CHAPTER

3GPP PHYSICAL LAYER SOLUTIONS FOR LTE AND THE EVOLUTION TOWARD NR

8

The purpose of this chapter is to describe the parts of the 3GPP long-term evolution (LTE) specifications that are relevant for advanced antenna system (AAS). In Chapter 9, the corresponding new radio (NR) specifications are described. Also, in Chapter 6, the underlying concepts related to multi-antenna operation are introduced and motivated and this chapter describes how these concepts have been standardized and specified in *third generation partnership program* (3GPP) LTE specifications. It is recommended to be familiar with the key concepts in Chapter 6 to fully understand the motivation for multi-antenna standardization in 3GPP.

Moreover, to understand the motivation behind the concepts in the NR specifications, it is useful to get a grip on the fundamentals and shortcomings of LTE. This should motivate a reader interested in NR to also spend some time to understand LTE. Furthermore, central definitions such as antenna port and quasi colocation (QCL) are the same in LTE and NR and described more thoroughly in this LTE chapter.

The standardization work by 3GPP creates a multivendor solution by means of industry cooperation. This collaboration has created several successful telecommunication standards such as global system for mobile communications, wideband code division multiple access (WCDMA), LTE, and now NR. In the communication industry with global roaming of devices and many network components, there is no market without standards and interoperability. The key to the success has been open specifications and industry cooperation where all parties involved pool ideas to ensure the solution has cutting-edge technology that is practical to implement and can scale globally. A drawback with such cooperation is that compromises must always be made in the decision process which in some cases leads to an excess in the number of possible configurations. A single vendor telecommunication system would likely have done some parts slightly more efficient and streamlined, but the benefits such as crowdsourcing of ideas, resulting in global roaming technology and benefits of scale would have been lost.

The chapter is organized as follows: the basic principles of LTE are provided in Section 8.1 and the history and evolution of LTE from release 8 to 15 is given in Section 8.2. For the interested reader, a more detailed description of the LTE physical layer is provided in Section 8.3. A summary of LTE and "lessons learned" from this LTE evolution can be found in Section 8.4.

Finally, note that this section will by no means provide a complete description of 3GPP LTE physical layer specifications. The focus here is an in-depth explanation of the multi-antenna aspects, and less focus is given in this section to control signaling and other aspects such as connection establishment (random access procedures). For a broader presentation of LTE specifications, refer to Ref. [1]. There are also verticals evolving that build on LTE, for example, see the use of

Advanced Antenna Systems for 5G Network Deployments. DOI: https://doi.org/10.1016/B978-0-12-820046-9.00008-3

LTE for *Internet of things* (IoT) and *machine-type communications* as described in detail in Ref. [2]. Such technologies will not be discussed in this book.

8.1 LTE PHYSICAL LAYER—BASIC PRINCIPLES

The purpose of the standard is to define a common understanding of signals, procedures, and protocols exchanged between the network [for LTE physical layer specifications, the evolved Node B (eNB)] and the terminals (the user equipment, UEs) to allow for interoperability between network components from different vendors and also enable global roaming.

Furthermore, the core of the physical layer standard specification for data transmission and reception can be categorized into three main areas of functionality:

- Functionality to provide *synchronization* between eNB and UE, for tracking of the fading radio channel, including *channel analysis* where the UE estimates large-scale channel properties (see Section 3.5). This functionality is enabled by *synchronization and tracking reference signals (RSs)*;
- Definition of the *transmission schemes* for the data transmission and the associated *demodulation reference signals*;
- Functionality to provide *link adaptation* through *channel measurements*, the associated reporting of these measurements from UE to eNB, and the associated *measurement RSs*. Link adaptation is the procedure to adjust the modulation scheme and code rate of the transmission to the quality of the link. Link adaptation also includes a recommendation from the UE to the eNB of a multi-antenna transmit precoder, or a preferred transmit beam.

A brief overview of LTE is first presented before the three main areas of functionality as outlined above are discussed further.

In Fig. 8.1, an eNB and UE are communicating, and the different physical layer channels are illustrated, consisting of downlink (DL; eNB to UE) and uplink (UL; UE to eNB) transmissions. Note that eNB and UE are 3GPP terminology that corresponds to base station and wireless terminal in other sections of this book.

On the highest level, channels can be divided into control channels and data channels, where the purpose of the control channel is to assist the data channel reception, for example, carrying a data channel scheduling assignment. Also, channels can be of broadcasted type (to many UEs simultaneously or to a single, yet unknown UE) or of unicast type, to a single specific UE.

Broadcast channels cover the entire served area in DL. They are used by UE to find the network and perform synchronization [provided by the primary and secondary synchronization signals (PSS, SSS)] and to obtain the basic information necessary to access the network, known as broadcast information [carried by the physical broadcast channel (PBCH)]. Part of the control information is also broadcasted [common physical downlink control channel (PDCCH) or physical control format indicator channel (PCFICH)], while the other part is unicasted [dedicated PDCCH or physical hybrid ARQ indicator channel (PHICH)], intended for a single UE only, such as containing scheduling of data transmission or reception.

The data transmission is carried by one of multiple different transmission schemes and is using the unicast physical downlink shared channel (PDSCH) and physical uplink shared channel (PUSCH) for DL and UL, respectively. Finally, the physical random access channel (PRACH) is transmitted in UL for random access, e.g. in the initial access procedure and PUCCH is used for UL control signaling. In addition to these defined channels, there are RSs transmitted to support synchronization, mobility, and measurements.

These channels will be described further in detail in this section and are summarized in Table 8.2.

LTE uses orthogonal frequency division multiplexing (OFDM) (Section 5.1.1) for the DL transmissions and discrete Fourier transform (DFT)-spread OFDM (Section 5.1.5) for the UL. The basic LTE physical resource can thus be illustrated in a frequency domain representation by a time−frequency grid as illustrated in Fig. 8.1. The smallest physical resource in the OFDM based system is one *resource element* (RE) and if OFDM without DFT precoding is used, each RE contains one modulated symbol containing a variable number of bits depending on the link adaptation, that is, link signal-to-noise ratio. The used modulation constellations in LTE are shown in Table 8.1. Refer

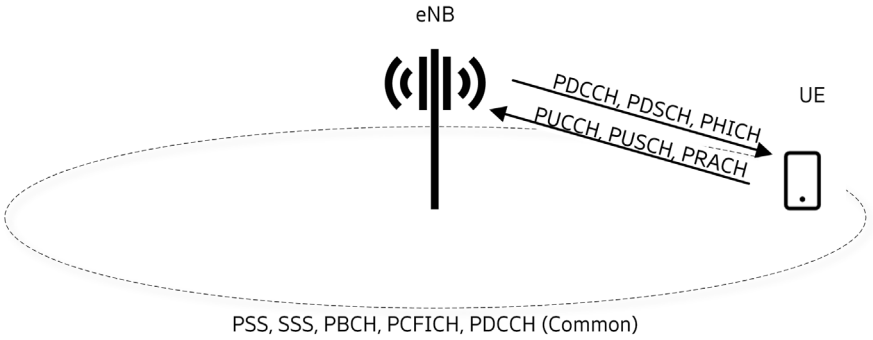

FIGURE 8.1

The downlink and uplink physical channels used for initial access, synchronization, broadcast, and unicast transmission in LTE. The unicast channels are only received by a single UE and can therefore be beamformed toward that UE. Other channels must have cell wide coverage as they may be received by many UEs simultanously.

Table 8.1 Modulation Constellations in LTE.

Modulation	Bits per Constellation Symbol
BPSK	1
QPSK	2
16-QAM	4
64-QAM	6
256-QAM	8

to Section 5.1 for an introduction to quadrature amplitude modulation (QAM) modulation. Each RE corresponds to one OFDM subcarrier during one OFDM symbol interval.

Furthermore, when scheduling data (see Section 2.1.1.2), that is, the PDSCH and PUSCH, the resource allocation is described in multiples of (physical) *resource blocks* (RBs) in frequency domain consisting of 12 contiguous subcarriers and a duration of 1-ms subframe in time domain (although later LTE versions introduced shorter scheduling durations to improve latency). The RB contains a control region in the beginning of the subframe (which can vary between 1 and 4 OFDM symbols) and the remainder of the subframe is a data region (see Fig. 8.2).

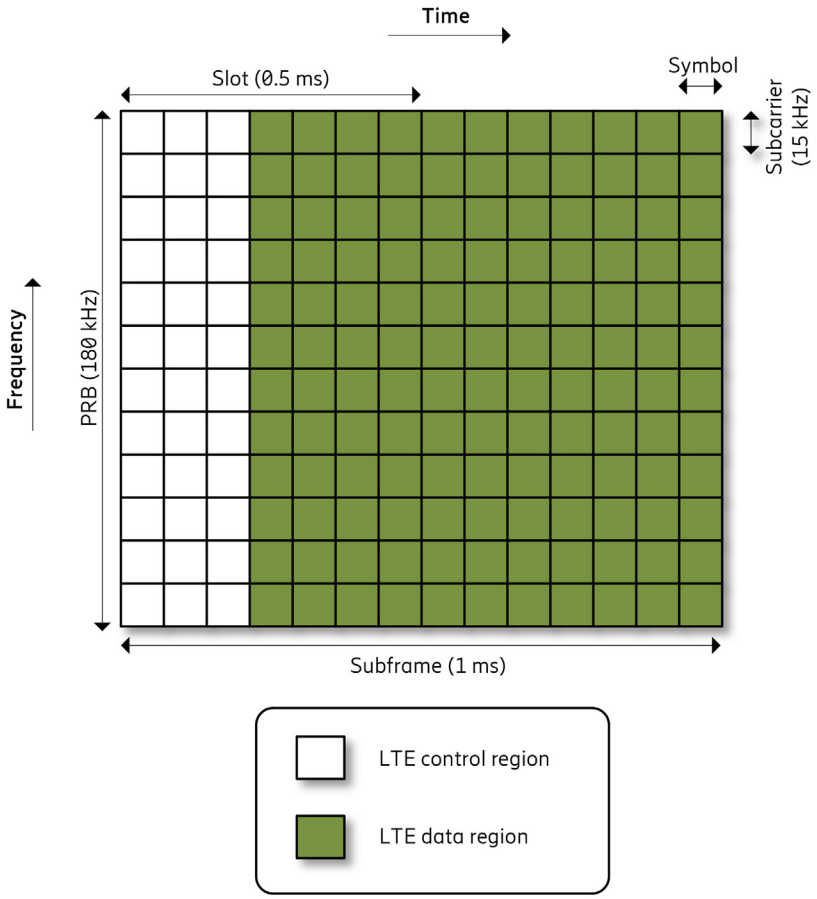

FIGURE 8.2

The LTE downlink physical resources showing an RB of 12 subcarriers and one LTE subframe (1 ms) of 14 OFDM symbols. It is divided into a control region and a data region (assuming FDD) where in this figure a control region length of three OFDM symbols is illustrated. LTE supports between one and four OFDM symbols for the control region.

Data can be scheduled contiguously (multiple adjacent RBs) or non-contiguously, to a given UE in the DL, and this allows for user multiplexing in frequency (see Fig. 8.3). The non-contiguous resource allocation is useful in case the channel has a higher signal-to-noise ratio (SNR) in some RBs than others (and the scheduler is SNR aware), then the transmission can use only these, so-called opportunistic frequency scheduling.

In the time domain, LTE DL subframes are organized into radio frames of 10 ms and each radio frame consisting of ten subframes of length 1 ms each (see Fig. 8.4).

The physical channels introduced in the initial release of LTE, release 8, are shown in Table 8.2 with a brief description of the content and use. The RS used to receive the channel is also given in the table.

For a UE to receive data in a subframe, the UE first decodes the PCFICH located in the first symbol of the subframe, to determine the length of the control region. This length can be

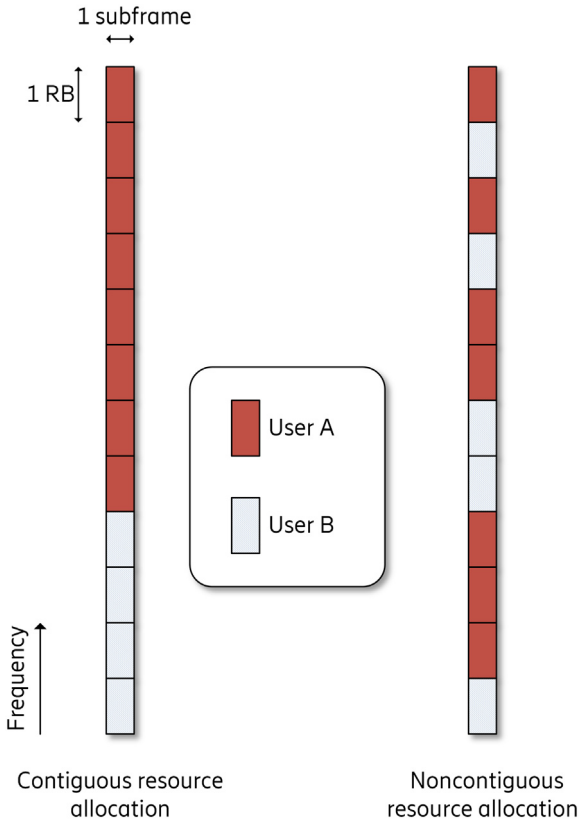

FIGURE 8.3

Resource allocation in basic LTE is performed in units of resource blocks in frequency and subframes in time. Resource allocations in the DL to a user can be both contiguous and non-contiguous.

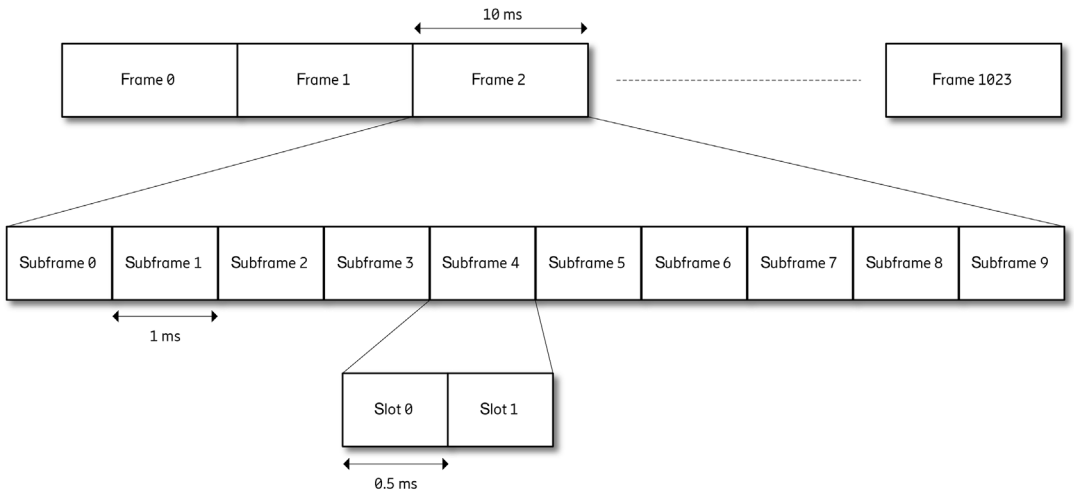

FIGURE 8.4

LTE time-domain frame, subframe, and slot structure.

Table 8.2 Physical Channels in LTE Release 8.

Physical Channel	Name of Physical Channel	Content	Demodulation Reference Signal
PBCH	Broadcast channel	Part of the system information to all UEs in the cell	CRS
PDCCH	Downlink control channel	DCI, for example, for scheduling PDSCH and PUSCH	CRS
PCFICH	Control format indicator channel	DCI to indicate the extent of the PDCCH in the same subframe	CRS
PHICH	HARQ indicator channel	DCI to indicate HARQ-ACK for PUSCH	CRS
PDSCH	Downlink shared channel	Downlink unicast data, paging, and part of the system information	CRS or DM-RS
PMCH	Multicast channel	Multicast data	MBSFN RS
PUSCH	Uplink shared channel	Scheduled uplink data transmission	DM-RS
PUCCH	Uplink control channel	UCI to indicate HARQ-ACK for PDSCH and for CSI feedback	DM-RS
PRACH	Random access channel	Nonscheduled uplink channel used, for example, in random access	N/A

dynamically changed by the network to 1, 2, or 3 OFDM symbols depending on the needed control channel (i.e., PDCCH) capacity (where four control channel symbols are only available for the smallest LTE system bandwidth of 1.4 MHz).

A PDCCH transmission contains downlink control information (DCI) to schedule the DL or the UL and is mapped to the REs in the control region, where these REs are distributed across the full system bandwidth to achieve diversity against fading and interference. The eNB can transmit multiple PDCCH simultaneously in nonoverlapping REs, and a UE decodes a set of PDCCHs in each subframe. This is known as blind decoding since the UE must decode multiple PDCCHs before it can determine whether any PDCCH included a DCI that was intended for the UE. The DCI may contain a scheduling assignment to receive a PDSCH in the data region of the same subframe or to transmit a PUSCH in UL in a later subframe. For frequency division duplex (FDD), the PUSCH is scheduled four subframes later than the reception of the PDCCH, for the UE to have time to prepare the UL transmission. See Fig. 8.5 for the timing for PDSCH and PUSCH (see also Section 8.3.3 on PDCCH transmission procedures and Section 8.3.4 on PDSCH transmission procedures). The acknowledgment (positive) (ACK) or negative ACK (NACK) (i.e., hybrid automatic repeat request (HARQ)-ACK) for the PDSCH reception is then reported to the eNB using uplink control information (UCI) mapped to either PUCCH or PUSCH, and for FDD, it is reported four subframes after the PDSCH (see Fig. 8.5). For time division duplex (TDD), the scheme gets more complicated as when determining the distance between PDSCH and ACK/NACK, then whether the ACK/NACK transmission subframe is a UL, a DL, or a mixed (special) subframe must be considered.

After scheduling and reception of a PUSCH at the eNB, the associated ACK or NACK (i.e., whether to perform a retransmission) may be conveyed to the UE, either by the PHICH or indirectly by a new PDCCH.

The UE also receives a broadcast channel, the PBCH that carries the master information block, that contains the information necessary to connect to the cell and to receive, for example, PDCCH. The PBCH is not scheduled by a PDCCH but is instead transmitted in a fixed location and format.

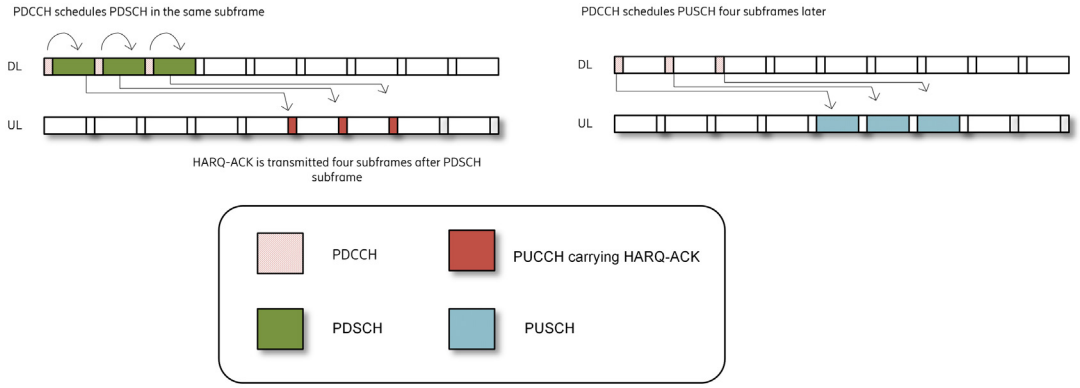

FIGURE 8.5

Scheduling timing for downlink (left) and uplink (right) for FDD.

