a consecutive set of ECCEs in which the number of ECCEs used for an EPDCCH can be 1, 2, 4, 8, 16, or 32 although not all these values are possible for all configurations. Note though that the ECCE numbering is in the *logical* domain; consecutive ECCEs do not necessarily imply transmission on consecutive PRB pairs.

Each ECCE in turn consists of four[34] EREGs, where an EREG in essence corresponds to nine[35] resource elements in one PRB pair. To define an EREG, number all resource elements in a PRB pair cyclically in a frequency-first manner from 0 to 15, excluding resource-

[34]Four holds for normal cyclic prefix; for extended cyclic prefix and some special subframe configurations in normal cyclic prefix there are eight EREGs per ECCE.
[35]Eight in case of extended cyclic prefix.

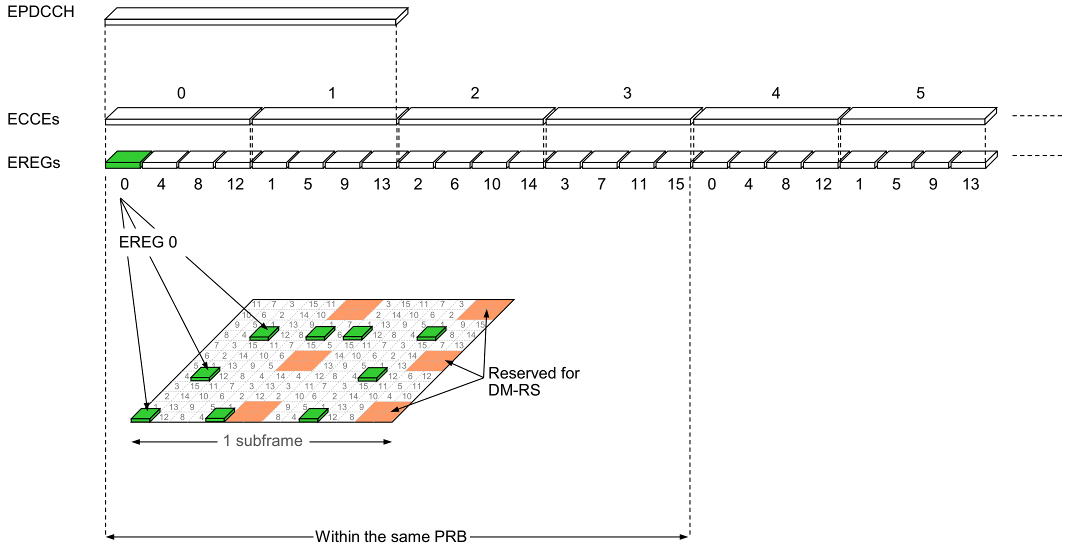

FIGURE 6.30

Example of relation between EPDCCH, ECCEs, and EREGs for localized mapping with aggregation level 2.

elements used for DM-RS. EREG number *i* consists of all resource elements with number *i* in that PRB pair and there are thus 16 EREGs in one PRB pair. However, note that not all nine resource elements in an EREG may be available for EPDCCH usage as some resource elements are occupied, for example, by the PDCCH control region, CRS, or CSI-RS. In Figure 6.30, an example with EREG 0 in one PRB pair is shown.

The mapping of ECCEs to EREGs is different for localized and distributed transmission (see Figure 6.31).

For localized transmission, the intention is to provide possibilities to select the physical resources and antenna precoding based on the instantaneous channel conditions. This is useful when exploiting channel-dependent scheduling or in conjunction with multi-antenna schemes such as CoMP (see Chapter 13). Hence, an ECCE is mapped to EREGs in the *same* PRB pair, which allows the eNodeB to select the appropriate frequency region to which to transmit the EPDCCH. Only if one PRB pair is not sufficient to to carry one EPDDCH—that is, for the highest aggregation levels—a second PRB pair is used. A single antenna port is used for transmission of the EPDCCH. The DM-RS associated with the EPDCCH transmission[36] is a function of the ECCE index and the C-RNTI. This is useful to support MU-MIMO (see Section 6.3.5.1) in which multiple EPDCCHs intended for different spatially separated devices are transmitted using the *same* time−frequency resources but with

[36]In the specifications, this is described as selecting *one* of *four* antenna ports for transmission of the EPDCCH, where each antenna port has an associated orthogonal reference signal.