In addition, the PMCH is a channel that optionally can be transmitted by the network, to support multicast services. Finally, PRACH is defined for UL transmissions to perform, for example, initial access to the network.

Now, how does a UE receive the DL transmissions with good quality? For coherent detection, the UE needs to know the channel over which the data has propagated so it can compensate for, for example, phase rotations induced by the channel. For this purpose, RSs are defined in specifications and transmitted by the eNB, and as these are known to the UE, the UE can then use these RSs to estimate the channel that they have propagated through. If the REs containing RS are dense enough, the UE can interpolate the channel for REs in between the RS and thus estimate the channel for each RE in between, i.e. for each transmitted modulated data symbol (e.g., quadrature phase shift keying, QPSK). The frequency and time separation between RSs in the time-frequency grid must be smaller than the coherence bandwidth and coherence time of the channel, respectively. Refer to Section 3.6 on multi-path propagation and channel coherence.

In LTE, there are two alternatives for RS that can be used for PDSCH demodulation, either *cell-specific RS* or *UE−specific RS*. The *cell-specific reference signals* (CRSs) transmitted in DL are shared by all the UEs served by the eNB (cell centric design), while the *UE−specific demodulation reference signal* (DM-RS) is dedicated for the demodulation of the transmission to or from one UE only (UE centric design). As this is an important differentiation, these two approaches will be further elaborated in the following paragraphs. The first LTE release relied heavily on the cell-specific RS, while a paradigm shift to UE−specific RSs occurred in LTE release 10. The UE−specific RSs were later used as a key principle of the design of NR (see Chapter 9).

8.1.1 KEY FUNCTIONALITIES

A first key functionality is *synchronization*, where the UE aligns the time−frequency grid of REs with the eNB transmitter; hence frequency offset estimation and time delay need to be estimated and compensated. In addition, the UE needs to find the frame and subframe boundaries as in Fig. 8.4 to be able to find, for example, the PDCCH so that the eNB can reach the UE with a scheduled message.

It is also beneficial for the UE to perform tracking of channel estimates over time, that is, across subframes, as this provides better performance compared to estimating the channel independently, on a per subframe basis. In addition, the receiver benefits from knowing the channels' large-scale properties, such as the Doppler shift and average delay (see further Section 3.5 on channel large-scale properties). If the UE knows these statistical properties of the channel prior to performing channel estimation, it can tune its channel estimation filter by using this as a priori information. This is important to ensure good reception performance, especially when higher-order modulation (e.g., 64- or 256-QAM) is used as these are more sensitive to channel estimation errors and transmitter-receiver misalignment.

For an LTE UE, this tracking and analysis functionality is always performed using measurements on the CRSs, irrespective of the configured transmission mode (TM). The CRS is present in every subframe and distributed across both the time and frequency, that is, with a density within the channel coherence limits (see Chapter 3) and is therefore a reliable, always present reference to base a channel analysis on.

Prior to LTE release 11, it was not specified that which RS the UE shall use to estimate the large-scale channel parameters; hence this was left open for UE implementation. However, since CRS is always present it is a reliable RS for this purpose and is thus typically used. In the later LTE releases and when TMs using DM-RS for demodulation was introduced, the specifications instead describe which RS the UE can use (or better use) for an estimate of these large-scale parameters. For this purpose, five such large-scale channel−fading parameters were defined in LTE release 11. As will be explained later, this was necessary to enable support for more advanced network deployments such as shared cells (see Section 6.6.1).

An LTE UE can be configured to operate according to 1 of 10 different *transmission modes* (TM) (see Section 8.3.4 for an overview of the TMs). If the UE supports LTE release 11 and when TM 9 or TM 10 is configured, not all large-scale channel−fading parameters should be derived from CRS, some should instead be derived from CSI-RS, and this is controlled by the introduction of the *quasi colocation* (QCL) feature in release 11. As will be seen in Section 9.3.1.6 for NR, a QCL framework is even more important and extensive, especially for high-band operation. Hence, understanding the basics of QCL principles is necessary for grasping the technical details of NR physical layer design (refer to Section 8.3.1.2 for a first introduction).

A stream of information bits is encoded by a Turbo encoder and then modulated (e.g., to QPSK, 16-QAM, 64-QAM, or 256-QAM) and mapped to the one or multiple multiple-input, multiple-output (MIMO) layers, thus supporting, for example, spatial multiplexing as described in Section 6.3. A *transmission scheme* for PDSCH or PUSCH describes this as the code word to layer mapping and the MIMO precoding. This mapping and precoding can be performed in various ways to achieve different purposes (e.g., maximize spectral efficiency or maximize diversity) (see Section 6.3.2.1). Hence, multiple such transmission schemes are specified in LTE. Each transmission scheme has a main design target, for example, providing maximal spectral efficiency, providing robustness by antenna diversity or providing robustness combined with increased spectral efficiency which is useful at higher UE velocities.

One scheme to maximize diversity in LTE is the well-known Alamouti diversity encoding [3] specified for LTE PDSCH and PDCCH. It can be described as a fixed MIMO linear dispersion block code [4], which can further be viewed as a generalization of the precoder concept in Section 6.3.3 to precoders spanning both the spatial and another domain such as frequency or time. Another scheme to achieve diversity that also supports spatial multiplexing is the large delay *cyclic delay diversity* [5] described as a set of precoders where one precoder from this set is cyclically selected and used per subcarrier, across the scheduled bandwidth.

When CRS is the RS used for demodulation of a certain transmission scheme, one unique CRS (from an available set of CRSs) is typically transmitted per eNB antenna. A typical case where CRS-based schemes are used is the conventional system with up to four radios and where the antenna consists of an antenna column with antenna elements of the same polarization. The mapping/precoding of a code word to these CRSs that is performed by the eNB, that is, the transmission scheme, must be known to the UE for proper PDSCH demodulation.

For transmission schemes where instead DM-RS is used for data demodulation, a set of DM-RS is used, and one unique DM-RS is associated with each PDSCH spatial layer. The same mapping of DM-RS as the data for the layer is used, which makes the de-mapping trivial and the used precoding by the eNB is then transparent to the UE. This is known as *noncodebook-based* transmission (see the concept described in Section 6.4.1.2). This also allows for the more advanced precoding

techniques such as the singular value decomposition and zero-forcing precoding (see Sections 6.3.3.2 and 6.3.3.3, respectively).

Therefore the DM-RS-based transmission schemes give greater flexibility to the transmitter to optimize the transmission and improve the link performance. Still, the transmitter needs to acquire some information on how to select the precoder and there are two main approaches to this as discussed in Section 6.4. Either obtained by measurements by the UE in the DL see Section 6.4.1, where the UE chooses a preferred precoder and uses feedback reporting to the network known as channel state information (CSI) reporting. Alternatively, this information can be obtained by the transmitter (eNB) itself by utilizing channel reciprocity (see Section 6.4.2) and thus by UL measurements using *sounding reference signals* (SRS).

The first approach, using CSI feedback will now be discussed further, as a key functionality of LTE. Note that even in the case of reciprocity-based operation, some CSI feedback as in the following section may still be necessary as a complement as discussed in Section 6.4.2.

Channel state information (CSI) measurements and reporting from the UE are necessary as described in Section 6.4.1, and in LTE it can be configured to assist the eNB in determining the precoder. Implicit feedback is used in LTE (see Section 6.4.1.2) and the most basic CSI report contains the channel quality indicator (CQI), a precoding matrix indicator (PMI), and a rank indicator (RI).

CSI measurements can be done using the CRS for CRS-based transmission schemes, or using dedicated CSI measurements signals, the CSI-RS for the DM-RS-based transmission schemes.

Fig. 8.6 shows the CSI feedback for a *CRS-based transmission scheme* (i.e., LTE TM 4), where in this example, four CRS are transmitted (see the related description of the *antenna ports* in Section 8.3.1 to fully understand the notion of a CRS port as in Fig. 8.6).

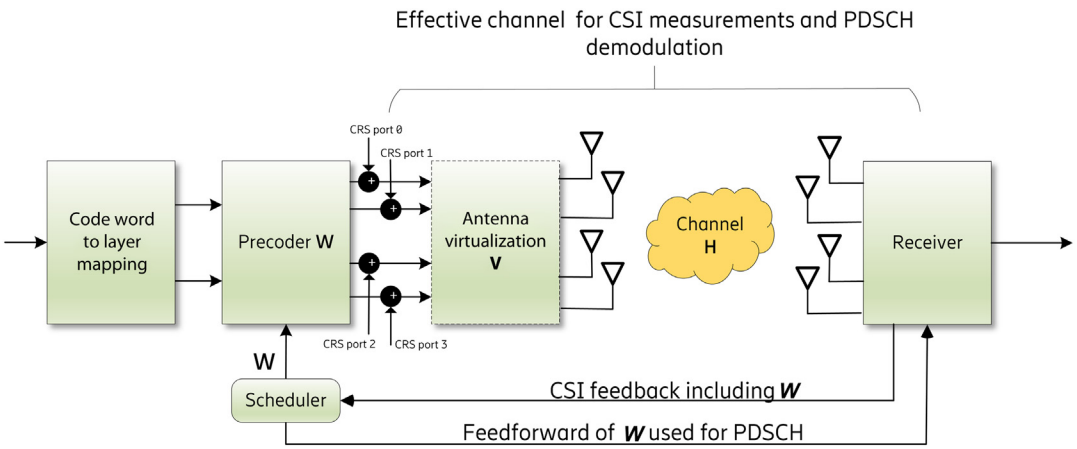

FIGURE 8.6

CSI feedback for a CRS-based transmission scheme, also illustrating the insertion point of CRS. The precoder is not part of the estimated combined channel $\boldsymbol{HV}$, and therefore the used precoder $\boldsymbol{W}$ must be indicated to the UE when scheduling PDSCH.

Antenna virtualization is used which maps the measurement signal (CRS in this case) to the four (or more) physical antennas. This virtualization step allows for cell shaping (cell-specific beamforming, see Section 6.2). Since the antenna virtualization is inserted *after* the measurement RSs (CRS in this case) in the direction of the signal flow, the used antenna virtualization matrix $\boldsymbol{V}$ is transparent to the UE as it becomes part of the combined channel which is the channel the UE observed by measuring on these CRSs. Hence, the UE cannot distinguish a change in $\boldsymbol{V}$ from a change in $\boldsymbol{H}$ as they are lumped together from a UE perspective. In the CRS-based scheme, the antenna virtualization $\boldsymbol{V}$ is the same for all UEs and typically does not change over time.

Antenna virtualization is necessary in the case where the eNB has more physical antennas than the maximum number of CRS that can be four. The input to the antenna virtualization can thus be described as a "virtual" antenna or antenna port. In the example shown in Fig. 8.6, there are however only four physical antennas illustrated and a one-to-one mapping can be used; one CRS can be transmitted from each of the four physical antennas of the eNB (i.e., the antenna virtualization matrix $\boldsymbol{V}$ is an identity matrix).

As described in Section 6.4.1, the UE performs channel estimation as well as interference estimation, in this case using the CRS. It selects the preferred precoder $\boldsymbol{W}$ from the specified precoder codebook and it selects the preferred rank. For the selected precoder and rank, the UE determines an associated CQI per code word (either one or two code words depending on the rank). It then feeds back the precoder, rank, and CQI information in the CSI report.

The receive data model from Section 6.3.1 is now used in the following description. When eNB decides to schedule a PDSCH transmission, the used precoder $\boldsymbol{W}$ from the codebook, which typically is obtained from the most recent CSI report, is indicated to the UE in the scheduling message. The UE then performs the CRS-based channel estimation again to obtain an estimate of the combined channel $\boldsymbol{HV}$ per subcarrier and then internally multiplies $\boldsymbol{HV}$ with the indicated precoder $\boldsymbol{W}$ from the eNB, to arrive at the effective channel $\boldsymbol{H}_{\mathrm{e}} = \boldsymbol{HVW}$, which is a matrix with the same number of columns as the number of layers, which describes the effective channel of each transmitted PDSCH layer.

Note that the CRS is used for *both* measurements and demodulation in this case. Also, since the used precoder $\boldsymbol{W}$ is not part of the combined channel that the UE estimates, it must explicitly be indicated to the UE in the scheduling (DCI) message.

Fig. 8.7 shows the CSI feedback for a *DM-RS-based transmission scheme* (e.g., LTE TM 9) where antenna virtualization is again used, which maps the measurement signal (CSI-RS in this case, which is a UE−specific RS) to the physical antennas. Since CSI-RS is configured per UE, different UE can use different CSI-RS and this allows for beamforming of the RSs, for example, UE−specific beamforming of CSI-RS, tailored for a specific UE. Hence, the antenna virtualization can when CSI-RS is used be dynamic and change, as opposed to the antenna virtualization in the CRS-based scheme. As antenna virtualization is inserted after the measurement RSs, the used $\boldsymbol{V}$ is also in this case transparent to the UE as it becomes part of the effective channel.

The antenna virtualization $\boldsymbol{V}$ can be wideband and fixed by antenna implementation as in the CRS-based case, or allow some functionality to provide for some adaptation of virtualization on a slow basis, hence to be semi-static over a long time. As mentioned, for DM-RS-based schemes it may vary dynamically, although still with some restriction on its time dynamics. The reason for such restriction is if the UE performs a channel measurement when a certain antenna virtualization

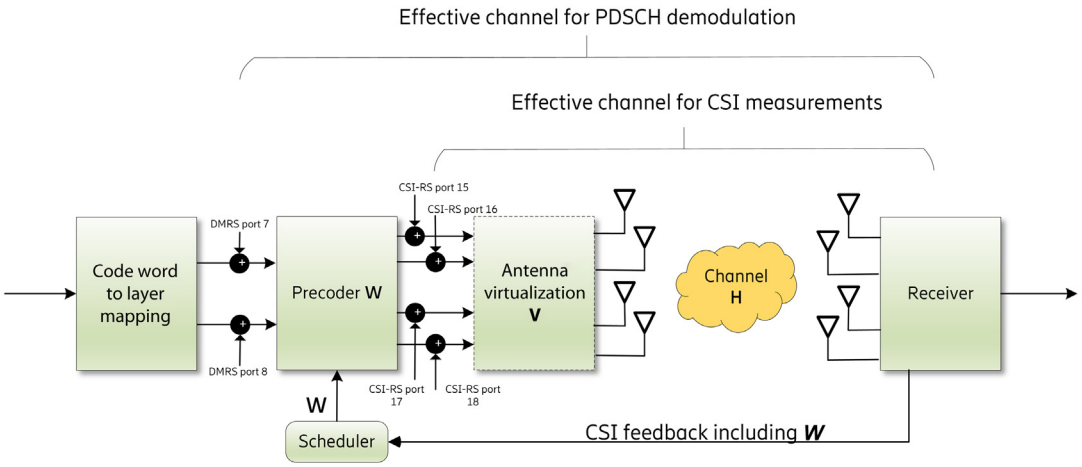

FIGURE 8.7

CSI feedback for a DM-RS-based transmission scheme illustrating the different insertion points of the DM-RS ports and CSI-RS ports. The precoder $\boldsymbol{W}$ is part of the effective channel for PDSCH demodulation; hence there is no need to indicate the choice of precoder to the UE and this gives full freedom to the network to select $\boldsymbol{W}$. Note that even in DM-RS-based transmission schemes in LTE, the CRS ports are always transmitted as well since they are cell-specific.

$\boldsymbol{V}$ is applied, then the same antenna virtualization $\boldsymbol{V}$ must be used in the eNB when transmitting the subsequent PDSCH. Otherwise, if a new virtualization $\boldsymbol{V}'$ is applied when PDSCH is transmitted, then the CSI report (especially the CQI and rank) and hence PDSCH link adaptation is valid for the wrong, that is, the previous virtualization $\boldsymbol{V}$.

To summarize, the CSI-RS is mapped to the virtual antennas, and the virtual antennas are further mapped to one or more physical antennas by a specification transparent mapping $\boldsymbol{V}$. Such virtual mapping could, for example, be a UE−specific beamforming of the CSI-RS in some direction. This is known as Class B operation in LTE CSI configuration. It could also be a static virtualization, typically using a one-to-one mapping between virtual antennas and physical antennas. This is referred to as Class A operation. In this case, as there is no UE−specific virtualization, typically all UEs in the cell share the same CSI-RS for their measurements of the DL channel. For Class A, the number of simultaneous CSI-RS is usually large since typically one measurement RS is needed per physical antenna. LTE supports up to 32 such simultaneous CSI-RS (referred to as 32 CSI-RS *antenna ports* in Section 8.3.1).

The CSI estimation principles for the DM-RS-based transmission schemes are the same as in the CRS-based schemes, using implicit CSI reporting, as defined in Section 6.4.1.2. The UE measures the channel by the CSI-RS over the antenna virtualization; it is thus measuring the combined radio channel $\boldsymbol{HV}$. It suggests a preferred precoder $\boldsymbol{W}$, preferred rank, and associated CQI computed based on the effective channel $\boldsymbol{H}_e = \boldsymbol{HVW}$ per subcarrier and feeds back the jointly determined precoder(s), rank and channel quality indicator(s) (CQI) to the eNB.

The eNB then schedules PDSCH and transmits one DM-RS per PDSCH layer and each DM-RS precoded in the same way as each associated PDSCH layer and the UE then estimates the effective channel $\boldsymbol{H}_e = \boldsymbol{HVW}$. The choice of $\boldsymbol{W}$ for PDSCH transmission is transparent to the receiver, giving full freedom to the eNB to modify. This is different compared to the CRS-based scheme where $\boldsymbol{W}$ is bound to a specified codebook. Note however that if eNB changes $\boldsymbol{W}$, or even changes the rank of the transmission, compared to what was indicated in the CSI report, then the CQI in the report is no longer valid and the eNB needs to recompute CQI to adjust the modulation and code rate of the transmission. This recomputation will then be an approximation since the exact effect on CQI when changing the precoder is not known to the eNB.

In addition, since one DM-RS is associated with each layer, the DM-RS overhead is in principle proportional to the number of MIMO layers. For CRS-based demodulation on the other hand, the number of CRSs is the same as the number of base station antennas (or virtual antennas in case of antenna virtualization); hence four CRSs are needed for a base station with four antennas even if only a single layer is transmitted, which gives an overhead drawback for low-rank transmissions.