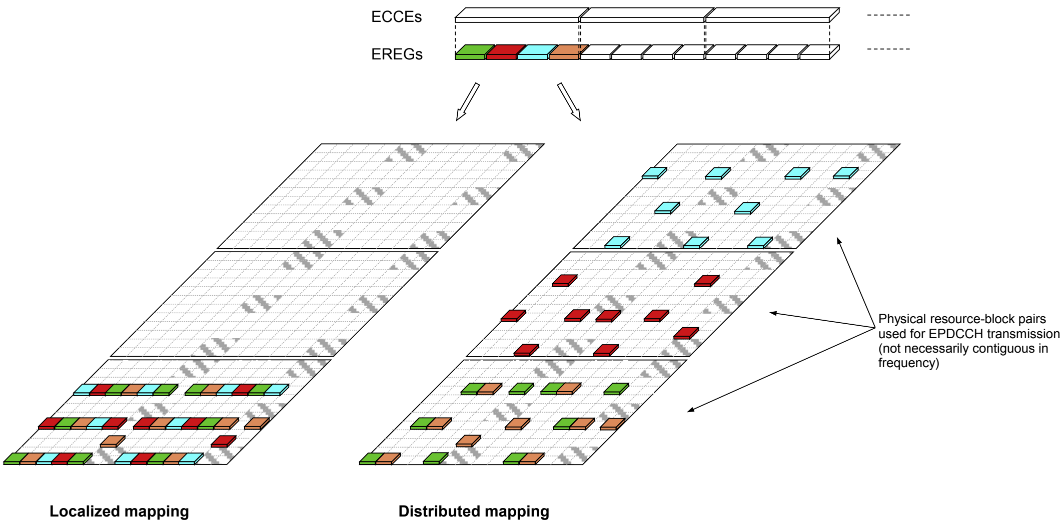

FIGURE 6.31

Example of localized and distributed mapping of ECCEs (three PRB pairs for EPDCCH assumed in this example).

different DM-RS sequences. Up to four different orthogonal DM-RS sequences are available, implying that up to four different devices can be multiplexed in this manner.

For distributed transmission, the intention is to maximize diversity to provide diversity gain in situations in which it is not possible or desirable to exploit the instantaneous channel conditions. Consequently, an ECCE is mapped to EREGs in *different* PRB pairs, where the PRB pairs preferably are configured to be far apart in the frequency domain. To further increase the diversity order and exploit multiple transmission antennas, the resource elements in each EREG is alternating between two antenna ports to provide spatial diversity.

Quasi-colocation configuration is supported for the EPDCCH. In transmission mode 10, each EPDCCH set is associated with one of the four *PDSCH to RE mapping and quasi-colocation states* (see Section 5.3.1). This can be used to enable CoMP and dynamic point selection (see Chapter 13). Each EPDCCH set is associated with a certain transmission point and dynamic point selection is achieved by selecting from which set to transmit a DCI message. Handling different sizes of the PDCCH control region and locations of CRS from the different transmission points is possible since these pieces of information are part of the four states.

Triggering of retransmissions is done in the same manner as for the PDCCH case—that is by using the PHICH which in essence is a very compact uplink grant for retransmissions. In principle an EPHICH could be considered, but since each EPHICH would need its own set of DM-RS, the resulting structure would not be as compact as the PHICH in which multiple devices share the same set of CRS. Hence, either the PHICH is used or, if the dependency on CRS is not desirable, an EPDCCH can be used to schedule a retransmission.

6.4.5 BLIND DECODING OF PDCCHS AND EPDCCHS

As described in the previous section, each PDCCH or EPDCCH supports multiple DCI formats and the format used is a priori unknown to the device. Therefore, the device needs to blindly detect the format of the (E)PDCCHs. The CCE and ECCE structures described in the previous section help in reducing the number of blind decoding attempts, but are not sufficient. Hence, it is required to have mechanisms to limit the number of CCE/ECCE aggregations that the device is supposed to monitor. Clearly, from a scheduling point of view, restrictions in the allowed aggregations are undesirable as they may influence the scheduling flexibility and require additional processing at the transmitter side. At the same time, requiring the device to monitor all possible CCE/ECCE aggregations, also for the larger cell bandwidths and EPDCCH set sizes, is not attractive from a device-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 device, LTE defines so-called *search spaces*. A search space is a set of candidate control channels formed by CCEs (or ECCEs) at a given aggregation level, which the device is supposed to attempt to decode. As there are multiple aggregation levels a device has multiple search spaces. The search space concept is applied to both PDCCH and EPDCCH decoding, although there are differences between the two. In the following section, search spaces are described, starting with the case of PDCCHs on a single component carrier and later extended to EPDCCHs as well as multiple component carriers.

The PDCCH supports four different aggregation levels corresponding to one, two, four, and eight CCEs. In each subframe, the devices 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 device and the device processes the information (scheduling assignment, scheduling grants, and so on). The network can only address a device if the control information is transmitted on a PDCCH formed by the CCEs in one of the device's search spaces. For example, device A in Figure 6.32 cannot be addressed on a PDCCH starting at CCE number 20, whereas device B can. Furthermore, if device A is using CCEs 16–23, device 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 devices. From this it can be intuitively understood that for efficient utilization of the CCEs in the system, the search spaces should differ between devices. Each device in the system therefore has a *device-specific* search space at each aggregation level.

As the device-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 device-specific search space for each aggregation level. One possibility would be to let the network configure the device-specific search space in each

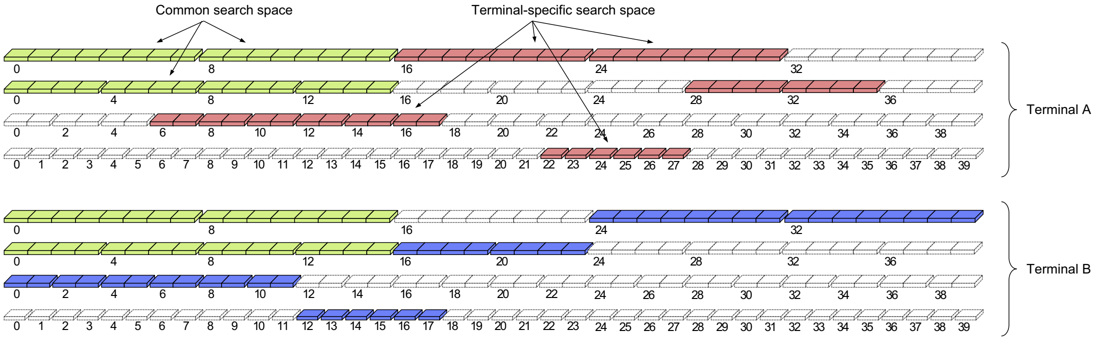

FIGURE 6.32

Illustration of principle for PDCCH search spaces in two devices.

device. However, this would require explicit signaling to each of the devices and possibly reconfiguration at handover. Instead, the device-specific search spaces for PDCCH are defined without explicit signaling through a function of the device identity and implicitly the subframe number. Dependence on the subframe number results in the device-specific search spaces being time varying, which helps resolve blocking between devices. If a given device cannot be scheduled in a subframe as all the CCEs that the device is monitoring have already been used for scheduling other devices in the same subframe, the time-varying definition of the device-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, devices in the system. One example is dynamic scheduling of system information; another is transmission of paging messages, both described in Chapter 11. Transmission of explicit power-control commands to a group of devices is a third example. To allow multiple devices to be addressed at the same time, LTE has defined *common search spaces* for the PDCCH in addition to the device-specific search spaces. A common search space is, as the name implies, common, and all devices in the cell monitor the CCEs in the common search spaces for PDCCH control information. Although the motivation for the common search space is primarily transmission of various system messages, it can be used to schedule individual devices as well. Thus, it can be used to resolve situations where scheduling of one device is blocked due to lack of available resources in the device-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 device, system messages typically need to reach the cell border. 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 device used for monitoring the common search space.

Figure 6.32 illustrates the device-specific and common search spaces for PDCCH in two devices in a certain subframe. The device-specific search spaces are different in the two devices and will, as described earlier, vary from subframe to subframe. Furthermore, the device-specific search spaces partially overlap between the two devices in this subframe (CCEs 24–31 on aggregation level 8) but, as the device-specific search space varies between subframes, the overlap in the next subframe is most likely different. There are 16 PDCCH candidates in the device-specific search spaces, mainly allocated to the smaller aggregation levels, and six PDCCH candidates in the common search spaces.

EPDCCH blind decoding in general follows the same principles as the PDCCH—that is, the device will attempt to decode all the EPDCCHs that can be formed from the ECCEs in each of its search spaces. However, only the device-specific search spaces are supported for EPDCCH. Thus, if a device has been configured to use the EPDCCH, it monitors the EPDCCH device-specific search spaces *instead of* the PDCCH device-specific search spaces. The common search spaces for PDCCH are monitored irrespective of whether an EPDCCH has been configured or not. The reason for not defining common search spaces for EPDCCH is that system information needs to be provided to all devices, including those not supporting EPDCCH, and hence the PDCCH needs to be used. The lack of a common search space for the EPDCCH is also the reason why DCI formats 3/3A and 1C are not supported on the EPDCCH.

The device-specific EPDCCH search spaces for an EPDCCH set are randomly varying within the set of PRB pairs configured for EPDCCH monitoring in a device using similar principles as the PDCCH. The two EPDCCH sets have independent pseudo-random sequences, initialized through RRC signaling. Having multiple EPDCCH sets is beneficial as it reduces the blocking probability: compare with the PDCCH, where one PDCCH on aggregation level eight may block all the PDCCH candidates on aggregation level one.

For a localized EPDCCH set, the EPDCCH candidates are spread over as many PRB pairs as possible. The intention behind this is to allow the device to be addressed over a wide frequency range in order not to limit the benefits of channel-dependent scheduling of EPDCCH.

The number of blind decoding attempts is the same irrespective of whether the device is monitoring the EPDCCH or PDCCH search spaces. Thus, there are 16 EPDCCH candidates[37] to attempt to decode, and for each candidate two different DCI formats need to be considered. The 16 EPDCCH candidates are distributed across the two EPDDCH sets, roughly following the principle that the number of candidates is proportional to the number of PRB pairs. Furthermore, for a localized set, more candidates are allocated to lower aggregation levels suitable for good channel conditions, motivated by the assumption that channel-dependent scheduling is used for localized transmission. For distributed transmission, the opposite holds—that is, more candidates are allocated to the higher aggregation levels than the lower ones to provide robustness as channel-dependent scheduling is typically not used for distributed transmission.

[37]For a small number of PRB pairs configured for EPDCCH there may be less than 16 EPDCCH candidates.