8.2 LTE HISTORY AND EVOLUTION

In this section, the evolution in 3GPP of LTE multi-antenna features from the first to current release will be described, and in Fig. 8.8, this evolution is summarized in an illustration. Refer also to Section 6.3 for a general discussion to multi-antenna principles and Section 8.3.4 for a release-by-release introduction of LTE enhancements.

The first three releases of a radio access technology generation are the most important and contain the basic functionality. Although features added in later releases are well motivated and useful, they are most often optionally supported by the UE, so they have a higher threshold to be implemented and deployed. Also, they are in some cases functionality-wise hampered by the backward compatibility constraints. This is unfortunate, since very good technology components can be added in a later release but never realized in implementation. A striking example is the TM 10 in LTE, which has superior performance compared to TM 9 and "older" TMs, but it is not used in real networks due to lack of UE implementation support. However, the key TM 10 concepts made a strong "come back" in NR as the first NR release is functionality-wise based on LTE TM 10.

Multi-antenna-related features have been extensively studied in 3GPP over the years and every release includes enhancements in this area. In 3GPP, a release takes roughly a year to complete and is organized into time-limited projects denoted *work items* (WI). These are used to enhance a specific technical area or a feature and can be carried out by one or multiple 3GPP working groups. Sometimes, benefits of a proposed enhancement are less known, and it is too uncertain to start a related WI. It is then desirable to first carry out an initial feasibility study, resulting in a technical report. This is undertaken in the context of a study item (SI) and may include commercial as well as technical considerations. If the result of the SI is positive, it is likely to be followed by a WI, leading to new or an update to an existing technical specification.

There have been three occasions of multi-antenna-related SI for LTE (in release 9, 11, and 13, respectively). For NR, an SI took place in release 14 that preceded the NR release 15 WI that carried out the NR standardization. In addition to technical enhancements, there have been two SI on spatial channel modeling in release 12 and 14 that have relevance for AAS since they extended

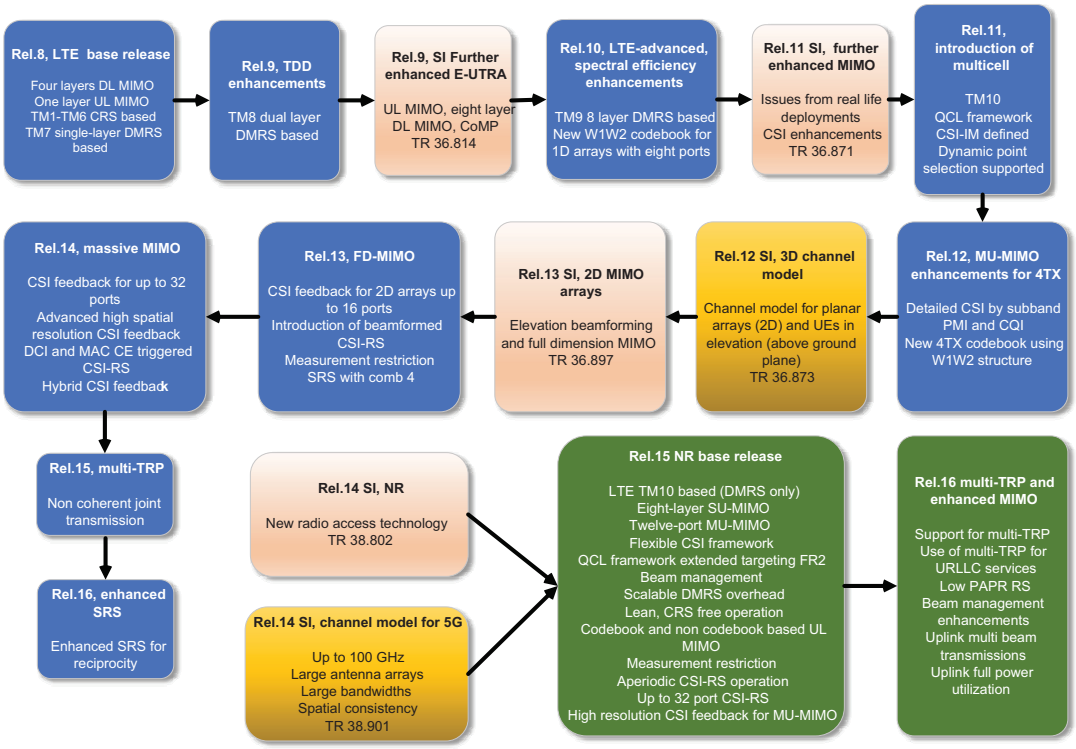

FIGURE 8.8

Overview of multi-antenna evolution in 3GPP, release by release for LTE and toward and including the introduction of NR. Blue box indicates an LTE work item, green box an NR work item, yellow box are channel model study items (SI), and peach-colored boxes represent LTE/NR SIs.

existing channel model to a three-dimensional (3D) channel model and further extended to the 5G channel models, respectively. See Section 3.6 for more in-depth discussion of channel modeling.

Release 8 of LTE specification was designed based on the requirements to support the operation of a single carrier component in a frequency band assigned for either FDD or TDD. The first release provided frequency flexibility supporting carrier bandwidths between 1.4 and 20 MHz multi-antenna transmission in DL was an integral part of release 8, supporting spatial multiplexing of up to four layers in the DL to the same UE, that is, single-user MIMO (SU-MIMO), and some basic multiple user MIMO (MU-MIMO) functionality. For the UL, release 8 did not contain any SU-MIMO transmission feature, although UL MU-MIMO is possible from release 8. Even in this first release, there was a large degree of configurability and hence a toolbox of features with seven different DL TMs, targeting different deployments and operating points.

LTE multi-antenna operation was then evolved during the following releases adding support for up to eight-layer reception in DL to a single-user, four-layer transmission in UL, support for measurements of eNB antenna arrays with up to 32 antenna ports (where antenna ports are defined in

Section 8.3.1) and in the later releases support for multipoint transmission of data using either *dynamic point selection* (DPS) or *noncoherent joint transmission* (NC-JT) which are flavors of coordinated multipoint transmission (see Section 6.6).

In release 10, IMT-Advanced requirements were defined by ITU, having a key feature of 1 GBit/s for low-mobility UEs, which then led 3GPP to conclude 3GPP requirements [6] for release 10 on a high-peak spectral efficiency of 30 bps/Hz for DL and 15 bps/Hz for UL.

The first LTE release did not fulfill the requirements for IMT-Advanced and enhancements of spectral efficiency was necessary. Therefore, the scope of release 10 included introduction of up to eight layer SU-MIMO in DL and four layers in UL. This release was preceded with a release 9 SI on physical layer enhancements in which multi-antenna aspects were included [7].

In release 11, yet another SI [8] concluded that MIMO operation could be further enhanced considering realistic deployment aspects such as noncolocated antennas at the network side (see Section 6.6). It was also observed that proposed methods to enhance the CSI feedback gave DL performance benefits at the expense of more feedback overhead. These findings were then explored in the WI of the following releases.

Prior to release 13, the CSI feedback framework supported only one-dimensional (1D) transmission steerability (i.e., in either vertical or in horizontal domain depending on how the antenna array is oriented relative to the horizon). In the scoping of release 13 enhancements of LTE, it was observed that the popularity of smartphone applications led to dramatic increase of wireless data traffic and MIMO was identified as one of the key technologies to address the increased capacity demands.

It was observed that to consider only the horizontal dimension to achieve the MIMO benefits is insufficient, especially in dense urban areas where most of the traffic demands arise, and where traffic distribution is both in vertical and horizontal dimensions. Therefore it was in release 13 decided to introduce standard support for CSI feedback for a two-dimensional (2D) planar antenna array to also exploit the vertical dimension. In parallel to the 3GPP working group (WG1), WG4 started an AAS SI in release 11, which turned into a WI in release 12. This work continued in release 13 where the first release of AAS specification was finalized (see Section 11.4). For details on how and when to utilize one or two dimensions, see the network performance analysis in Chapters 12 and 13.

At the time of writing, 3GPP has developed a large toolbox of features facilitating efficient multi-antenna operation for LTE. In Fig. 8.9, the 3GPP release toolbox for DL MIMO transmission for LTE is schematically described, where in a certain deployment and for a certain UE, the operator needs to choose which features should be enabled and which should remain disabled. A feature can be enabled only if the UE has reported capability to support the feature and if the network has implemented the feature. See a discussion on mandatory versus optional features in Section 10.1.3.

In Fig. 8.9, the 3GPP LTE feature tools have been grouped into which area of multi-antenna operation they mainly enhance. Note that the groups show enhancements over the baseline, for example, MU-MIMO can be used from release 8 with TM 7 and 8, while dedicated MU-MIMO enhancements were introduced in release 12 and 14.

In the following section, the tools specified in the LTE evolution from release 8 to release 15 are described in more detail. How and why these tools were developed and added to the 3GPP specifications will also be discussed.

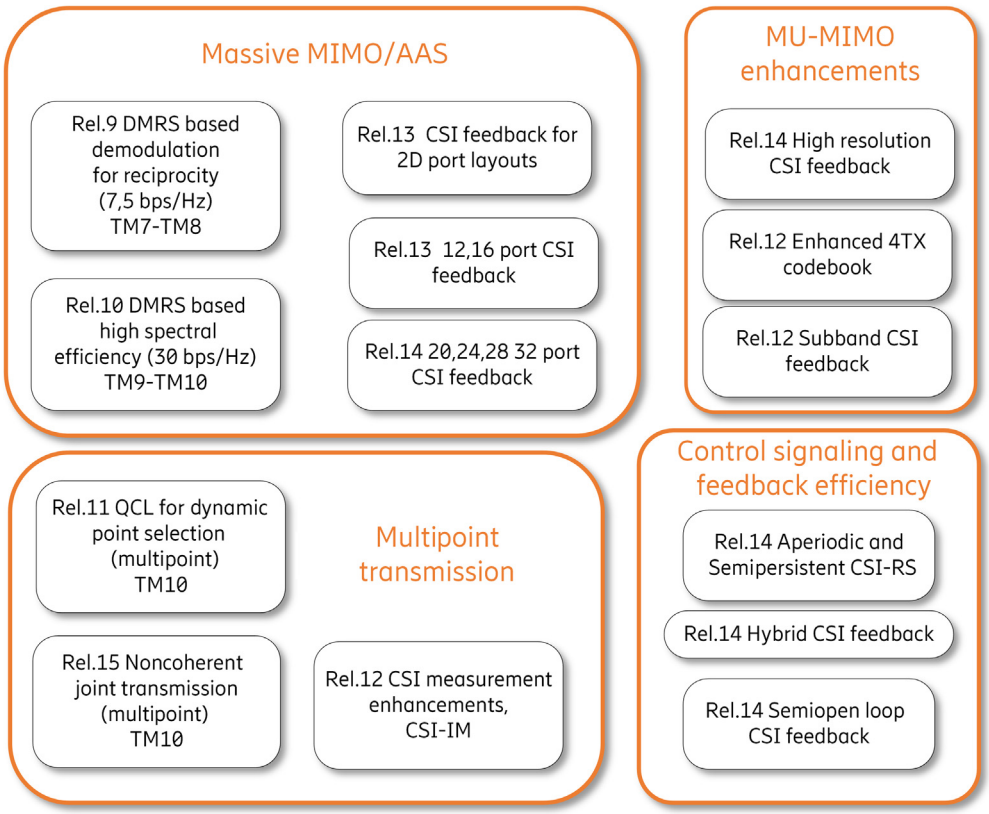

FIGURE 8.9

Overview of the multi-antenna toolbox provided by 3GPP standardization for LTE release 8—release 15. Note that since TM 9 and TM 10 support up to eight-layer SU-MIMO, the spectral efficiency per UE is as high as 30 bps/Hz, while TM 8 can only provide 7.5 bps/Hz per UE as there are at most two layers per UE. If MU-MIMO is used, then the sum spectral efficiency can be higher than the values indicated here.

8.3 LTE PHYSICAL LAYER SPECIFICATIONS FOR AAS

This section describes in more detail the specified functionality for the LTE physical layer where focus has been given to parts that relate to AAS.

8.3.1 ANTENNA PORTS, REFERENCE SIGNALS, AND QUASI COLOCATION

The channel that the UE estimates using transmitted RSs includes not only the propagation channel, but also the characteristics of the radio hardware such as the antenna elements, feeder cables, and

power amplifiers (PA), on both the transmitter and the receiver side, plus the possible use of antenna virtualization as discussed in Section 8.1.1.

Hence, from the UE radio channel estimation perspective and in 3GPP physical layer specifications (as defined by 3GPP RAN1), it is undesirable and unnecessary to define "physical antenna," since what matters for demodulation is to have an accurate enough representation of the channel for which a channel (e.g., PDSCH) is transmitted over, the UE does not care how an RS has been mapped to one or more physical antennas. Specifying physical antennas would restrict the flexibility in applying the specifications to the real-world base station designs as such antennas need to consider many factors (such as multi-band arrays, combined LTE and 3G operation, etc.).

Therefore 3GPP has in the physical layer specifications instead introduced the notion of *antenna ports* as a logical concept. The 3GPP definition is [9]:

> An antenna port is defined such that the channel over which a symbol on the antenna port is conveyed can be inferred from the channel over which another symbol on the same antenna port is conveyed.

Such definition allows for assuming and specifying relations between different channels and RSs without explicitly mentioning channel estimation in specifications. The term antenna port can therefore be an identifier for a channel. A modulated symbol of one or more physical channels is transmitted "on" a certain antenna port, which means the UE can use the estimated channel using the RS symbols associated with that antenna port to demodulate the physical channel symbols mapped on the same antenna port.

Consequently, when a UE is said to measure the channel from an eNB antenna port, it measures the effective radio channel at the receiver using the transmitted RS(s) associated with that antenna port. Refer to Fig. 8.7 where the different CSI-RS mapped to the input of the antenna virtualization would in a 3GPP specification correspond to different CSI-RS antenna port numbers.

Some antenna ports are in LTE specifications numbered as in Table 8.3. In addition, there are RSs in LTE for special purposes such as port 4 for demodulating PMCH, port 6 used for positioning measurements, and ports 107−110 for the enhanced control channel [enhanced PDCCH (EPDCCH)] DM-RS in release 11.

Finally, note that for radio performance requirements and conformance testing (see Chapter 10), the term antenna reference point is used, which should not be confused with the logical concept of antenna port defined here.

Table 8.3 Region of Validity for Some LTE Antenna Ports in Downlink.

Port Number	Downlink RS	Validity Bandwidth	Validity Time Duration
0−3	CRS	Whole system bandwidth	Across all time
7−14	PDSCH DM-RS	One PRG (1, 2, or 3 RB depending on system bandwidth)	One subframe
15−46	CSI-RS	Whole system bandwidth	Across all time *without* measurement restriction enabled One subframe *with* measurement restriction enabled

8.3.1.1 Validity region of an antenna port

Another aspect of antenna ports that often leads to confusion is that, for example, port 7 can be used twice in two adjacent subframes but they are not actually the same antenna port since a port is only valid in a certain region (time and frequency). Hence, in this example, the UE is not allowed to use measurements of port 7 in one subframe to demodulate a PDSCH in the next subframe.

This is because antenna ports have a specified region of validity. If an antenna port with a given port number is used in two different validity regions, then the two estimated channels may be completely different since the eNB may have changed the precoder $\boldsymbol{W}$ between the two regions (e.g., subframes), and the UE shall therefore not interpolate these estimates across the regions. For the port 7 example, the validity region in time is one subframe, so cross-subframe interpolation is not allowed even though the number of the port is 7 in both subframes.

In addition, there is another source of confusion when it comes to how antenna ports are used in 3GPP specifications. A UE can be configured to measure on multiple CSI-RS resources simultaneously, where each such resource, for example, have four ports. The use of multiple CSI-RS resources is for example used in the CSI feedback mode of Class B, where each resource is transmitted in a different beam. The four antenna ports in the example will be numbered as 15, 16, 17, and 18 in all these CSI-RS resources, but this does not mean that port 15 in the first resource is the same as port 15 in the second CSI-RS resource and so on. Antenna ports belonging to different CSI-RS resources are completely independent even if the antenna port number is the same.

For PDSCH DM-RS, the *precoder resource group* (PRG), defined as 1, 2, or 3 RB wide in frequency, is introduced to define the validity region of the port in the frequency direction. The purpose is that the eNB may use a different precoder in each PRG to optimize the precoding gains (also known as sub-band precoding) and hence the effective channel is different for each PRG.

An overview of the region of validity for some DL antenna ports is given in Table 8.3. Note that the release 13 feature *measurement restriction* provides a means for eNB to control the validity region in time for CSI-RS, Measurement restriction configurations are discussed more in Section 8.3.4. Note also that the validity region of CRS is unlimited, the UE can average the estimates across time and frequency as it desires, leading to very good channel estimation performance on CRS.

8.3.1.2 Quasi co-location

The UE receiver algorithm typically performs some channel analysis prior to channel estimation in order to tune the channel estimator filters and to set the correct gain of the receiver front end to utilize the full dynamic range of the receiver. For example, it is useful for a channel estimation algorithm to know the delay spread of the channel (see Section 3.6.11), as this a priori information generally improves the estimation performance.

Delay spread can be measured using a wide bandwidth CSI-RS and used to demodulate a PDSCH even if it is only scheduled to be transmitted in a single physical resource block (PRB). Using the PDSCH DM-RS only, for delay spread estimation purpose would lead to a poor estimate in this example since the DM-RS is only a single RB wide. Hence, there is a need to specify an association between different RSs, such as the wide bandwidth CSI-RS and the variable bandwidth PDSCH DM-RS, to make clear which channel properties estimated from one RS may be exploited to improve the estimate of a channel from another RS.

The 3GPP LTE specification therefore introduces the concept of *quasi co-location* (QCL) between antenna ports of the DL transmissions, represented by five large-scale channel parameters: delay spread, Doppler spread, Doppler shift, average gain, and average delay. Hence, loosely speaking, the channel of two antenna ports has, if they are QCL, the same channel *property/properties*, although the *actual channel realization* is not the same. This strong relation of properties is useful information for the receiver in the UE.

When two antenna ports are QCL with respect to one or more of these five parameters, then the UE may estimate those large-scale parameters by a measurement of the first antenna port. When the UE, possibly at a later point in time, desires to perform channel estimation of the channel using the second port, then it can base the selection of the channel estimation filter (e.g., using the estimated delay spread) on these estimated large-scale parameters. Hence, the UE can prepare the filter in advance to receiving the second port since it already knows these large-scale parameters (i.e., the channel properties) of the channel for the second port.

Note also that the QCL relation works in both directions so the second and first ports could in principle be exchanged in the previous paragraph. However, it is common that one port is more suitable than the other to determine large-scale channel properties (see Section 3.5), for example, one port may have a higher time and frequency density or have a larger bandwidth and may be periodically transmitted. So, in practical use cases, QCL relations are one-directional and sometimes the term source and target RSs are used to denote the first and second RS, particularly when they are of different type (e.g., CSI-RS (source) and DM-RS (target), respectively).

As an example of a QCL relation, the LTE specification describes that the UE may assume that CRS ports 0−3 are mutually QCL with respect to delay spread, Doppler spread, Doppler shift, average gain, and average delay. For example, delay spread of CRS port 0 and 1 is assumed to be the same.

The consequence of such specified QCL relation is that the specifications do not guarantee the support of a network deployment where CRS port 0 and port 1 are transmitted from two largely separated physical locations. This is because that such deployment may imply that, for example, average gain and average delay of the two-channel measurements in the UE receiver are far from being the same; hence the specified QCL relation of at least "average gain" between CRS port 0 and 1 would be violated.

Therefore the network has incentives to ensure that two ports in this example have the same value of the large-scale parameters, which can be ensured by having them transmitted physically close to each other (i.e., two neighboring antenna elements) or at least the channels measured from these two ports at the receiver side must behave as they are transmitted from antennas physically close to each other in order for performance to be guaranteed. This is also the origin of the term "quasi co-located," which comes from the Latin word *quasi*, meaning "as it were."

8.3.1.3 Downlink cell-specific reference signals

The CRSs are fundamental in LTE from the first release as to provide a stable, dense, and always present DL RS which the UE uses for multiple and widely different purposes. These purposes are fine time and frequency synchronization, estimation of large-scale channel parameters, measurements for CSI feedback, as well as serving as demodulation reference for receiving the DL physical channels PBCH, PDSCH, PDCCH, PCFICH, and PHICH.

Since the CRS is used for all different purposes it is designed to be very dense in time and frequency to perform well also in the most demanding case of data channel demodulation at high UE speed where the channel coherence in time (see Chapter 3) can be very short. A CRS port of the CRS occupies every sixth subcarrier in frequency (i.e., 90 kHz distance between frequency samples) and in total two or four OFDM symbols within a subframe (see Fig. 8.10).

The CRS is not configured by dedicated signaling from the network to the UE, it is instead implicitly configured. This allows UEs even in IDLE mode to determine and use the CRS of any cell. A UE can therefore read the information in the PBCH to obtain the system information of any cell without any action by the eNB.

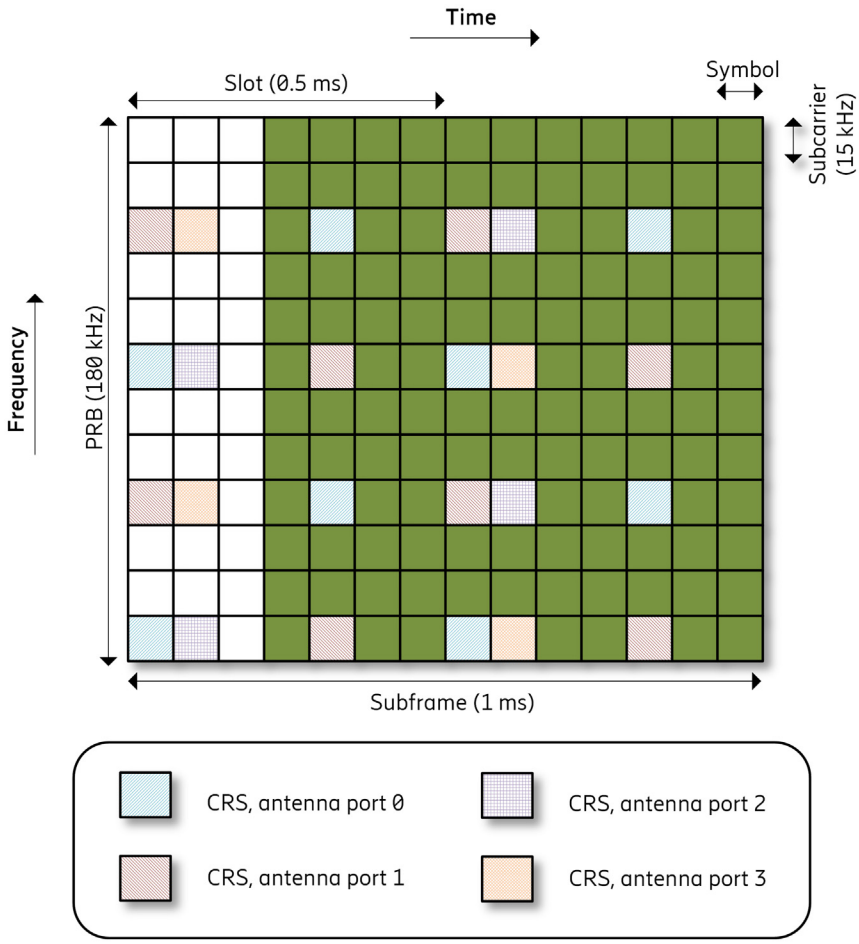

FIGURE 8.10

LTE CRS mapping to RE in subframe and RB assuming a control region of three symbols and a data region of 11 symbols. The figure shows four CRS ports in a normal subframe and with a subcarrier offset zero. Alternative CRS mappings is obtained by shifting the CRS pattern one or two subcarriers

The network configures the CRS transmission by the number of CRS ports (which can be a single port, two or four ports) and a subcarrier offset of one to three or one to six subcarriers. The available offsets depend on the number of configured CRS ports and which offset is used is linked to the physical cell ID (PCI) which the UE obtains from the PSS and SSS. The QPSK modulated symbol sequence used for CRS is obtained from a pseudo-noise (PN)-generated sequence for which initialization also depends on the PCI. In 3GPP specifications, the PN sequence is obtained by a specified pseudo-random sequence generator.

Hence, after the UE has synchronized to a cell by detecting the PSS and SSS, the PCI is obtained and also the CRS position and sequence become known to the UE. However, the number of ports the CRS is configured with is signaled in the PBCH message. To demodulate the PBCH, CRS is used, which leads to a chicken and egg problem as the number of CRS ports is needed to know the REs where the PBCH is mapped in the time−frequency grid. To resolve this paradox, the UE always assumes four ports for the PBCH mapping to REs, even though in reality fewer CRS ports may be transmitted by eNB.

CRS ports 2 and 3 are typically configured to be transmitted in case the eNB has capability to transmit (and UE to receive) more than two layers of PDSCH using a CRS-based transmission scheme. These ports have a lower time density compared to port 0 and 1 with only two OFDM symbols per subframe (see Fig. 8.10). The motivation was that higher-order MIMO transmission is only used by UEs with lower UE speed and thus two-channel time samples per subframe are enough, thereby saving some CRS overhead. Ports 2 and 3 are not used in mobility and reference signal received power measurements even if they are present, only ports 0 and 1 are used for this functionality.

Since the CRS is present in all DL subframes and covers the entire served cell, the UE can continuously track the channel, leading to excellent channel estimation performance and robustness for the CRS-based transmission schemes.

However, there is also a downside to this “always present” CRS. As LTE networks evolved in the subsequent LTE releases, the paradigm shifts toward using UE−specific RS for demodulation and dedicated CSI-RS for measurements instead of the “all-purpose” CRS took place. As there are legacy terminals connecting to the cell that use CRS and do not support the DM-RS-based TMs introduced in later releases of LTE, and since also IDLE UEs use CRS when measuring and connecting to a cell, the CRS must be always transmitted by the eNB in every subframe.

This “always on” CRS transmission leads to energy consumption by the base station even where there is currently no served UE. In addition, the CRS is an unnecessary overhead for UEs configured with the DM-RS and CSI-RS-based transmission schemes, since for such a UE, the CRS is only used for fine synchronization and for this purpose a dense RS in every subframe is unnecessary.

Moreover, the CRS from neighboring cells creates interference to PDSCH and PDCCH if the CRS has a subcarrier offset compared to the CRS of the serving cell, even if the interfering cell is not transmitting any data. On the other hand, if the same subcarrier shift is used, then there is another drawback related to CSI reporting. Since it is common for the UE to measure inter-cell interference as the residual power on the received serving cell CRS it means that the UE will measure high inter-cell interference even if the neighboring cell is not transmitting data, which gives incorrect CQI. These problems were one reason why dedicated interference measurements resources were introduced in TM 10 (see Section 8.3.4.3).

As mentioned earlier, the CRS ports 0−3 of a serving cell are QCL with respect to *delay spread, Doppler spread, Doppler shift, average gain, and average delay*, implying that these ports are all transmitted from the same eNB and with the same radiation pattern (in order to fulfill the QCL assumption on same average gain).

8.3.1.4 Downlink channel state information reference signals

The CSI-RS is introduced as a dedicated RS used for CSI measurement purposes and replaces the CRS for measurement purpose, at least for some TMs. When evolving LTE network deployments for heterogeneous networks and when introducing beamforming with advanced antenna arrays it was observed that the CRS-based measurement signals were too inflexible and restrictive. The LTE-Advanced requirements of a peak spectral efficiency of 30 bps/Hz implied a requirement of eight-layer transmission and extending the CRS-based transmission for this purpose (i.e., increasing the number of CRS ports to eight) was inefficient.

In addition, CRSs are shared by all UEs in the cell and hence, a CRS transmission cannot be channel-dependent, that is, it cannot be beamformed to a specific UE. It makes it difficult to fully utilize the antenna gains of AAS since the measurement RS will not benefit from beamforming gain. Moreover, to fully benefit from reciprocity-based operation, channel-dependent precoding/beamforming toward a specific UE was seen necessary *both* for the measurement signals and for the PDSCH transmission. When using TM 8 with reciprocity, the full benefit is not achieved since measurements are made on the CRS.

Therefore a UE specifically configured and *dedicated measurement signal*, the CSI-RS, was introduced in release 10 with 1, 2, 4, or 8 orthogonal antenna ports. In release 11, there was a need to be able to configure multiple CSI-RS to a UE simultaneously. This enables the UE to perform measurements and reporting of CSI for more than one transmission point for example. Therefore the concept of a *CSI-RS resource* was defined and a UE in TM 10 can have multiple such resources configured simultaneously and perform independent measurements on each of them.

Furthermore, in release 13 the definition of a CSI-RS resource was extended to have a larger number of antenna ports. The need for more ports was motivated by the introduction of 2D antenna port layouts. The 2D layouts allow precoding or beamforming in a given direction defined by both a vertical and a horizontal angle.

Note that the 2D antenna port layout (i.e., $N_1 \times N_2$ ports in first and second dimensions) is not part of the CSI-RS resource configuration, but instead it is part of the MIMO precoder codebook configuration for CSI feedback. The number of ports N_1 and N_2 in each dimension is configured by the network for the UE to use the correct codebook.

Although it is not specified that how CSI-RS ports are mapped to antennas (cf. the transparent antenna virtualization block in Fig. 8.7), the ports are numbered and the codebook is designed assuming a convention of mapping ports in a first dimension first, then a second dimension, and lastly across polarization. See Fig. 8.11 for the used convention.

The CSI-RS is UE specifically configured, that is, per dedicated signaling to each UE. However, the same CSI-RS resource may be configured to many or all UEs served by the eNB to save CSI-RS overhead, that is, they all measure on the same CSI-RS resource. In addition, the sequence used for CSI-RS is a PN sequence and the initialization is configured by dedicated radio resource control (RRC) signaling, hence it is not necessarily tied to the PCI (it is however possible to use the PCI as the initialization, if desired).

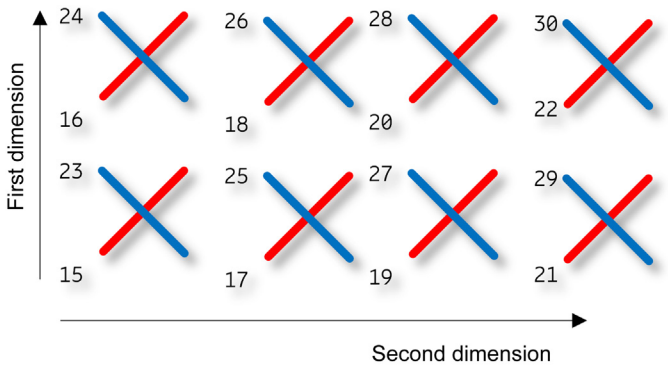

FIGURE 8.11

3GPP convention of CSI-RS port numbering and mapping to antennas for the 2×4 port layout. Numbering is across first dimension first (for one polarization) followed by second spatial dimension and lastly across polarizations. The UE is placed at the position of the reader, where port 15 is in the lower left corner. This example shows an antenna port layout with 2 and 4 ports in first and second dimensions, respectively, a total of 16 ports. Note that CSI-RS port numbering for a CSI-RS in LTE always begins with port 15.

Since CSI-RS is used only for measurements and not for demodulation or synchronization, the density is much lower than CRS; only 1 RE/port/PRB pair is used compared to 8 RE/port/PRB pair for CRS port 0. This density was seen enough during standardization since the CSI report is anyway quantized and there was not more performance benefit of having a larger density.

Instead, in release 14, even further reduced density for CSI-RS was introduced to reduce overhead allowing CSI-RS to be configured with 1/2 and 1/3 RE/port/PRB pair, but only for CSI-RS resources with 20 ports or more.

Fig. 8.12 shows a PRB pair and the total set of 40 REs that are available for a CSI-RS resource to be configured. As the first 1, 2, or 3 OFDM symbols in the subframe may carry control signaling, they are excluded from REs available for CSI-RS configuration. Also, symbols that may contain CRS are avoided. A CSI-RS resource is always allocated as symbol pairs and are thus located in OFDM symbol {6,7}, {10,11}, or {13,14}, respectively.

For example, if a resource with one or two ports is configured, only 2 REs out of the 40 available is used while if a CSI-RS resource of larger number of ports is required, then more REs will be used for the resource.

The CSI-RS, when configured, spans the whole system bandwidth and as can be seen in Table 8.3, the UE can utilize RS measurements from the whole bandwidth to estimate the channel of every subcarrier.

In specifications, the CSI-RS used to measure the channel is known as the *non-zero power* (NZP) CSI-RS. It is also possible to configure a CSI-RS that occupies the configured RE but the eNB does not transmit any energy in these RE, that is, they are empty. These REs are known as a *zero power* (ZP) CSI-RS resource, which has several use cases. For example, the NZP CSI-RS transmitted from cell A can be overlapped with a ZP CSI-RS from cell B. When the UE is measuring the channel using NZP CSI-RS, nothing is transmitted from cell B on these Res; hence there is

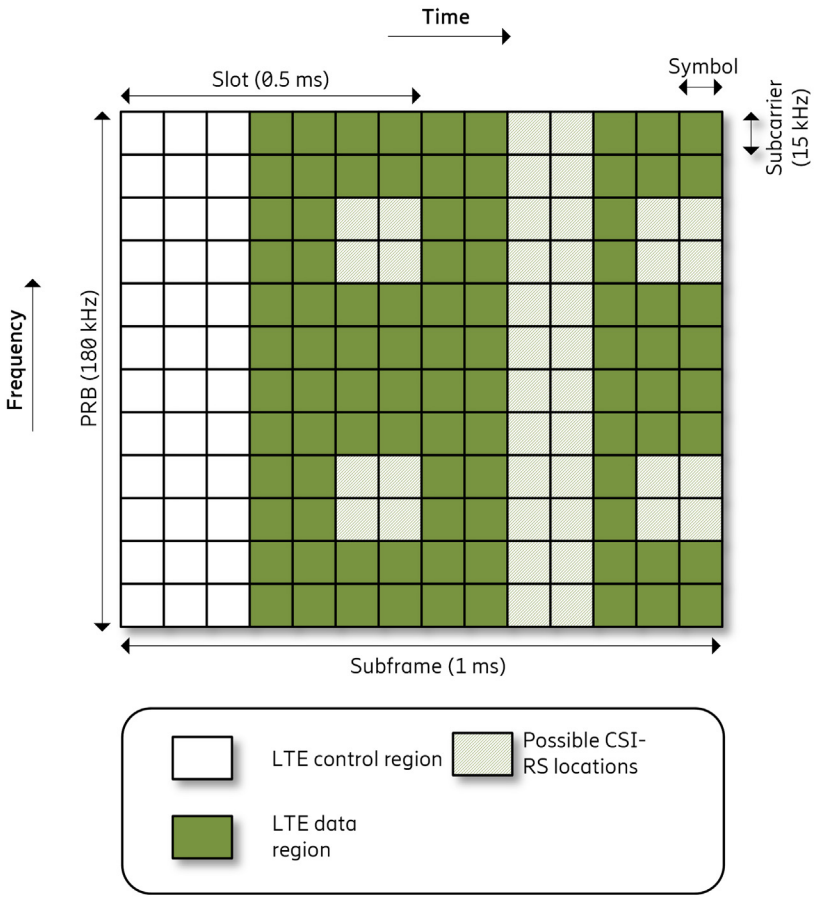

FIGURE 8.12

Possible REs where the LTE CSI-RS can be mapped in a subframe and one RB. For example, a 2 port CSI-RS use two of these 40 possible REs.When configured and present in a subframe, it is present in all RB in the system bandwidth.

no interference from cell B (provided that the propagation delay from these cells is comparable so the resources overlap). This improves the measurement performance of the channel from cell A. Hence, the ZP CSI-RS can be used *to protect* a configured NZP CSI-RS transmission in an adjacent cell.

In addition, the ZP CSI-RS is also used to define an *interference measurement resource* (CSI-IM). This provides the tool for making a "resource element hole" in the PDSCH transmission of the serving cell. The UE can measure the received power in this "hole" and thus it measures the level of interference from ongoing transmissions in neighbor cells without measuring the power received from the own cell (provided that a "hole" is not configured also for the interfering cell,

which is another possibility, It all depends on which *transmission hypothesis* the network wants the UE to feed back).

Finally, a note on QCL for NZP CSI-RS. The baseline is that CRS and CSI-RS are QCL with respect to Doppler shift, Doppler spread, average delay, and delay spread; this is denoted as *QCL Type A* in LTE. However, in TM 10 (introduced in Section 8.3.4), the UE can alternatively be configured so the CRS and CSI-RS ports are only QCL with respect to the large-scale parameters Doppler spread and Doppler shift (see Section 3.5). This is known as *QCL Type B* in LTE and is useful if *dynamic point selection* (DPS) feature is deployed, in which case CRS is transmitted from either or both transmission points and one CSI-RS is transmitted from each transmission point respectively.

By configuring QCL Type B, it implies that the UE cannot estimate, for example, average gain from the CRS when receiving CSI-RS of the serving cell as CRS and CSI-RS are not QCL with respect to average gain. From the deployment aspect, this means that the CSI-RS may be transmitted from a non-colocated position compared to the position where CRS is transmitted, for example, a different eNB than the one transmitting CRS. It enables the possibility to acquire a CSI report from the UE assuming PDSCH transmission from a different eNB than the serving eNB.

8.3.1.5 Demodulation reference signals for physical downlink control channel

The PDSCH is the physical channel carrying the data payload to the UE. The DL RS for demodulation (DM-RS) is a reference signal that is associated to PDSCH. It is only present in the subframe and RB where the PDSCH is transmitted and it is only used by the UE that receives the PDSCH. Hence, it is a UE−specific RS and each PDSCH layer is associated with a DM-RS port as discussed earlier in Section 8.3.