As mentioned, the device is monitoring *either* the device-specific PDCCH search spaces *or* the device-specific EPDCCH search spaces. The basic principle is that the device is monitoring EPDCCHs in all subframes whenever EPDCCH support is enabled, except for the special subframe in configurations where there are no DM-RS supported and hence no possibility to receive the EPDCCH. However, to provide additional control of when the EPDCCH is monitored, it is possible to provide a bitmap to the device, indicating in which subframes it should monitor EPDCCHs and in which subframes it should monitor PDCCHs. One possible use case for this is PDSCH transmission in MBSFN subframes. In MBSFN subframes where the PMCH is transmitted, it is not possible to transmit the EPDCCH and the device needs to be addressed using the PDCCH, while in MBSFN subframes not used for PMCH transmission, the EPDCCH can be used.

In the preceding paragraphs, the search spaces for PDCCH and EPDCCH in terms of control channel candidates are described. However, to determine whether a control channel candidate contains relevant downlink control information or not, the contents have to be decoded. If the CRC, which includes the device identity, checks, then the content of the control channel is declared as valid for this device and the device processes the information (scheduling assignment, scheduling grants, and so on). Thus, for each control-channel candidate, the device needs to attempt to decode the contents once for each supported DCI format. The downlink DCI formats to decode in the device-specific search spaces depend on the *transmission mode* configured for the device. Transmission modes are described in Section 6.3 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 device has not been configured for spatial multiplexing, which helps to reduce the number of blind decoding attempts. The DCI formats a device addressed using the C-RNTI should monitor as a function of the transmission mode are listed in Table 6.5. Note that DCI format 1C is monitored in the common search spaces as well, although not with the C-RNTI identity. As seen from the table, there are two DCI formats to monitor in the device-specific search spaces and one in the common search spaces. In addition, a device also needs to monitor DCI format 1C in the common search space. Hence, with 16 PDCCH/EPDCCH candidates in the device-specific search spaces and six in the common search spaces, a device needs to execute $2 \cdot 16 + 2 \cdot 6 = 44$ blind decoding attempts in each subframe. With uplink spatial multiplexing, introduced in release 10, an additional uplink DCI format needs to be monitored in the device-specific search spaces, increasing the number of blind decoding attempts to $3 \cdot 16 + 2 \cdot 6 = 60$. These numbers are for a single component carrier; in case of carrier aggregation the number of blind decodes is increased further as discussed in Chapter 12. Finally, note that a certain device in some circumstances may be addressed with different RNTIs. For example, DCI format 1A in the common search spaces may be used either with the C-RNTI for normal scheduling purposes or with the SI-RNTI for scheduling of system information. This does not affect the number of blind decoding attempts, as they are related to

Table 6.5 Downlink DCI Formats Monitored in Different Search Spaces for C-RNTI. Note that DCI Format 1C is Monitored in the Common Search Space as Well, Although Not with the C-RNTI Identity

Mode	Search Space			Description	Release
	Common (PDCCH)	Device-Specific (PDCCH or EPDCCH)			
1	1A	1A	1	Single antenna transmission	8
2			1	Transmit diversity	
3			2A	Open-loop spatial multiplexing	
4			2	Closed-loop spatial multiplexing	
5			1D	MU-MIMO	
6			1B	Single-layer codebook-based precoding	
7			1	Single-layer transmission using DM-RS	
8			2B	Dual-layer transmission using DM-RS	9
9			2C	Multi-layer transmission using DM-RS	10
10			2D	Multi-layer transmission using DM-RS	11

the DCI format; checking two different RNTIs—that is, checking two different CRCs, after decoding is a very low-complexity operation.

Configuration of the transmission mode is done via RRC signaling. As the exact subframe number when this configuration takes effect in the device is not specified and may vary depending on, for example, RLC retransmissions, there is a (short) period when the network and the device may have different understandings of which transmission mode is configured. Therefore, in order not to lose the possibility of communicating with the device, 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 device 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.

6.4.6 DOWNLINK SCHEDULING ASSIGNMENTS

Having described the transmission of downlink control information on PDCCH and EPDCCH, the detailed contents of the control information can be discussed, starting with the downlink scheduling assignments. 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, 2C, or 2D and the DCI formats used depend on the transmission mode configured (see Table 6.5 for the relation between DCI formats and transmission modes). 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 6.6, but 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 noncontiguous allocations of resource blocks and the full range of modulation-and-coding schemes.