The PDSCH DM-RS existed already in LTE release 8 but only for a single-layer PDSCH transmission (port 5). In release 9, it was extended to two ports (7 and 8) and later in release 10 a further extension of up to 8 ports (port 7−15) was specified, to support the IMT-Advanced requirement of eight-layer SU-MIMO.

The DM-RS overhead is small, when the number of transmitted PDSCH layers is one or two, and the DM-RS is self-contained within the scheduled PDSCH resources. Hence there is no transmission outside the PDSCH RBs that could cause interference or prevent scheduling of transmissions to other UEs, which is a principle that also was adopted for NR.

In Fig. 8.13, the DM-RS ports mapped to RE in a PRB pair is illustrated (for FDD or for a DL subframe in TDD). As can be seen, a DM-RS port is mapped to three subcarriers in the last two symbols of each slot. Four ports are mapped to 12 REs, hence the density is three times higher than for CSI-RS. Note that the REs used for DM-RS do not collide with CRS or CSI-RS which was a design requirement since CRS and sometimes CSI-RS will be transmitted in the same subframe as the PDSCH DM-RS.

If the PDSCH is scheduled on one or two layers, then port 7 or port 7 and 9 is used and the DM-RS uses 12 REs, while if more than two-layer PDSCH is scheduled, then 24 REs are used for DM-RS since more than two ports are needed for the demodulation. For rank 5 and higher, the overhead remains at 24 REs used for DM-RS and additional DM-RS ports are then obtained by the *code division multiplexing* (CDM) of DM-RS ports. The CDM is implemented by an *orthogonal cover code* (OCC) which in this case is applied across all four DM-RS RE in a subcarrier to effectively obtain up to four DM-RS ports per 12 RE. Since the OCC code length is across RE in

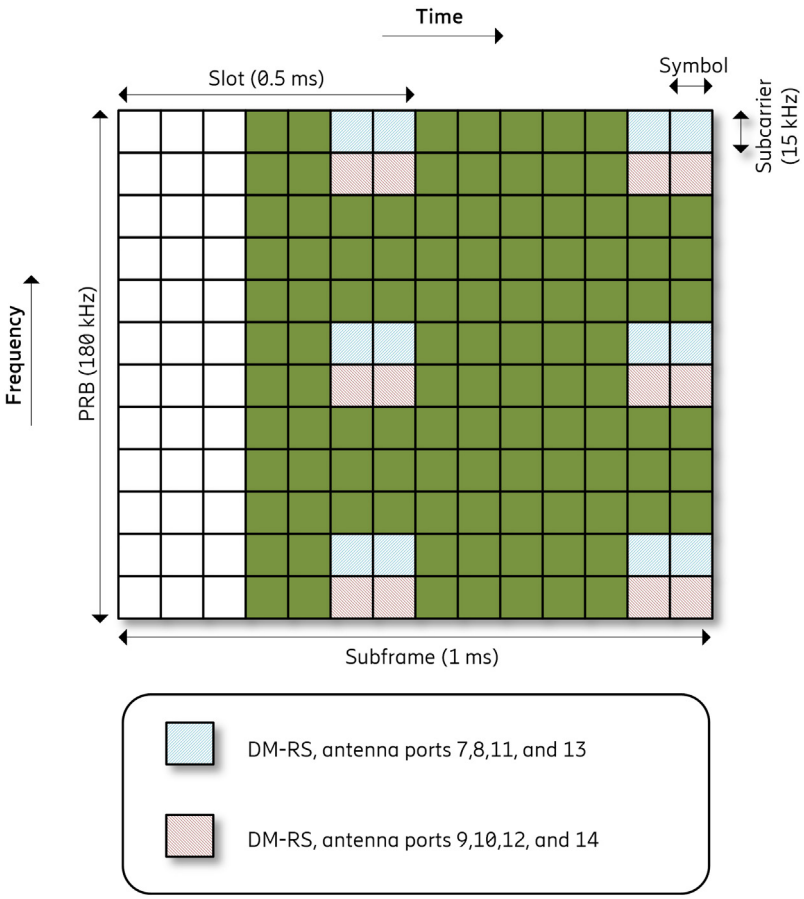

FIGURE 8.13

LTE DM-RS pattern for one subframe and one RB. Note that the CRS is not shown in this figure but will consume additional overhead as it is present in all subframes.

OFDM symbols of both slots, it means that such high-rank transmission is more sensitive to time variations in the channel as the channel should ideally be constant across all REs occupied by the OCC.

The sequence used for PDSCH DM-RS is a PN sequence and two sequence initialization seeds are configured by dedicated RRC signaling per UE. When scheduling the UE, the scheduling control message contains a $n_{SCID} = \{0,1\}$ parameter that selects either of the two PN sequences to be used to generate DM-RS of the scheduled PDSCH. If two DM-RS use different n_{SCID}, then these DM-RS are nonorthogonal.

When scheduling multi-user (MU) MIMO, that is, two or more UEs simultaneously in the same physical resources, then it is beneficial for channel estimation performance if orthogonal DM-RS

ports are used. For example, one UE is scheduled on one layer with port 7 and the other UE on one layer with port 8. For both UEs, the same n_{SCID} value is indicated in the scheduling message to ensure orthogonality between DM-RS ports.

In LTE release 10, the use of MU-MIMO is restricted to two UEs if orthogonal ports are used, since if a single PDSCH layer is scheduled for a UE, then it is only possible to indicate port 7 or port 8. Hence, it is not possible to schedule, for example, three UEs with orthogonal ports since the standard does not support signaling of port 9 as a single-layer PDSCH to a third UE.

There is however a possibility to use nonorthogonal DM-RS ports and schedule additional UEs in MU-MIMO since the n_{SCID} parameter can be utilized. In this case, one UE uses port 7 with $n_{SCID} = 0$, while another UE uses port 7 with $n_{SCID} = 1$. Using nonorthogonal ports may have poor performance unless the multi-antenna precoder ensures that the spatial suppression of cross-interference between the transmission to the two UEs is good enough.

In LTE release 12, MU-MIMO enhancements were introduced and the signaling framework for DM-RS antenna ports in the scheduling DCI for TM 9 and 10 was extended so that up to four UEs could be scheduled in MU-MIMO also with orthogonal DM-RS ports. There were also possibilities to schedule, for example, one UE with two layers and two UEs with single layer, and still maintain orthogonal DM-RS ports between these UEs. This should be compared to TM 8 of release 9, where nonorthogonal ports must be used between UEs for the same example.

As a final note on DL DM-RS; in release 11, the *enhanced PDCCH* (EPDCCH) was introduced to be able to benefit from UE−specific beamforming also for the control channel in addition to the already supported feature for the data channel. Hence, EPDCCH uses DM-RS and the same RE mapping as PDSCH DM-RS, but to avoid confusion, these ports were numbered as 107−110 instead of ports 7−10. The reason for having four ports for EPDCCH is not to enable spatial multiplexing of control channel transmission; instead they are used to support up to four EPDCCH transmissions in one PRB pair with individual and orthogonal ports so that they can be transmitted to four different UEs simultaneously.

8.3.1.6 Demodulation reference signals for physical uplink shared channel

The DM-RS for PUSCH is similar in functionality as the PDSCH DM-RS introduced in release 10, where each layer has its own DM-RS port. However, the UL transmission is codebook based, where the eNB selects the precoding matrix the UE should use for the PUSCH transmission. Note that even if the UE has four physical antennas, if a single PUSCH layer is transmitted with a precoder across the four antennas, then a single DM-RS port is transmitted by the UE.

Since PUSCH in LTE is based on DFT-spread OFDM, which generates a waveform with low *peak-to-average power ratio* (PAPR), the PUSCH DM-RS is designed with a low PAPR property as well.

The set of orthogonal DM-RS ports is obtained by a time-domain cyclic shift (CS) of the DM-RS sequence and by using an OCC across the two OFDM symbols carrying the DM-RS in the subframe. Due to the properties of the DM-RS sequence, using carefully selected CS, orthogonality between the DM-RS ports is ensured provided that the different RSs are time-aligned at the receiver and that the channel is sufficiently nonfrequency selective.

In case of a very frequency selective fading channel (large channel delay spread), ports separated by CS will not be orthogonal at the receiver and there will thus be cross-port interference. As an alternative, port separation using the time-domain OCC can be used instead. Each scheduling

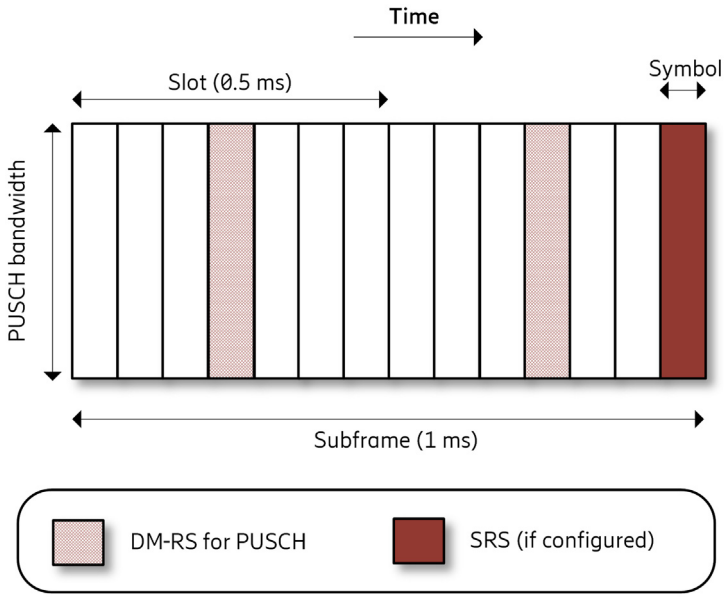

FIGURE 8.14

DM-RS and SRS positions in the subframe for the LTE uplink in release 8. In release 14, a comb-based mapping to subcarriers was introduced for DM-RS.

grant of the UL selects a CS and one out of the two available OCC codes, that is, a DM-RS port, for the PUSCH DM-RS.

The PUSCH DM-RS maps to the fourth and eleventh OFDM symbol in the subframe, and it maps to all subcarriers for the scheduled PUSCH (see Fig. 8.14).

In LTE release 14, the number of available DM-RS ports was increased further by introducing a comb-based DM-RS structure for PUSCH. This means that the RS of a DM-RS port is in this case mapped in an OFDM symbol to either odd or even subcarriers instead of every subcarrier. This theoretically doubles the number of PUSCH DM-RS ports, with the motivation of MU-MIMO with a larger number of simultaneously scheduled users in the UL. However, using a comb structure also increased the sensitivity to channel delay spread, like using CS, so the expected doubling of number of available ports is only valid for channels with very small delay spreads, for example, indoor deployments. To maintain the low PAPR property, the PUSCH and its associated comb-based DM-RS are not frequency multiplexed in the same OFDM symbol but are still time multiplexed as in Fig. 8.14.

8.3.1.7 Uplink sounding reference signals

A UL SRS transmission can be configured to assist in link adaptation for the UL (including determining a UL precoder) and for providing a UL channel estimate in reciprocity-based operation (see Section 6.4.2). The RS used for sounding the radio channel in the UL can be configured to be transmitted with a periodicity between 2 and 160 ms. The SRS is transmitted in the last symbol of the

subframe in case of FDD (see Fig. 8.14). UL sounding is also used in reciprocity-based schemes, to provide the channel for the DL.

In LTE release 10, aperiodic SRS was introduced in addition to the periodic SRS, where a UL measurement could be triggered by a DCI on a need basis. In addition, for special subframes in TDD which has both a DL and a UL part, more than one SRS symbol can be configured in the subframe. This is because for TDD configurations with mostly DL subframes, the possibilities for UL transmission are few and there is a need for more SRS transmission opportunities.

When using SRS, there is a tradeoff between coverage and sounding bandwidth (see also Section 6.4.2.1). Since the transmit power of a UE is limited, typically to 23 dBm, one way to improve SRS coverage is to configure the UE to transmit the SRS over a smaller bandwidth. Given a constant transmit power, the *power spectral density* (in W/Hz) of the SRS transmission is then increased and thus SRS coverage improves. The drawback is that only a part of the bandwidth is sounded and then multiple such SRS transmissions hopping across different parts of the band are needed. The smallest SRS transmission bandwidth is 4 RBs. Since each of these multiple narrow SRS transmissions are in different subframes, the channel may have changed, i.e. faded between each transmission. When the whole bandwidth has been sounded, some parts may be outdated while other parts are more recent and still a valid representation of the channel.

The closer the UE is to the eNB, the wider SRS bandwidth can be configured as the path loss in the channel is lower. Ideally, the UE transmits SRS across the whole bandwidth in a single transmission.

The release 8 SRS transmission is mapped to either odd or even subcarriers (i.e., one of two available *combs*) to increase the SRS capacity. The SRS use the same type of low PAPR sequence as PUSCH DM-RS and different SRS resources are thus separated by a CS and the comb identifier. Separating using a comb simplifies the receiver compared to separating using CS as the different combs are strictly orthogonal.

In release 13, the support for receiving SRS from an even larger number of UEs simultaneously was introduced. Theoretically, three times more SRS resources were obtained per OFDM symbol. This extends the support for UL MU-MIMO or extends use of reciprocity-based operation. The subcarrier mapping configuration to include also every fourth subcarrier and more CS was introduced. Note however that even if the SRS capacity increases by comb 4 and with additional CS, using these enhancements increases the sensitivity to large channel delay spreads. A sparser sampling in frequency domain (i.e. increasing comb value) and/or using a larger number of CS with smaller shift separation between each CS may lead to that cross-interference between different SRS resources occur. Hence these enhancements typically are targeting small cell or indoor deployments where small delay spreads are more commonly encountered. In NR, the same SRS structure of using comb 2 and comb 4 and CS is used as in LTE release 13 (refer to Fig. 9.9).

UEs often have more receive branches (receive antennas and associated baseband processing) than transmit branches since transmitters require power-consuming PA. See, for example Fig. 8.15, where the UE has a single PA for transmission but two antennas for reception.

To support reciprocity-based transmission in the DL, there is a need to sound the channel in the UL from all (receive) antennas at the UE and this becomes an issue in the typical case shown in Fig. 8.15 as the second antenna cannot transmit anything. In this case, the eNB is only able to get the partial UL channel by using the SRS and this degrades the performance of reciprocity-based operation and thus the DL PDSCH transmissions.

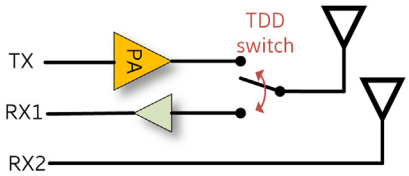

FIGURE 8.15

One transmit antenna and two receive antennas at the UE. The SRS can thus only be transmitted from one of the antennas and only partial channel sounding can be achieved. This UE thus supports 1T1R although it has two RX antennas.

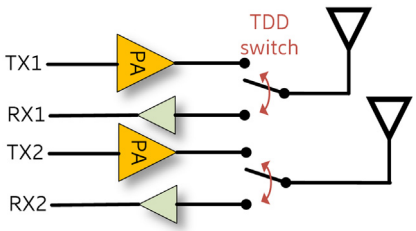

FIGURE 8.16

Full channel sounding with 2T2R, that is, both antennas can receive and transmit the SRS simultaneously.

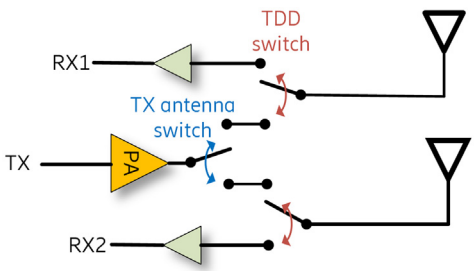

FIGURE 8.17

SRS antenna switching for 1T2R where the transmitter is dynamically switched to be connected to different antennas to allow full channel sounding.

It is thus desirable from reciprocity operation perspective (to improve DL performance) to support full channel sounding, where SRS can be transmitted from all antennas, as shown in Fig. 8.16.

In 3GPP, the used nomenclature for the case of Fig. 8.16 is 2T2R, which means the terminal has two transmit antenna and two receive antennas and thus supports full reciprocity. Another capability of a UE-supporting full reciprocity with four PA is 4T4R.

To support reciprocity-based operation also for terminals that do not have implemented the capability of full reciprocity channel sounding, antenna switching has been introduced for SRS. See Fig. 8.17 where the UE only has a single transmit PA. Here, UE capabilities that indicate that the UE support such switching are defined as 1T2R, 1T4R, or 2T4R. The scheme in Fig. 8.15 would imply that the UE needs to report 1T1R capability since it can only transmit SRS from one of the receive antennas as there is no antenna switching implemented.

The UE in the 1T2R case transmits SRS from the first antenna (corresponding to one of the two receive antennas) in a given subframe and then switches to the second antenna in a following subframe (assuming a single SRS symbol per subframe). Thereby the full UL channel can be sounded, although with at least 1-ms delay between the two measurements of the channels from the different antennas since this is the length of a subframe.

In TDD, the transmission of more than one SRS symbol in the same subframe is possible, and it is then possible to configure antenna switching within the same subframe, and in this case there is a guard (empty) symbol inserted between each SRS transmission to allow time for the UE to perform the antenna switching. The release 12 TDD supports 1 and 2 symbols used for SRS in TDD special subframes (known as UpPTS) and release 13 terminals support up to 6 SRS symbols in UpPTS.

In FDD, only a single symbol can be used for SRS per subframe although ongoing release 16 specification work supports UL subframes where all 14 OFDM symbols can be configured for SRS transmissions even from the same UE (a significant amount of SRS repetition in the same frequency band may thus be possible to configure for extended SRS coverage).

8.3.2 CSI REPORTING

8.3.2.1 Overview

The purpose of CSI feedback reporting from the UE to the eNB is to aid the scheduler in the DL link adaptation, that is, setting the modulation scheme and code rate for a scheduled PDSCH transmission. For the multi-antenna base station, the CSI can also include a *rank indicator* (RI), which indicates the preferred number of MIMO layers, and a preferred MIMO precoding matrix using a *precoding matrix indicator* (PMI) pointing to a matrix in a *codebook* of such matrices. In 3GPP, implicit CSI feedback is used (see a discussion of explicit vs implicit CSI feedback in Section 6.4.1).

The CSI reporting is computed under a hypothetical PDSCH transmission that can be wideband (the whole system bandwidth) or per sub-band. If sub-band CQI reporting is used, and the eNB decides to schedule more than one sub-band, then the eNB needs to perform additional calculations to combine these multiple per sub-band CQIs to obtain correct parameters, for example, code rate, for the link adaptation.

The bandwidth of a sub-band for CSI reporting depends on the used system bandwidth and ranges from 4 PRBs for bandwidths up to about 5 MHz up to eight PRBs for the largest bandwidths (where 20 MHz is the maximum bandwidth of an LTE carrier). This variable sub-band size keeps the CSI feedback overhead reasonably maintained even when the number of PRBs in the system is large.