Table 6.6 DCI Formats used for Downlink Scheduling

Field		DCI Format									
		1	1A	1B	1C	1D	2	2A	2B	2C	2D
Resource information	Carrier indicator	•	•	•		•	•	•	•	•	•
	Resource block assignment type	0/1	2	2	2’	2	0/1	0/1	0/1	0/1	0/1
HARQ process number		•	•	•		•	•	•	•	•	•
1st transport block	MCS	•	•	•	•	•	•	•	•	•	•
	RV	•	•	•		•	•	•	•	•	•
	NDI	•	•	•		•	•	•	•	•	•
2nd 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/ antenna ports									•	•
PDSCH mapping and quasi-colocation indicator											•
Downlink assignment index		•	•	•		•	•	•	•	•	•
PUCCH power control		•	•	•		•	•	•	•	•	•
SRS request[a]			F						T	T	T
ACK/NAK offset (EPDCCH only)			•	•		•	•	•	•	•	•
Flag for 0/1A differentiation			•								
Padding (only if needed)		(•)	(•)	(•)		(•)	(•)	(•)	(•)	(•)	(•)
Identity		•	•	•	•	•	•	•	•	•	

[a]*Format 1A for FDD and formats 2B, 2C, and 2D for TDD.*

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. Format 1A can also be used to trigger a contention-free random access (see Chapter 11), in which case some bit fields are used to convey the necessary random-access preamble information and the remaining bit fields are set to a specific combination.

DCI format 1B is used to support codebook-based beam-forming described in Section 6.3.3, 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 devices, 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, transmission of system information, MBMS-related signaling (see Chapter 19), and eIMTA support (Chapter 15). 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 MU-MIMO (transmission mode 5) scheduling of one codeword with precoder information. To support dynamic sharing of the transmission power between the devices sharing the same resource block in MU-MIMO, one bit of power offset information is included in DCI format 1D, as described in Section 6.3.5.2.

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 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; for two transmit antennas the number of layers is implicitly given by the number of transport blocks.

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 device 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 6.2.2.

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.

DCI format 2D was introduced in release 11 and is used to support spatial multiplexing using DM-RS (transmission mode 10). In essence it is an extension of format 2C to support signaling of quasi-colocation of antenna ports.

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. Furthermore, in later releases, some of the DCI formats are extended with additional bits. One example hereof is the addition of the carrier indicator in many DCI formats in release 10, as well as the inclusion of the SRS request in DCI formats 0 and 1A in release 10. Such extensions, for which different releases have different payload sizes for the same DCI format, are possible as long the extensions are used in the device-specific search spaces only where a single device is addressed by the DCI format in question. For DCI formats used to address multiple devices at the same time, for example, to broadcast system information, payload extensions are clearly not possible as devices from previous releases are not capable of decoding these extended formats. Consequently, the extensions are allowed in the device-specific search spaces only.

The information in the DCI formats used for downlink scheduling can be organized into different groups, as shown in Table 6.6, with the fields present varying between the DCI formats. A more detailed explanation of the contents of the different DCI formats is as follows:

- Resource information, consisting of:
  - Carrier indicator (0 or 3 bit). This field is present in releases 10 and beyond if cross-carrier scheduling is enabled via RRC signaling and is used to indicate the component carrier the downlink control information relates to (see Chapter 12). The carrier indicator is not present in the common search space as this would either impact compatibility with devices not capable of carrier aggregation or require additional blind decoding attempts.
  - Resource-block allocation. This field indicates the resource blocks on one component carrier upon which the device 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 6.4.6.1. Resource allocation types 0 and 1, which are the same size, support noncontiguous 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 device about the hybrid-ARQ process to use for soft combining. Not present in DCI format 1C as this DCI format is intended for scheduling of system information which does not use hybrid ARQ retransmissions.
- For the first (or only) transport block:[38]
  - Modulation-and-coding scheme (5 bit), used to provide the device with information about the modulation scheme, the code rate, and the transport-block size, as described later. DCI format 1C has a restricted size of this field as only QPSK is supported.
  - 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 6.5.
  - PMI confirmation (1 bit), present in format 1B only. Indicates whether the eNodeB uses the (frequency-selective) precoding matrix recommendation from the device or if the recommendation is overridden by the information in the PMI field.
  - Precoding information, providing information about the index of the precoding matrix used for the downlink transmission and, indirectly, about the number of transmission layers. This information is present in the DCI formats used for CRS-based transmission only; for DM-RS-based transmissions the precoder used is transparent for the device and consequently there is no need to signal this information in this case.
  - 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 CRS used to support dynamic power sharing between multiple devices for MU-MIMO.
  - Reference-signal scrambling sequence, used to control the generation of quasi-orthogonal DM-RS sequences, as discussed in Section 6.2.2.
  - Number of layers, reference-signal scrambling sequence and the set of antenna ports used for the transmission (jointly encoded information in releases 10 and later, 3 bits, possibility for extension to 4 bits in release 13).

[38] A transport block can be disabled by setting the modulation-and-coding scheme to zero and the RV to 1 in the DCI.

- PDSCH resource-element mapping and quasi-colocation indicator (2 bit), informing the device which set of parameters to assume when demodulating the PDSCH. Up to four different sets of parameters can be configured by RRC to support different CoMP schemes as discussed in Chapter 13.
- Downlink assignment index (2 bit), informing the device about the number of downlink transmissions for which a single hybrid-ARQ acknowledgment should be generated according to Section 8.1.3. Present for TDD only or for aggregation of more than 5 carriers in which case 4 bits are used.
- 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 *acknowledgment resource indicator* (ARI)—see Chapter 12.
- SRS request (1 bit). This field is used to trigger a one-shot transmission of a sounding reference signal in the uplink, a feature introduced in release 10 and an example of extending an existing DCI format in the device-specific search space with additional information fields in a later release. For FDD it is present in format 1A only, while for TDD it is present in formats 2B, 2C, and 2D only. For TDD, where short-term channel reciprocity can be used, it is motivated to include this field in the DCI formats used to support DM-RS-based multi-layer transmission in order to estimate the downlink channel conditions based on uplink channel sounding. For FDD, on the other hand, this would not be useful and the SRS request field is consequently not included. However, for DCI format 1A, including an SRS request bit comes "for free" as padding otherwise would have to be used to ensure the same payload size as for DCI format 0.
- ACK/NAK offset (2 bit). This field is present on EPDCCH only and thus supported in release 11 and later. It is used to dynamically control the PUCCH resource used for the hybrid-ARQ acknowledgment as discussed in Section 7.4.2.1.
- 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 that may occur simultaneously in the same search space (this is rarely required in practice as the payload sizes are different due to the different amounts of information). Finally, for PDCCH, padding is used to avoid certain DCI sizes that may cause ambiguous decoding.[39]

[39]For a small set of specific payload sizes, the control signaling on PDCCH 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 device. To overcome this, padding is used if necessary to avoid the problematic payload sizes. Note that this padding applied for PDCCH only.

- Identity (RNTI) of the device for which the PDSCH transmission is intended (16 bit). As described in Sections 6.4.3 and 6.4.4, 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.).

6.4.6.1 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 6.6. Resource-block allocation types 0 and 1 both support noncontiguous 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 device 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 device 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 6.33. 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 6.33, 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 address this by dividing the total number of resource blocks in the frequency domain into dispersed subsets, as shown in the middle of Figure 6.33. 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 6.33, there are two subsets, whereas for a cell bandwidth of 100 resource blocks

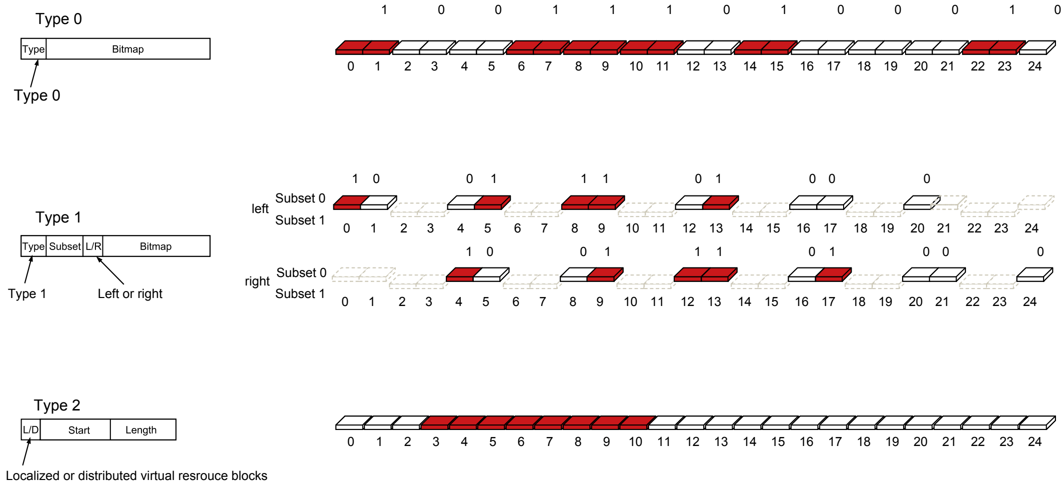

FIGURE 6.33

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

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 device whether resource allocation type 0 or 1 is used, the resource allocation field includes a flag for this purpose, denoted "type" in the leftmost part of Figure 6.33. For type 0, the only additional information is the bitmap discussed previously. For type 1, on the other hand, in addition to the bitmap 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,[40] 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 6.33.

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 6.34 and, as shown, the difference is fairly large for the larger cell bandwidths.

[40]Allowing different sizes would result in an increase in the number of blind decoding attempts required in the device.