Both periodic and aperiodic CSI reporting is supported, using PUCCH and PUSCH, respectively. For aperiodic CSI reporting over PUSCH, the report from the UE can be triggered by the

eNB sending PDCCH with a UL scheduling grant. The PUSCH containing the aperiodic CSI report is transmitted k subframes after the subframe of the triggering PDCCH. The distance $k = 4$ for FDD and for TDD, k is variable but at least 4. For TDD, the variable k is to ensure the indicated subframe is a subframe valid for UL transmissions.

The purpose of the aperiodic and periodic CSI feedback is different. The periodic feedback is carried by PUCCH and is intended to give eNB less detailed channel information; hence feedback is coarse, and payload is low. The periodic feedback is received by eNB even if there is no data to transmit to the UE and it is thus important to keep the overhead low, for example, long periodicity.

If data arrives to eNB for DL transmission to a UE, then PDSCH can be scheduled where link adaptation is based on the periodic CSI, while at the same time, an aperiodic CSI report can be triggered. The aperiodic CSI report provides the eNB with more detailed MIMO channel information, to maximize the DL throughput. Since the payload of the aperiodic CSI is large, it is triggered "on demand" only. The drawback of aperiodic reporting is of course the need for a trigger, that is, it occupies a PDCCH resource and a larger UL overhead that steals resources from UL data transmissions. In addition, since the aperiodic report is more detailed, the UE needs more time to compute it and there is a specified delay of at least four subframes between the trigger and the report is transmitted. Hence, it is only beneficial to trigger such report if the data packet to deliver to the UE is so large that it is expected to take more than four subframes. Otherwise the packet is completely delivered before the aperiodic CSI report has been received, using the periodic CSI for link adaptation.

For the aperiodic CSI reporting over PUSCH, several CSI reporting modes are defined (see Table 8.4), and the eNB configures the UE with one of these modes by higher-layer signaling. The PMI can be wideband (a single PMI) or per sub-band (multiple PMIs) where the latter has higher CSI feedback overhead but provides more detailed information to the eNB. There is also a possibility to configure modes with "No PMI," which is used for the single antenna eNB, transmit diversity, and open-loop precoding transmission schemes (e.g., TM 3), respectively.

Moreover, "No PMI" can also be configured for DM-RS-based transmission schemes, when used for reciprocity-based operation since the eNB computes the precoder based on SRS measurements in this case and the UE need not feedback PMI.

The channel quality information (CQI) is a 4-bit value that indicates the highest modulation and code rate for a received transport block that meets a block error rate target of at most 10% (as estimated by the UE). It can be reported either as a single wideband value or a wideband value plus per sub-band values where the sub-band option has two flavors, UE selected (where the UE also

Table 8.4 CQI and PMI Feedback Modes for CSI Reporting on PUSCH.

	No PMI	Single PMI	Multiple PMI
Wideband CQI	Mode 1-0	Mode 1-1	Mode 1-2
Sub-band CQI, UE selected sub-bands	Mode 2-0	X	Mode 2-2
Sub-band CQI, higher layer−configured sub-bands	Mode 3-0	Mode 3-1	Mode 3-2

All modes were introduced in release 8 except Mode 3-2, which was introduced in release 12.

Table 8.5 CQI and PMI Feedback Modes for CSI Reporting on PUCCH.

	No PMI	Single PMI
Wideband CQI	Mode 1-0	Mode 1-1
Sub-band CQI, UE selected sub-bands	Mode 2-0	Mode 2-1

reports which sub-bands have highest CQI) or higher-layer configured, which report CQI for all sub-bands.

PUSCH CSI reporting Mode 3-2, which is sub-band CQI plus sub-band PMI, was introduced in release 12 to better support MU-MIMO as this gives a somewhat more detailed CSI report. The decision to introduce this mode 3-2 was made after the 3GPP RAN1 SI on enhanced MIMO where it was concluded that more detailed CSI feedback gives DL performance benefits at the cost of more CSI signaling overhead.

Depending on the configured TM to the UE (see Section 8.3.4), the RS used for CSI measurements is different (either CRS or CSI-RS based) and the subset of CSI reporting modes (see Table 8.4) that is available for configuration depends also on the configured TM. For example, the TMs that are using codebook feedback (TM 4, TM 6, and TM 8−TM 10) allow configuration of CSI reporting modes 1-2, 2-2, 3-1, and 3-2.

Periodic CSI reporting over PUCCH is also supported, but due to the limited maximum payload of 11 bits in PUCCH in LTE release 8, simultaneous PMI reporting or CQI reporting for all sub-bands cannot be obtained and as mentioned above only coarse CSI is provided in the report. The PUCCH periodicity is configured by the network and typically used values are 20 or 40 ms.

To reduce the PMI payload, codebook subsampling is used for PUCCH reporting, where the UE cannot use the full codebook for the PMI reporting, but can only select from a subset of the precoding matrices in the codebook.

Table 8.5 shows the four different CSI reporting modes available for PUCCH reporting, where PMI can be included or not in the report, and CQI can be either wideband or per a preferred sub-band that the UE selects and indicates.

8.3.2.2 Codebooks in LTE release 8−release 12

Precoding for spatial multiplexing is introduced in Section 6.3. The purpose of a MIMO precoding codebook is to define *a set* of precoders, and one example is discussed Section 6.3.3.4. The codebook is thus standardized and increasing the size of the set (by proper codebook design) should ideally give better performance at the expense of increased feedback overhead since the precoding matrix indicator (PMI) is needed to select one precoder from the set.

Hence, in design of codebooks in standardization, the challenge is to find the best possible set of precoding matrices given the overhead constraints, that is, given the codebook size. A model-based approach can be taken, where a priori assumptions on port layouts (physical distance between phase center of each antenna port in one or two dimensions) are made, such as assuming dual-polarized antenna pairs and some typical channel models.

Typically, the receive signal correlation between RSs transmitted from closely spaced antennas is high, while correlation is low between signals transmitted from different polarizations. Such

knowledge is utilized in the codebook design and this model-based approach reduces the PMI feedback overhead.

A description of a precoder for a layer is represented by a $N_T \times 1$ vector of complex-valued elements with the same modulus, valid for either the full system bandwidth or per CSI reporting sub-band. When the reported rank is $r > 1$, a $N_T \times r$ matrix $\boldsymbol{W}$ instead of a vector represents the precoder, hence for r layers. Moreover, a finite number of such matrices $\boldsymbol{W}$ are defined per each rank r and number of ports N_T, and this superset of matrices $\boldsymbol{W}$ constitutes the entire LTE codebook.

If a CSI report is configured, the UE shall report a joint selection of preferred precoder(s) (PMI), rank indication (RI), and the largest CQI (per code word) that meets the 10% block error rate target.

In release 13, a *CSI-RS resource indicator* (CRI) was also added to the parameters in a CSI feedback report, which is present in the case a UE is configured to measure on more than one CSI-RS resource and select one of them for the CSI report. The motivation for introducing this is beamformed CSI-RS, for example, the eNB transmits four different CSI-RS resources in four different beam directions. The UE selects a preferred beam and feeds back the "normal" CSI report for a selected beam only (i.e., the CSI-RS resource) plus the CRI to indicate the selected resource. This is also known as Class B operation in LTE.

Sometimes, the network may want to restrict the precoders the UE can select from. Then there is a possibility for the eNB to restrict the UE to select only a subset of precoders within a codebook and/or a subset of the ranks. This is known as *codebook subset restriction* (CSR) and can be used if the network knows that it will never schedule a transmission in a certain direction (e.g., since this direction interferes heavily with users in an adjacent cell). Hence the precoders associated with this direction are restricted. Note that for simplicity, the use of CSR does not change the number of PMI and/or RI feedback bits, even if the number of "allowed" PMI and/or RI are significantly reduced.

In LTE, there are codebooks defined for $N_T = 2$ and 4 ports to be used with CRS-based measurements and $N_T = 2$, 4, 8, 12, 16, 20, 24, 28, and 32 ports with CSI-RS-based measurements. The antenna model used when designing the CSI-RS-based codebooks is uniformly spaced antenna ports in one or two dimensions, where the codebooks for 2D port layouts were introduced in release 13.

In the following, the basics of LTE codebooks will be presented, and general principles and key aspects will be discussed without digging into details. In the NR codebook description (Section 9.3.5), a more thorough walk through of the codebook design is given.

It is illustrative to first introduce codebooks with the LTE release 8 codebook for $N_T = 2$ CRS antenna ports, which for rank $r = 1$ are the following four vectors:

$$\boldsymbol{W} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\j\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-j\end{bmatrix}$$

Hence, the PMI, which in this case requires two bits, indicates the best possible cophasing of the two antenna ports using a QPSK alphabet. In cophasing transmission, the eNB transmits the same PDSCH layer from both antenna ports but where the transmission from the second antenna port applies the phase shift x as was recommended by the UE. Coherent combining of the two transmitted layers is then achieved at the receiver antenna in the UE.

For $N_T = 2$ CRS antenna ports and $r = 2$, these are the three specified precoding matrices in the LTE codebook

$$\boldsymbol{W} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix}$$

The first matrix simply indicates to transmit one layer per CRS antenna port, while the two others transmit each layer across both CRS antenna ports with a cophasing. Note that the columns of the precoding matrix are by design orthogonal and that each matrix element has the same amplitude (constant modulus requirement), which are general principles used for codebook design in 3GPP. In addition, the precoding matrices are power normalized, so each layer uses an equal portion of the total transmitter power.

For the release 8 codebook design for $N_T = 4$ CRS antenna ports, one underlying model assumption was dual-polarized antenna pairs spaced some distance apart. However, in case the distance between the two antennas is large, then the spatial correlation between the fading variations experienced between antennas of the same polarization in different pairs is low (see discussion in Section 3.7). In the release 8 codebook design, the aim was to design a codebook suitable for both high and low spatial correlation and it therefore contains a mix of precoding matrices where some are motivated by high-correlation and dual-polarized antenna setups, while others are not bound by this underlying model but more "random."

Ultimately, a compact way to describe such a codebook was by using a Householder transformation, since the whole codebook for rank 1−4 could simply be described by 16 different 4×1 vectors $\boldsymbol{u}_n, n = 0, \ldots, 15$ and for each $\boldsymbol{u}_n$ a 4×4 matrix is obtained as

$$\boldsymbol{W}_n = \boldsymbol{I} - \frac{2\boldsymbol{u}_n\boldsymbol{u}_n^*}{\boldsymbol{u}_n^*\boldsymbol{u}_n},$$

where $\boldsymbol{I}$ is the identity matrix. The codebooks for $r = 1$, 2, and 3 are then obtained by specifying which r columns to select from each $\boldsymbol{W}_n$. Note that the use of the Householder transformation implies a certain restriction on the possible precoders compared to individually optimizing each matrix without such constraint, but ease of specification and simple description was prioritized. Alternative and improved four-port codebooks were specified in release 12 as will be discussed further below.

In LTE release 10, the 3GPP specifications for multi-antenna transmissions were expanded to target up to eight-layer MIMO transmission. A new codebook structure was then introduced for eight CSI-RS ports and the codebook design principle was different from release 8, since for eight CSI-RS ports, it was seen as likely that an array antenna with four linearly spaced dual-polarized antenna pair elements would be used in real deployments and hence high correlation between spatially separated antenna ports and low correlation between polarization separated antenna ports (with no spatial separation) was exploited in the eight-port codebook design. This new structure has then been used since then in both LTE evolution and in NR.

The precoder $\boldsymbol{W}$ structure for this codebook is referred to as the factorized precoder structure as discussed in more depth and motivated in Section 6.4.1.2 and is defined as a matrix multiplication of an inner and outer precoding matrix:

$$\boldsymbol{W} = \boldsymbol{W}_1\boldsymbol{W}_2$$

where the first, inner, $N_T \times \tilde{N}_T$ precoder matrix $\boldsymbol{W}_1$ is block diagonal with a matrix $\boldsymbol{B}$ repeated twice in the matrix diagonal; hence

$$\boldsymbol{W}_1 = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}$$

and where $\tilde{N}_T$ is a design parameter which is related to the number of remaining beams after the channel dimension reduction by $\boldsymbol{W}_1$.

The matrix $\boldsymbol{B}$ is an $N_T/2 \times N_b$ matrix that contains N_b DFT vectors and the purpose of $\boldsymbol{B}$ is hence to match the spatial channel from a uniformly and linearly spaced array of equally polarized antenna elements. The use of two blocks is thus to match the channel from each of the two different polarizations respectively. The channel from closely spaced antenna ports of the same polarization are likely to be well approximated with a DFT vector at least for modest angle spread, and thus it is expected that $\boldsymbol{W}_1$ is similar over the system bandwidth and over time. Hence, $\boldsymbol{W}_1$ can be wideband and updated rather infrequently as discussed more in Section 6.4.1.2. One could also see the different DFT vectors of the columns of $\boldsymbol{B}$ as different beam directions, an analogy that holds perfectly well under the "ideal" conditions of line of sight channel and half wavelength−spaced antenna ports per polarization.

The outer precoder $\boldsymbol{W}_2$ has then the purpose of performing beam selection and cophasing of the beams from the two polarizations, that is, combining the two diagonal $\boldsymbol{B}$ matrices. As correlation between different polarizations typically is small, the outer precoder needs to be selected per subband to follow the channel variations across frequency.

Another interpretation (see Fig. 6.23 in Section 6.4.1.2) is that $\boldsymbol{W}_1$ serves to create a new effective (dimension reduced) $N_R \times \tilde{N}_T$ channel matrix as measured by the UE, $\boldsymbol{H}_{eff} = \boldsymbol{H}\boldsymbol{W}_1$ for the outer $\tilde{N}_T \times r$ precoder $\boldsymbol{W}_2$ to be applied to. The inner precoder thus introduced virtual antennas (one port per beam polarization if that analogy can be used). Since the number of virtual antennas $\tilde{N}_T$ is much smaller than the number of antenna ports N_T, this provides a considerable dimension reduction and thus a smaller codebook (fewer elements) for the outer precoder. As the outer precoder typically is reported per sub-band, the smaller outer precoder codebook helps to keep the total PMI overhead down.

For rank 1 and rank 2 codebooks, a grid of beams of 32 DFT vectors are used and since the release 10 LTE codebook is designed for a 1D antenna port layout with $N_T = 8$ antenna ports, there are $N_T/2 = 4$ spatially separated ports in the linear array (per polarization) implying four spatially orthogonal DFT vectors. In the codebook design, if only these four orthogonal DFT vectors would have been chosen to represent the different precoders, then there would have been a large signal power loss if the UE is in a direction between these orthogonal DFT beam directions. To mitigate this power loss, spatial oversampling has been introduced, where additional beams (DFT vectors) are placed in between orthogonal beams. Since the rank 1 and rank 2 codebooks for four linear and spatially separated ports (of same polarization) have 32 DFT vectors, the oversampling factor is $O = 32/4 = 8$, for the eight-port codebook defined in LTE.

Another benefit of such a large degree of spatial oversampling is that it makes the codebook less sensitive to the physical distance between the four antenna ports compared to if, for example, only the four orthogonal beams would have been used. Even if a physical spacing larger than the half wavelength is implemented in the AAS, the codebook performance is still reasonably good thanks to this spatial oversampling property.

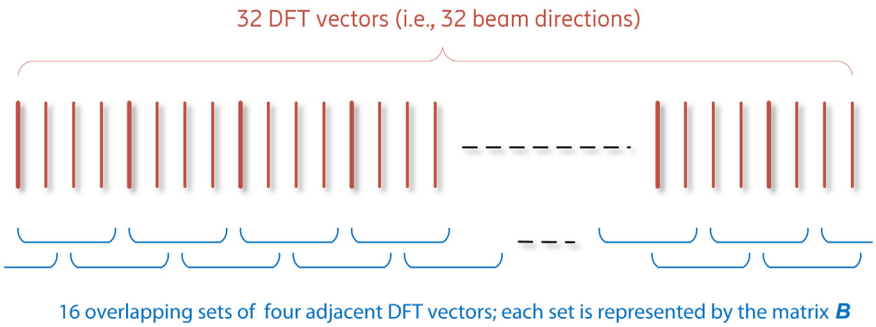

FIGURE 8.18

LTE codebook design for rank 1 and 2 where 32 DFT vectors are specified and the UE selects a subset of four adjacent DFT vectors for wideband feedback. The UE further selects one of the four, per sub-band.

From these 32 vectors, 16 different $\boldsymbol{B}$ matrices are defined in specifications where each contains $N_b = 4$ columns hence a sheaf of four DFT beam directions. The 16 matrices consist of subsets of the 32 DFT vectors as illustrated in Fig. 8.18.

The wideband PMI feedback, PMI_1, thus selects the inner precoder $\boldsymbol{W}_1$, that is, selects one of the 16 different $\boldsymbol{B}$ matrices and thus a sheaf of adjacent DFT beams, and the outer precoder $\boldsymbol{W}_2$ then performs selection of one column (i.e one beam) from each $\boldsymbol{B}$ so that effectively a dimension reduction from $N_T = 8$ to $\tilde{N}_T = 2$ is achieved, which simplifies the complexity of the following per sub-band cophasing.

The release 8 codebook for two ports is then re-used to perform the polarization cophasing. Hence, the outer precoder $\boldsymbol{W}_2$ performs both beam selection and cophasing. In general terms, the codebook for the outer precoder $\boldsymbol{W}_2$ can be written as

$$\boldsymbol{W}_2 = \begin{bmatrix} \boldsymbol{e}_{k_{\text{a},1}} & \boldsymbol{e}_{k_{\text{a},2}} & \cdots & \boldsymbol{e}_{k_{\text{a},r}} \\ e^{j\phi_1}\boldsymbol{e}_{k_{\text{b},1}} & e^{j\phi_2}\boldsymbol{e}_{k_{\text{b},2}} & \cdots & e^{j\phi_r}\boldsymbol{e}_{k_{\text{b},r}} \end{bmatrix}$$

where $\left\{\boldsymbol{e}_{k_{\text{a},l}}\right\}_{l=1}^{r}$ are 4×1 selection vectors with a single vector element $k_{\text{a},l}$ equal to one and the remaining elements equal to zero. The factors $e^{j\phi_i}$ are the polarization cophasing factors, per layer.

The rank 3 and 4 codebooks for eight ports are similar in principle although the oversampling factor is reduced, $O = 4$, and $N_b = 8$ DFT vectors are used in the matrix $\boldsymbol{B}$ since higher rank transmission requires selection of more spatially separated vectors. In addition, only four different $\boldsymbol{B}$ are defined.

For rank 5−8, there is no oversampling in matrix $\boldsymbol{B}$ design and there is no UE selection of cophasing or beam selection by $\boldsymbol{W}_2$ hence a fixed $\boldsymbol{W}_2$ is specified per rank. One out of four $\boldsymbol{W}_1$ matrices can be selected by the UE for rank 5−7 and for rank 8, the codebook consists of a single matrix $\boldsymbol{W}$. The reason is that the benefit of closed-loop precoding where UE selects beams and cophasing diminishes when the number of layers is above four for a system with eight transmit antennas and eight receive antennas. Hence, the PMI feedback overhead is wideband only and thus small for rank 5−8 feedback.