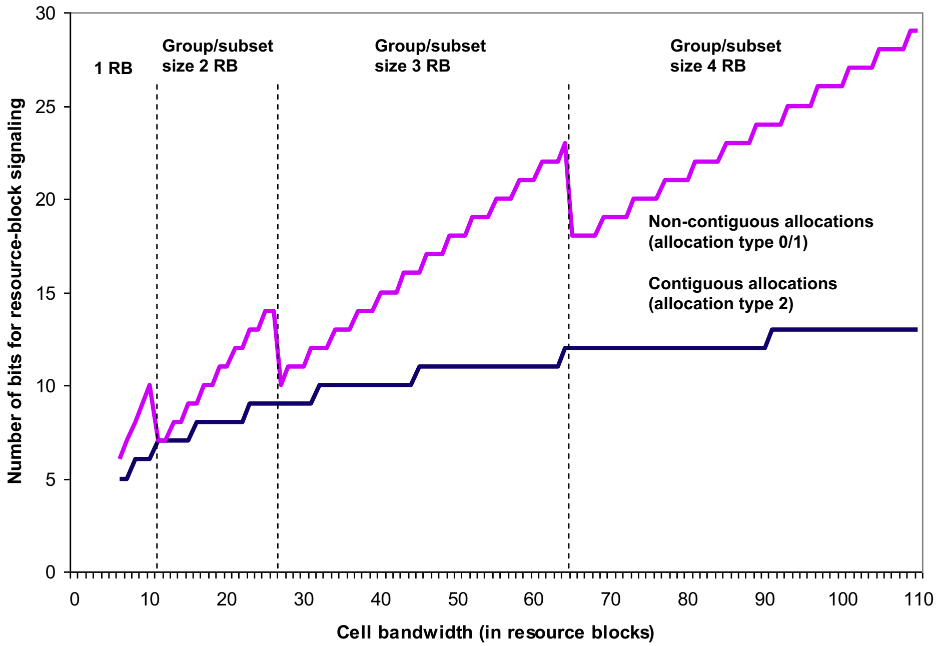

FIGURE 6.34

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

All three resource-allocation types refer to *VRBs* (see Section 6.1.1.8 for a discussion of resource-block types). For resource-allocation types 0 and 1, the VRBs are of localized type and the VRBs are directly mapped to PRBs. For resource-allocation type 2, on the other hand, both localized and distributed VRBs are supported. One bit in the resource allocation field indicates whether the allocation signaling refers to localized or distributed resource blocks.

6.4.6.2 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 5-bit field in the different DCI formats. Of the 32 combinations, 29 are used to signal the modulation-and-coding scheme whereas 3 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). For devices configured with support for 256QAM, 4 of the 32 combinations are reserved and an alternative 28 × 110 table is used instead[41] to support the larger transport block sizes. The principles in the following apply to both these cases, although the numbers given below assume no 256QAM support.

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 (e.g., 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 6.1.1.3). 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 said requirements. 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.[42] 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

[41]The alternative table is only used for transmissions scheduled with the C-RNTI. Random-access response and system information cannot use 256QAM for backward-compatibility reasons.

[42]For 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.

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 (with 256QAM the upper limit is 7.4 bit/s/symbol).[43] 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, and 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 of this section, those entries can be used for retransmissions only. In the case of a retransmission, the 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, or 16QAM or 64QAM,[44] 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 device 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 6.35.

6.4.7 UPLINK SCHEDULING GRANTS

Uplink scheduling grants use one of DCI formats 0 or 4; 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 device

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

[44]Four reserved values representing QPSK, 16QAM, 64QAM, and 256QAM for the alternative table.

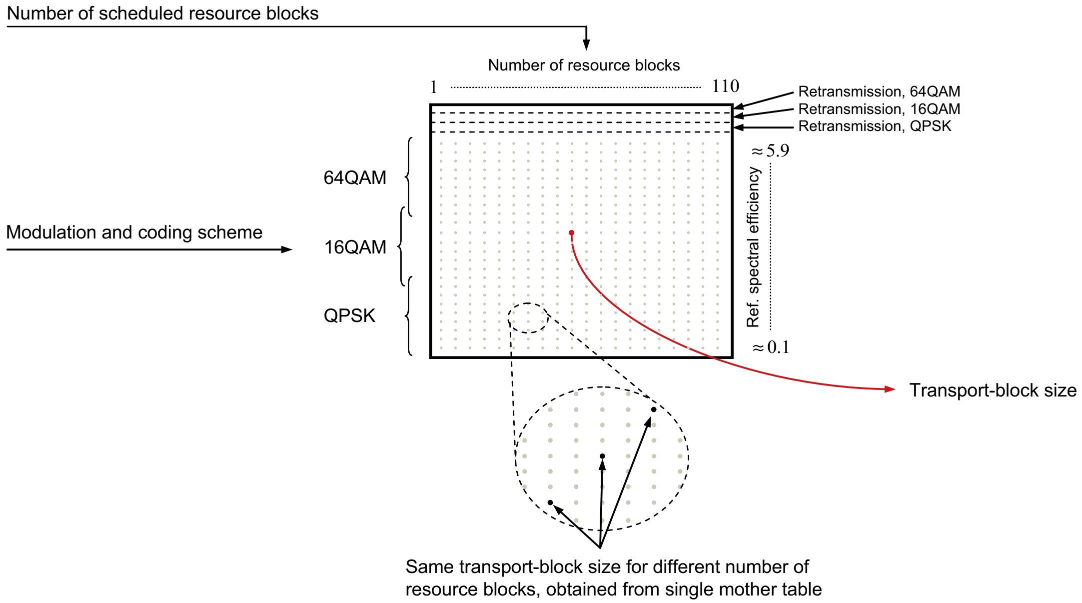

FIGURE 6.35

Computing the transport-block size (no 256QAM configured).

whether the message is an uplink scheduling grant (DCI format 0) or a downlink scheduling assignment (DCI format 1A).