In release 12, an alternative four-port codebook was specified for the DM-RS-based transmission schemes. It uses the $\boldsymbol{W} = \boldsymbol{W}_1\boldsymbol{W}_2$ structure that can replace the Householder based codebook

design from release 8. This codebook was specifically designed for two closely spaced cross-pole antennas, as well as for rank 1 transmission of four linearly spaced antennas with the same polarization, and thus has some advantage over the release 8 codebook for such deployments. It also reduces the CSI feedback overhead for large scheduling bandwidths since the release 8 codebook requires 4 bits per sub-band, while the new release 12 codebook has one wideband component $\boldsymbol{W}_1$ plus a low overhead cophasing component $\boldsymbol{W}_2$ per sub-band. A design difference compared to the eight-port codebook in release 10 is that this four-port codebook uses orthogonal, that is, nonadjacent beams in the block matrix $\boldsymbol{B}$. The motivation was that the beam width per DFT beam is doubled compared to the eight-port case and thus the use of adjacent beams in $\boldsymbol{B}$ is less motivated as they are rather similar. For rank 3 and 4, there was no change and the release 8 codebook is used but incorporated in the new codebook framework by using release 8 Householder codebook for $\boldsymbol{W}_2$ (allowing per sub-band reporting) and setting the identity matrix for $\boldsymbol{W}_1$.

8.3.2.3 Introducing two classes of channel state information feedback in long-term evolution release 13

In release 13, the LTE CSI framework was extended to 2D port layouts, and two classes of CSI feedback were defined, A and B, where Class A corresponded to the existing framework and Class B was a newly introduced framework that targeted beamformed CSI-RS.

The reason was that 3GPP concluded an SI on elevation beamforming and full-dimensional beamforming that followed the release 12 completion of a 3D channel model. The SI also included modeling of deployments with 2D antenna arrays. Prior to this SI, the used channel models assumed that all UEs are 1.5 m above ground, hence only distributed in horizontal as seen from the base station. The new 3D channel model allowed for UEs to be placed in buildings, indoor on different floors, which more realistically models an urban deployment.

MIMO in LTE had prior to release 13 been limited to adaptability of transmission in either the vertical or the horizontal domain and codebooks were designed for ID port layouts (uniform linear array of cross-pole antennas) that allowed beam steerability in one dimension only. Prior to release 13, there was no support for using horizontal and vertical beam steering simultaneously.

The release 13 SI report [10] concluded that significant throughput gains were achievable in realistic non-full-buffer evaluation with the joint vertical and horizontal beamforming, over the baseline, which was release 12−based MIMO features.

The concept of beamformed CSI-RS was introduced, motivated by the SI conclusion. One motivation was coverage, and another was that an envisioned product would create multiple vertical beams and in each of these beams there is a 1D horizontal port layout. The UE would then select a preferred vertical beam and use the "normal" CSI feedback assuming the codebook designed for a 1D port layout, as had been specified in release 10, within the selected beam. These principles laid the foundation for Class B type of operation.

Class A, also known as nonprecoded CSI-RS, is based on the principle of the same number of CSI-RS antenna ports as the number of antennas (although an "antenna" in this context may consist of multiple antenna elements in a subarray) (see Section 4.8 or the input to the antenna virtualization, according to below). The principle of Class A is thus that each CSI-RS port covers the whole serving cell (in contrast to Class B, see below).

The UE then computes the preferred precoder for a given rank, assuming transmission from these ports. Class A thus resembles what had been assumed for the codebook-based feedback in all

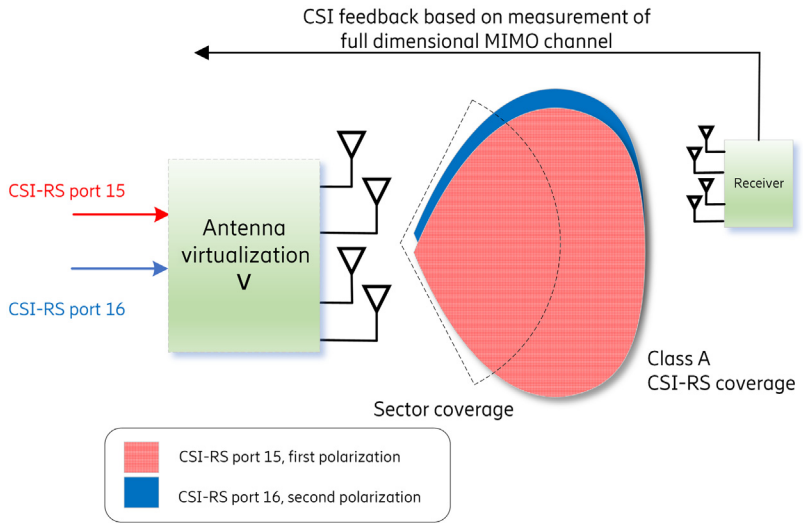

FIGURE 8.19

The Class A CSI feedback defined in release 13, also known as nonprecoded CSI-RS, in which each CSI-RS port has full sector coverage. The UE feeds back a precoder based on a full-dimensional channel since the UE can measure the full MIMO channel from all virtual antenna elements used by CSI-RS. The antenna virtualization is in this case static and same for all UEs served in the sector.

the previous LTE releases. In release 13, the introduced Class A operation also supports 2D CSI-RS antenna port layouts, and hence 3GPP introduced a corresponding codebook of 2D precoders (Fig. 8.19).

Class B is the beamformed CSI-RS where a transmitted CSI-RS does not need to have full sector coverage in vertical and/or horizontal. Hence, the eNB can be seen to perform a spatial prefiltering by the antenna virtualization, and the UE computes the preferred precoder on this spatially filtered channel. This operation requires either some knowledge at the eNB about a suitable spatial filter/beam (e.g., using reciprocity), or by providing multiple, different, beamformed CSI-RS resources (different virtualizations $\mathbf{V}$) and let the UE select and feedback a preferred resource that is, beam (Fig. 8.20).

The Class A configuration on the other hand always uses a single CSI-RS resource in release 13 of 8, 12, or 16 ports and the codebook uses the $\boldsymbol{W}_1\boldsymbol{W}_2$ precoder structure as introduced in Section 8.3.2.2.

To support this new functionality, codebooks were specified in release 13 for closely spaced cross-poles in various 1D as well as 2D port layouts from rank 1 up to rank 8. The 2D port layout defines N_1 and N_2 antenna ports per first and second dimensions, respectively, where it is up to each deployment whether to use N_1 ports in horizontal or vertical dimensions and vice versa for N_2.

In a similar way as the 1D codebook introduced in release 10, the 2D codebook uses DFT beams with spatial oversampling but with beams pointing in both a vertical and a horizontal angle. The spatial oversampling factor O_1 and O_2 was defined as configuration parameters for the first

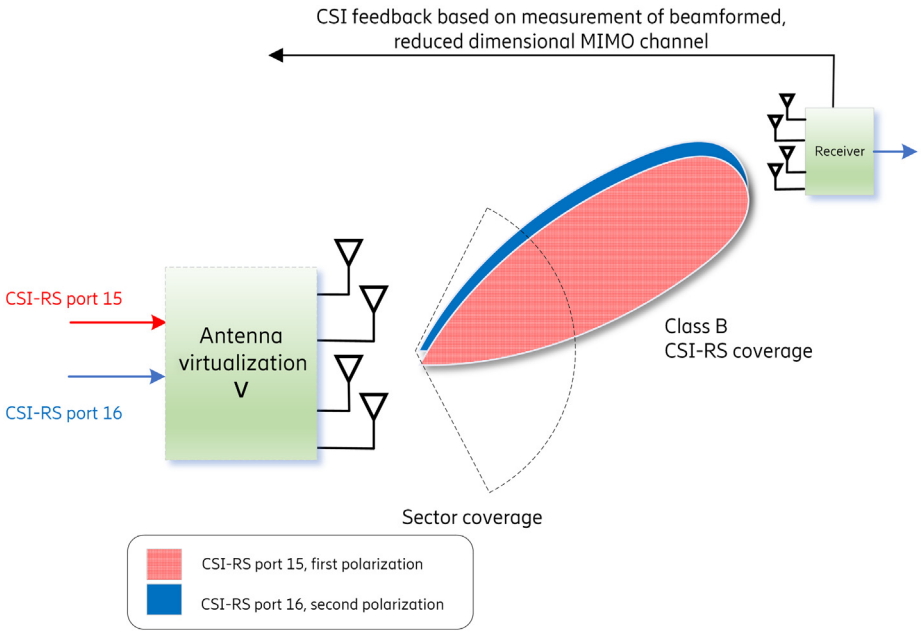

FIGURE 8.20

The Class B CSI feedback introduced in release 13, also known as beamformed CSI-RS, in which each CSI-RS resource is spatially filtered (e.g. beamformed) and has only partial sector coverage. The UE feeds back a precoder based on a low dimensional channel since the eNB has already reduced the number of dimensions by the spatial prefiltering. The antenna virtualization is in this case dynamic and can even be UE−specific.

and second dimensions, respectively. Note that the specification is agnostic to whether N_1 and N_2 correspond to vertical and horizontal dimensions of the antenna arrays, or vice versa, as this is up to the deployment and thus the physical rotation of the antenna array at installation.

The total number of beams for a given 2D codebook configuration in the two dimensions is $N_1O_1N_2O_2$ and where each of these four parameters are configurable. More beams imply more feedback overhead but potentially better performance.

The more detailed 2D codebook structure will be elaborated for the NR codebook in Section 9.3.5.3 since it is based on the same principles as LTE, using a grid of DFT beams in two dimensions, that is, vertical and horizontal.

For the LTE 2D codebook there is a great deal of configurability in the number of beams in each dimension (due to spatial oversampling), and different port layouts. Table 8.6 shows the supported configurations for the release 13 Class A codebooks where for instance (2,3) is interpreted as a 2D antenna array with 2 and 3 ports per polarization, respectively (thus in total $2 \times 3 \times 2 = 12$ ports), and where spatial oversampling can be configured to be either eight times in first dimension and four times in second dimension or eight times in both dimensions.

As discussed above, Class B codebooks are designed for beamformed CSI-RS, where a UE can be configured with multiple, $K \geq 1$ CSI-RS resources. In Class B operation, $\boldsymbol{W}_1$ is set to the identity

Table 8.6 Supported Codebook Configurations for LTE Release 13 Two-Dimensional Port Layouts.

(N_1,N_2)	(O_1,O_2)
(8,1)	(4,-), (8,-)
(2,2)	(4,4), (8,8)
(2,3)	(8,4), (8,8)
(3,2)	(8,4), (4,4)
(2,4)	(8,4), (8,8)
(4,2)	(8,4), (4,4)

N_1 and N_2 are the number of ports in first and second dimensions, respectively (assuming same polarization), and O_1 and O_2 are the spatial oversampling factors per dimension. The two dimensions represents horisontal and vertical, respectively, or vice versa.

matrix in the feedback. This is rather natural since eNB precodes the CSI-RS to mimic $\boldsymbol{W}_1$ in a transparent manner in Class B, as opposed to the codebook-based UE selection of $\boldsymbol{W}_1$ when using the Class A codebook.

In addition, two different configurations are possible for Class B operation, $K = 1$ resource or $K > 1$ resources, where each of resource $k = 0,1,\ldots,K-1$ is configured with N_k ports where $N_k = \{1,2,4,8\}$.

The Class B, $K > 1$ codebook uses the release 10 (i.e., 1D port layout) codebook for the selected CSI-RS resource. Up to $K = 8$ CSI-RS can be configured, and a 1D port layout is used per CSI-RS resource. For this operation with $K > 1$, the eNB beamforms each CSI-RS resource in a beam (e.g., vertical beam) and the UE selects a preferred beam on a wideband basis, that is, one of the K resources and feeds back its selection (by CRI, the CSI-RS resource indicator) together with CSI for the selected resource, including a PMI for the N_k ports (which then may represent the horizontal precoding).

Alternatively, if a single resource $K = 1$ is configured with Class B operation, the intended use case targets UE−specific beamforming, potentially where channel reciprocity is utilized by the eNB to acquire a set of useful beam directions for the UE. Here, $\boldsymbol{W}_1$ equals the identity matrix (and thus not included in CSI feedback) while $\boldsymbol{W}_2$ is used to performing beam selection (out of a set of four dual-polarized beams) and cophasing of polarizations within the selected beam. For this cophasing reporting purpose, it is possible to use the release 10 codebooks, or a new release 13 codebook tailored for this feedback configuration is used.

Together with the introduction of beamformed CSI-RS in release 13, *measurement restriction* was introduced as a new feature. One use case with a beamformed CSI-RS is that a periodic CSI-RS resource is configured and the eNB then changes the beam direction at each instant of that CSI-RS transmission, so that multiple UEs at different directions can measure on the same configured CSI-RS resource. This reuse scheme saves overhead but also introduced a problem since the UE may apply filtering across multiple instances of the CSI-RS. If the eNB changes the beam at each instant, the averaging of the channel across these subframes does not make sense. This problem led to the introduction of measurement restriction configuration to the UE, where if configured, the UE shall only use a single instance of a periodic CSI-RS when requested to feed back a CSI report. By

configuring measurement restriction, the problem of UE averaging over time is removed and the eNB knows which beam the UE has measured for each report. In release 13, measurement restriction for channel and/or interference measurements is supported.

In the following release, release 14, the work continued toward supporting massive MIMO, and Class A codebooks were further extended with 20, 24, 28, and 32 CSI-RS ports for both 1D and 2D port layouts.

8.3.2.4 Advanced channel state information feedback

Channel reciprocity-based operation is beneficial, and as seen in the performance evaluations in Section 13.6, and mainly beneficial for MU-MIMO scheduling. But it relies on many critical circumstances for it to work (see the discussion in Section 6.7). As an alternative to reciprocity based operation, one could rely on richer channel feedback using the CSI framework, in order to reach MU-MIMO performance similar to reciprocity-based operation. This feature thus enables very good MU-MIMO performance independently of whether FDD or TDD is deployed.

Hence, in release 14 the *advanced* CSI feedback reporting was introduced for 4−32 CSI-RS ports, with the aim to provide more detailed, high-resolution CSI feedback to the network. Only rank 1 and 2 feedback is supported in this advanced mode. The goal is to achieve comparable channel knowledge for an FDD system using CSI feedback as reciprocity-based SRS measurements can provide for TDD. Note that the advanced feedback is then also useful for TDD systems where channel reciprocity is not available, for example, due to the partial channel reciprocity problem discussed in Section 6.7.1.2. Using advanced feedback is more costly in feedback overhead and targets MU-MIMO performance gains although it should be noted in this context that reciprocity-based operation also consumes a lot of UL overhead since each UE needs to transmit SRS frequently.

Feeding back a single DFT beam direction as is used in feedback design of the previous codebooks for rank 1 and 2 is in many cases insufficient for MU-MIMO. The reason is that for MU-MIMO precoding, not only the strongest DFT beam direction of the channel matters, but also other directions since these other directions are relevant for interference toward co-scheduled UEs.

Hence, it was observed that a "richer" multi-antenna channel feedback could provide a rather large performance increase for MU-MIMO. This resulted in advanced CSI feedback using a 2D codebook.

The design principle of this codebook is illustrated in Fig. 8.21.

Two DFT beams ($\begin{bmatrix} \boldsymbol{b}_0 & \boldsymbol{b}_1 \end{bmatrix}$) are in a first step chosen from a group of beams (constituting an orthogonal 2D DFT basis of the channel space) to represent the two strongest directions of the spatial channel. The power of the second beam is scaled by $\sqrt{p_1}$ to represent the different propagation conditions, that is, path loss between the two beams.

These two steps will generate the outer, wideband matrix $\boldsymbol{W}_1$ followed by a per sub-band cophasing using $\boldsymbol{W}_2$. This co-phasing using $e^{j\alpha_1}$ of the second beam will also be carried out over the polarizations even if that is not illustrated in the figure.

For dual-layer transmission, that is, rank 2, individual beam co-phasing is used for the two layers. Note that the "normal" CSI feedback is then a special case of the advanced feedback where only a single beam is selected in the first step.

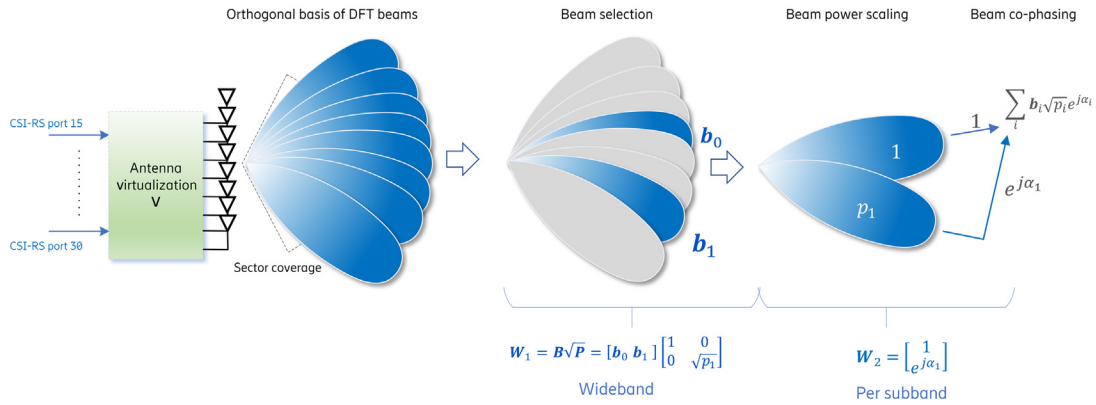

FIGURE 8.21

Codebook for advanced CSI feedback. The illustration only shows the cophasing of beams for one of the two polarizations (for simplicity).

8.3.3 PHYSICAL DOWNLINK CONTROL CHANNEL TRANSMISSION PROCEDURES

There are multiple TMs in LTE and the UE is simultaneously configured with one DL and one UL TM. The modes will be introduced in the next section. Prior to that, the control signaling is briefly introduced as it is important to understand how TM switching is performed without losing the connection to the UE and to briefly discuss how scheduling is performed.

DL control signaling in LTE and NR used to schedule UL and DL transmissions uses *downlink control information* (DCI), which is mapped onto the *physical downlink control channel* (PDCCH) in the control region of the subframe (see Fig. 8.2). For PDCCH, link adaptation is used where a PDCCH can consist of 1, 2, 4, or 8 *control channel elements* (CCEs). By selecting the number of CCEs (referred to as the aggregation level) for a PDCCH transmission the coverage can be controlled.

In each subframe, the UE shall decode a set of potential PDCCH transmissions, known as PDCCH candidates, for each aggregation level. The set of PDCCH candidates and their positions on the subframe is known as the PDCCH search space. Since there are multiple candidates, the network can choose which one to use for PDCCH transmission to a specific UE, and thereby avoid collision with another PDCCH transmitted simultaneously to another UE that is scheduled in the same subframe.

Note that the UE may also receive more than one PDCCH in the same subframe to schedule, for example, PDSCH and PUSCH simultaneously. Hence, in addition to blindly decoding a number of PDCCH candidates for each aggregation level, the UE shall also for each PDCCH candidate decode twice for two different information payloads, that is, a UL DCI payload and a DL DCI payload. These two DCIs are potentially scheduling PUSCH and PDSCH, respectively.

For DL scheduling, there is both a small DCI payload format and a large DCI payload format. The larger DCI payload has a variable size depending on the used TM, depending on which RRC configured features are enabled, such as carrier aggregation and SRS triggering. The smaller DCI payload is always the same across all LTE releases and irrespectively of any enabled features. The small payload has restrictions as spatial multiplexing cannot be scheduled from this smaller DCI format. Hence, it is always possible to reach the UE with a scheduling command with the smaller DCI payload, which is useful when the UE is reconfigured to a different TM as this reconfiguration takes a few milliseconds and it is ambiguous exactly when the UE has changed to the new TM. The smaller DCI is therefore often called the fallback DCI.

The fallback DCI for DL scheduling has the same payload size as the UL DCI for PUSCH scheduling, except when UL MIMO is configured. Hence, the UE can, after detecting the fallback DCI, further detect whether it is a UL or a DL DCI, without the need to blindly decode yet another DCI payload size.

When UL MIMO as introduced in release 10 is enabled for the UE, then a third payload size is used. UEs with UL MIMO enabled must therefore decode three different payload sizes, for DL, UL MIMO, and fallback, respectively.

Search spaces come in two types, the *common search space* (CSS) and the *UE−specific search space* (UESS). The CSS is used for broadcast messages, such as system information, paging, and random access response messages. Also, the fallback DCI is monitored by the UE in the CSS only. The large payload DCI is monitored in the UESS.

In release 11, the EPDCCH was introduced, allowing for beamforming of the control channel to the UE, but EPDCCH only supports the UESS; hence it can be configured as an alternative. The CSS must be broadcasted covering the whole cell and cannot benefit from UE−specific channel aware transmissions. For some of the messages scheduled from CSS, the preferred "beam" toward the UE is unknown (as in paging) or the message is a multicast message (such as system information).

8.3.4 PDCCH TRANSMISSION PROCEDURES AND DOWNLINK TRANSMISSION MODES

In LTE, there are 10 different *transmission modes* (TMs) and the first LTE release, release 8, contained the first seven, while the other three were added during LTE evolution as new features were introduced.

This section will describe these TMs. The reason for many modes in the initial release is mainly due to the use of CRS-based demodulation, which requires one mode for each transmission scheme as the actual mapping of bits to antennas must be clearly specified (e.g., single antenna, spatial diversity, spatial multiplexing, etc.). In hindsight, it may have been done differently and with fewer modes by allowing additional configurations on top of a TM, and this philosophy was also adopted in later LTE releases beyond release 11 where no new modes have been introduced.

In the first release of LTE, CRS was the all-purpose RS, used for fine synchronization, channel analysis, measurements, and demodulation. Hence, TM 1−6 all use CRS for demodulation, and to support different transmission schemes one mode was needed for each scheme. These are the release 8 TMs:

- **TM 1** and **TM 2** are single-layer schemes;
  - TM1 is configured when eNB only has a single CRS port, that is, a single transmit antenna;
  - TM2 is single-layer transmit diversity across two or four CRS ports;
- **TM 3** and **TM 4** are spatial multiplexing schemes used to increase spectral efficiency and are based on open-loop and closed-loop MIMO transmission, respectively;
  - TM 3 uses a predefined cycling of a precoding matrix selection from a codebook of matrices for a given rank, where cycling implies that one precoder from the codebook is selected pseudo-randomly for each RE and these precoders are cycled through within the PDSCH transmission to get spatial diversity for the PDSCH. The cycling pattern is given by the specifications and hence known to both eNB and UE. Supported transmission rank is 1, 2, 3, and 4;
  - TM 4 uses a single eNB selected precoder (wideband or per PRG) from a codebook of predefined matrices in the specifications. Hence, TM 4 is a codebook-based TM and supported transmission rank is 1, 2, 3, and 4;
- **TM 5** and **TM 6** are special configurations of TM 4 where only a single layer can be scheduled; hence transmission rank is restricted to 1. TM 5 has the possibility to signal in DCI a PDSCH power reduction of 3 dB. This is used for simultaneous transmission to another UE since then the power per UE is halved. TM 5 is therefore intended for MU-MIMO scheduling of two UEs with one layer each;

- **TM 7** is a single-layer transmission using DM-RS instead of CRS for demodulation. It can be used for UE−specific beamforming and reciprocity operation as the used precoder is transparent to the UE and not bound to be taken from a codebook. Hence, this is noncodebook-based operation since the codebook used for PDSCH transmission is not specified in the standards.

8.3.4.1 Release 9: Time division duplex enhancements and TM 8

In release 9, the scope of the enhancement was to improve the spectral efficiency for reciprocity-based operation, hence mainly targeting TDD. It was observed that most eNB have dual-polarized antennas and UEs have two RX antennas, yet only single-layer transmission using UE−specific RSs for demodulation was supported by release 8 (by TM 7).

Hence, dual-layer transmission was introduced in this release with new UE−specific RSs for demodulation, associated with a new TM, TM 8. As in TM 7, this is also a noncodebook-based TM and the precoder $\boldsymbol{W}$ is noncodebook based and hence derived by the eNB from UL measurements. During specification it was foreseen that even more layers with DM-RS would be introduced in release 10, so the release 9 DM-RS design for TM 8 was constructed to be forward compatible, that is, could be extended with more ports in a future release.

Note that no CSI feedback enhancements were made in this release, so even if reciprocity-based operation is used with TM 8, the CSI feedback is based on CRS measurements. Hence, if UE−specific beamforming is used for PDSCH, then the eNB needs to perform further calculations based on the CSI report to obtain a new CSI that better matches the link quality of a beamformed PDSCH. This problem was addressed in the next release by the introduction of dedicated measurement signals, the CSI-RS.

8.3.4.2 Release 10: Spectral efficiency enhancements and TM 9

Release 10, which was the LTE-Advanced release, took a big step in spectral efficiency by extending from four- to eight-layer transmission to the same UE. The UE−specific RS paradigm was used, and TM 9 was introduced, which, as TM 7 and TM 8, is a noncodebook-based TM. Also, CSI-RS was introduced to allow more flexibility in obtaining a measurement, specific for a certain UE (i.e., using channel-dependent beamforming of CSI-RS).

Due to these enhancements, the IMT-Advanced peak spectral efficiency target of 30 bps/Hz could be met. Furthermore, MU-MIMO operation was improved since the DM-RS is precoded in the same way as the associated data layer and MU-MIMO can therefore be supported with one layer per UE using orthogonal DM-RS or two layers per UE using different pseudo-orthogonal DM-RS sequences (see Section 8.3.1.5).

The DM-RS always have the same power as the PDSCH layer; hence there is no need to signal the power offset between DM-RS and PDSCH as was needed for the CRS-based MU-MIMO mode of TM 5. Precoding matrix codebooks were designed for CSI feedback with eight TX antennas (i.e., eight CSI-RS ports) at the eNB as was discussed in Section 8.3.2.2.

8.3.4.3 Release 11: Dynamic point selection and TM 10

In release 11, the scope of MIMO enhancements was targeting multipoint transmissions and improved CSI feedback. A definition of how the UE measures interference was specified and support for *dynamic point selection* (DPS), where the eNB that transmits data to the UE can

dynamically change from one PDSCH transmission to the next, was introduced. Yet another enhancement to the noncodebook-based transmission was specified by a new mode, TM 10. It basically has all the features of TM 9, using the same codebooks for CSI feedback but additionally defined *CSI interference measurement resources* (CSI-IM) to precisely specify the REs the UE uses for interference estimation. Refer to the problem of CRS-based interference measurements in Section 8.3.1.3.

The CSI-IM gives network information on how the UE estimates SINR and thus CQI reporting. Up until TM 9, the interference measurement procedure for the UE was not specified. The introduction of CSI-IM was important as it improves network performance with more advanced deployments.

The concept of CSI processes was also introduced in release 11, where up to four CSI processes can be configured to the UE so that the network can get CSI feedback regarding transmission not only from the serving eNB but also for DPS transmissions from other eNB. Each process thus reflects CSI for a transmission from one eNB.

To fully support the DPS use case, the QCL framework was introduced, since the UE needs to perform synchronization and channel analysis for possible upcoming transmissions from not only one but any one in a set of multiple different eNB (see Section 8.3.1.2 for introduction of QCL).

8.3.4.4 Release 12: Further MU-MIMO enhancements

In release 12, a new precoding matrix codebook was introduced for rank 1 and 2 in 4 TX operation in TM 8, 9, and 10, that is, the TMs that use DM-RS for demodulation. At this time, 4TX eNB were gaining popularity in the market and there was a need to further optimize the performance by a new codebook that is better matched to actual products in the field. The new codebook has the same $\boldsymbol{W}_1\boldsymbol{W}_2$ structure assuming grid of beams as the 8 TX codebook that was introduced in release 10.

In addition, a new CSI feedback mode (denoted mode 3-2) was introduced for TM 4, 6, 8, 9, and 10, which provides precoding matrix index (PMI) selection as well as CQI per sub-band across the system bandwidth. This higher resolution feedback provided some benefit in performance, especially for MU-MIMO scheduling.

In release 12, a 3GPP SI completed a 3D channel model that allowed to continue exploiting multi-antenna transmission benefits in vertical domain. The previous channel model assumed that all UEs are on the "ground plane", while in reality they are distributed vertically, for example, on different floors in a building facing the eNB. The new 3GPP 3D UMi and UMa models distributed UEs between 1.5 and 22.5 meters height. Moreover, support for modeling of two-dimensional antenna arrays was added.

8.3.4.5 Release 13: Full-dimensional MIMO enhancements

In release 13, the term *full-dimensional MIMO* (FD-MIMO) was introduced, with the purpose of exploiting the vertical domain as well as the horizontal domain in the precoding and focusing of radiated energy. Hence, release 13 is the first release that aims at exploiting the benefits of massively large antenna arrays with many steerable antennas. TM 10 supports all features introduced for FD-MIMO, while TM 9 supports a subset of these new features. Hence, new CSI feedback precoding matrix codebooks were defined for 8, 12, and 16 TX antennas where both 1D and 2D array

layouts were supported (assuming a model of equally spaced antennas). Hence, 2D precoding codebooks were defined as well.

There was also in release 13 scope an aim to reduce the UE complexity in measuring so many antenna ports for CSI measurement, by introducing precoded or beamformed CSI-RS, which also improves the coverage for each port. The principle is that the eNB may roughly know the preferred channel direction to the UE seen from the eNB (e.g., the vertical angle) and can then transmit one or multiple beams with CSI-RS to the UE (e.g., in different horizontal angles), and then the UE selects a preferred beam before reporting CSI for that selected beam. The number of CSI-RS ports per beam can then be kept small and since the eNB does part of the precoder down selection (e.g., by selecting vertical beam), the complexity is somewhat reduced for the UE while somewhat increased for the eNB. To support this in release 13, a Class A and Class B CSI feedback terminology as well as measurement restriction was introduced (see Section 8.3.2.3).

In addition, SRS was enhanced in release 13 for small cell operation (short delay spreads) and when using TM 3, 8, 9, and 10. The enhancement involved the use of a lower density SRS (a comb structure with every fourth subcarrier) allowing increased SRS capacity.

8.3.4.6 Release 14: Massive MIMO enhancements

In release 14, the number of CSI-RS ports was increased to 32 with associated precoding matrix codebooks for CSI feedback, supporting both 1D and 2D antenna arrays (modeled as equally spaced with dual-polarized antennas). The CSI-RS overhead with so many ports becomes large and therefore an optional reduced-density CSI-RS was introduced. For Class B operation, some further enhancements to reduce CSI-RS overhead were introduced, including aperiodic CSI-RS transmission triggering in any subframe and medium access controlled (MAC) control element (CE)–based signaling to disable and enable a periodic CSI-RS transmission. Using MAC CE is faster than the RRC control supported by earlier releases and uses the PDSCH to carry the control information. This MAC CE–controlled transmission was termed *semipersistent* CSI-RS transmission.

Release 14 also introduced advanced CSI feedback (see Section 8.3.2.4), developed mainly for enhancing MU-MIMO performance by enabling a richer, more detailed channel feedback at the expense of increased feedback overhead. A hybrid CSI feedback mode was also introduced, which allows for switching back and forth between Class A and Class B type of feedback, to save CSI-RS overhead. Another CSI feedback mode introduced in release 14 was the semiopen-loop CSI feedback mode, where only the long-term channel properties (precoder) are fed back to the network.

8.3.4.7 Release 15: Multi-point transmission enhancements

In release 15, support for *non-coherent joint transmission* (NC-JT) was introduced, where a UE can receive one PDSCH containing layers from two different cells simultaneously. This required some extensions of the QCL framework where PDSCH DM-RS can be split into two groups where each group has its own source RS for QCL purpose.

8.3.4.8 Release 16: Enhancement targeting reciprocity-based operation

In release 16, SRS enhancements were targeted, with aim to increase the SRS capacity to better support reciprocity-based MIMO operation for DL transmission. This is achieved by allocating additional SRS symbols in a subframe where a release 16 UE can use a configurable number of symbols for SRS, including all the symbols in a subframe, if desirable.

8.3.5 PUSCH TRANSMISSION PROCEDURES AND UPLINK TRANSMISSION MODES

For UL, release 8 supported only single-layer PUSCH transmission, but in release 10, two TMs (TM 1 and TM 2) were defined for PUSCH, the single-port TM 1 (single-layer data) and the closed-loop spatial multiplexing mode that allows for up to four-layer PUSCH transmission (TM 2). LTE uses DFT-precoded OFDM for PUSCH (see Chapter 5) to reduce the PAPR of the waveform, sometimes referred to as the "single carrier property."

In release 8, two UEs could not be paired for MU-MIMO in the UL if their scheduled bandwidths were partially overlapping (i.e., the two UEs had different scheduling bandwidths). This is because UL DM-RS sequences with different CS for PUSCH are only orthogonal if they have the same length. Thus only CS can be used to separate users.

In release 10, a length 2 OCC was introduced for DM-RS to separate two DM-RS ports by OCC in addition to CS (see Section 8.3.1.6). Hence, MU-MIMO multiplexing capacity using orthogonal DM-RS was doubled compared to previous release, and partially overlapped MU-MIMO scheduling was possible if UEs are assigned different OCCs. Since one UE may be power-limited scenario, it can only transmit on a few RBs, while another UE closer to the eNB can use more; the situation that two UEs with different scheduling bandwidths are scheduled simultaneously is not uncommon. Hence, support for partially overlapping frequency resources was motivated.

In release 14, the UL DM-RS was further enhanced with the aim to support UL MU-MIMO with even higher multiplexing capacity. This was enabled by a comb-based DM-RS, in which a DM-RS sequence for an antenna port is mapped to either odd or even subcarriers (instead of all subcarriers in the previous releases). Two UEs with DM-RS on different combs can thus be scheduled with independently assigned resource allocations without losing DM-RS orthogonality.

8.4 LTE SUMMARY

An LTE evolution stretching over eight releases and several SI has resulted in a rich multi-antenna toolbox containing a variety of features including those useful for AAS deployments. Codebook-based, noncodebook-based, and reciprocity-based MIMO are supported for both TDD and FDD. The evolution moved from a robust design with few transmitting antenna ports to a very large number of antenna ports allowing for both horizontal and vertical beamforming simultaneously.

Current LTE has a limitation to two MU-MIMO scheduled UEs if orthogonal DM-RS ports are to be used. Hence, LTE was not primarily designed for MU-MIMO with many layers although the basic functionality is supported and somewhat enhanced to four UEs in a later release. Hence, non-orthogonal DM-RS must be used for higher-order MU-MIMO.

There are strong dependencies on UE capability support for reaching optimum performance since most UEs cannot implement the full toolbox but only the basic features from early releases.

An LTE problem in its evolution is the always present CRS, which causes a large overhead for DM-RS-based TMs. For these DM-RS based TMs, the UE does not need CRS for PDSCH demodulation or measurements, only for fine synchronization. When used only for this purpose, the CRS density is unnecessary high.

In addition, the CRS transmitted in neighboring cells causes interference and is transmitted even when there is no traffic in a cell, and the CRS cannot be disabled since it is a common signal to the entire cell. Hence, the full potential of MIMO and AAS scalability to many steerable transmit antennas cannot fully be exploited with LTE due to the interference and overhead by presence of CRS.

8.4.1 LTE—WHAT IS IN THE PIPE?

It is likely that LTE and NR in carrier frequencies below 6 GHz will evolve side by side when it comes to AAS functionality. The reason is that the same radio hardware and baseband software are commonly shared between the two radio access technologies, and it makes sense that functionality is similar for the two.

As discussed, LTE has several limitations, and the most critical ones are related to the CRS. Over time, traffic will be carried over to NR networks and dynamic spectrum sharing between LTE and NR is likely to be used for a foreseeable time.

However, one obstacle for LTE MIMO evolution toward AAS is the lack of commercial availability of UEs that support an advanced MIMO feature (e.g., 32-port CSI-RS) as those are introduced in the later releases.

REFERENCES

[1] E. Dahlman,, S. Parkvall,, J. Sköld,, 4G LTE/LTE-Advanced for Mobile Broadband, Academic Press, 2013.
[2] O. Liberg,, et al., Cellular Internet of Things: Technologies, Standards, and Performance, Academic Press, 2019.
[3] S.M. Alamouti, A simple transmit diversity technique for wireless communications, IEEE J. Sel. Areas Commun. 16 (8) (1998) 1451−1458.
[4] B. Hassibi, B. Hochwald, Linear dispersion codes, in: Proceedings of 2001 IEEE International Symposium on Information Theory, 2001.
[5] G. Bauch, T. Abe, On the parameter choice for cyclic delay diversity based precoding with spatial multiplexing, in: Proceedings of IEEE GLOBECOM 2009.
[6] 3GPP TR 36.913, V8.0.1 (2009-03), Requirements for further advancements for evolved universal terrestrial radio access (E-UTRA).
[7] 3GPP TR 36.814, V2.0.1. (2010-03), Further advancements for E-UTRA physical layer aspects.
[8] 3GPP TR 36.871, V11.0.0. (2011-12), Downlink multiple input multiple output (MIMO) enhancement for LTE-Advanced.
[9] 3GPP TR 38.211, V15.5.0. (2019-03), NR; Physical channels and modulation.
[10] 3GPP TR 36.897, V1.0.1. (2015-06), Study on elevation beamforming/full-dimension (FD) MIMO for LTE.