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 6.7. Similarly to the downlink scheduling assignments, some DCI formats have been extended with additional bits in later releases. The contents of the different DCI formats are explained in more detail in the following:

- Resource information, consisting of:
  - Carrier indicator (0 or 3 bit). This field is present in releases 10 and beyond only if cross-carrier scheduling is enabled via RRC signaling and is used to indicate the uplink component carrier the grant relates to (see Chapter 12). The carrier indicator is not present in the common search space as this would either impact compatibility with devices not capable of carrier aggregation or require additional blind decoding attempts.
  - Resource allocation type or multi-cluster flag (1 bit), indicating whether resource allocation type 0 (one cluster of resource blocks is used for uplink transmissions) or type 1 (two clusters of resource blocks are used for the uplink transmission) is used.

Table 6.7 DCI Formats for Uplink Scheduling Grants

Field		DCI Format 0	DCI Format 4
Resource information	Carrier indicator	•	•
	Resource allocation type	•	•
	Resource block assignment	0/(1)	0/1
1st transport block	MCS/RV	•	•
	NDI	•	•
2nd transport block	MCS/RV		•
	NDI		•
DM-RS phase rotation and OCC index		•	•
Precoding information			•
CSI request		•	•
SRS request		•	•
Uplink index/DAI (TDD only)		•	•
PUSCH power control		•	•
Flag for 0/1A differentiation		•	
Padding (only if needed)		(•)	(•)
Identity		•	•

This flag is not present in releases prior to release 10. In previous releases, the downlink bandwidth was, in practice, always at least as large as the uplink bandwidth, implying that one padding bit was 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 backward compatibility. In DCI format 4, supported in releases 10 and later, the multi-cluster flag is always present.

- Resource-block allocation, including hopping indication. This field indicates the resource blocks upon which the device should transmit the PUSCH using uplink resource-allocation type 0 (DCI format 0) or type 1 (DCI format 4), as described in Section 6.4.7.1. The size of the field depends on the cell bandwidth. For single-cluster allocations, uplink frequency hopping, as described in Chapter 7, 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 device 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 devices,

devices 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 RV, as described later. 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 device 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 signal (3 bit), used to support MU-MIMO, as described in Chapter 7. By assigning different reference-signal phase rotations to devices scheduled on the same time—frequency resources, the eNodeB can estimate the uplink channel response from each device and suppress the interdevice interference by the appropriate processing. In releases 10 and later, it also controls the orthogonal cover sequence, see Section 7.2.
- Precoding information, used to signal the precoder to use for the uplink transmission in releases 10 and later.
- Channel-state request flag (1, 2, or 3 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 of up to five carriers, 2 bits are used to indicate which downlink component carrier the CSI should be reported for (see Chapter 10), a number increased to 3 bits if more than five carriers are configured in the device.
- SRS request (2 bit), used to trigger aperiodic sounding using one of up to three preconfigured settings, as discussed in Chapter 7. The SRS request, introduced in release 10, is supported in the device-specific search space only for reasons already described.
- Uplink index/DAI (2, 4 bit for carrier aggregation of more than five carriers). This field is present only when operating in TDD or when aggregating more than five carriers. 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 9. For other uplink—downlink configurations, it is used as downlink assignment index to indicate the number of downlink transmissions the eNodeB expects hybrid-ARQ acknowledgment 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.
- 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 DCI formats 0 and 4 (this is rarely required in practice as the payload sizes are different due to the different amounts of information). Finally, for PDCCH, padding is used to avoid certain DCI sizes that may cause ambiguous decoding.

- Identity (RNTI) of the device for which the grant is intended (16 bit). As described in Sections 6.4.3 and 6.4.4, 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 acknowledgment on the PHICH and not explicitly scheduled as for downlink data transmissions. However, as described in Chapter 8, 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 version 1, 2, and 3. If one of those values is signaled, the device 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.

6.4.7.1 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 releases 10 and later also provide 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 6.4.6.1 except that the single-bit flag indicating localized/distributed transmission is replaced by a single-bit frequency hopping flag. The resource allocation field in the DCI provides the set of VRBs to use for uplink transmission. The set of PRBs to use in the two slots of a subframe is controlled by the hopping flag, as described in Chapter 7.

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

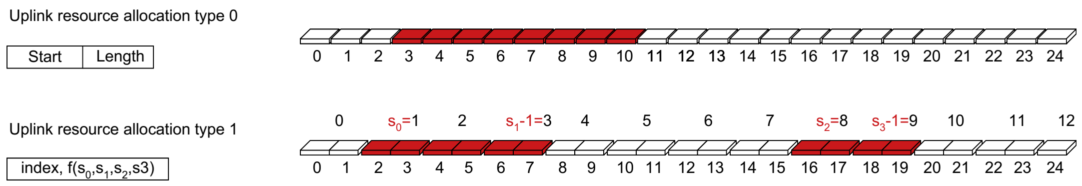

FIGURE 6.36

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

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 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 6.36, the cell bandwidth is 25 resource blocks, implying a group size of two resource blocks.

6.4.8 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 (2-bit command per device) or 3A (single-bit command per device). 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 devices using an RNTI specific for that group. Each device 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 devices, each device 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.