CHAPTER

13 Power Control, Scheduling, and Interference Handling

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

13.1 UPLINK POWER CONTROL

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

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

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

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

4G LTE/LTE-Advanced for Mobile Broadband.

13.1.1 Uplink Power Control – Some Basic Rules

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

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

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

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

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

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

If PUCCH is not transmitted in the subframe but L1/L2 control signaling is multiplexed on to PUSCH, the PUSCH carrying the L1/L2 control signaling is first assigned the power determined by its corresponding power-control algorithm, before any power is assigned to any other PUSCH to be transmitted in parallel. Once again, this ensures that L1/L2 control signaling is assigned the power assumed to be needed before any power is assigned for other PUSCH transmissions only carrying UL-SCH. Note that, in the case of transmission of multiple PUSCH in parallel (carrier aggregation), at most one PUSCH may include L1/L2 control signaling. Also, there cannot be PUCCH transmission and L1/L2 control signaling multiplexed on to PUSCH in the same subframe. Thus, there will never be any conflict between the above rules.

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

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

$$\sum_i w_c \cdot P_{\mathrm{PUSCH},c} \leq P_{\mathrm{TMAX}} - P_{\mathrm{PUCCH}}, \tag{13.1}$$

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

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

13.1.2 Power Control for PUCCH

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

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

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

$$P_{\mathrm{PUCCH}} = \min\left\{P_{\mathrm{CMAX},c}, P_{0,\mathrm{PUCCH}} + PL_{\mathrm{DL}} + \Delta_{\mathrm{Format}} + \delta\right\}. \tag{13.2}$$

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

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

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

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

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

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

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

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

13.1.3 Power Control for PUSCH

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

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

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

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

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

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

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

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

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

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

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

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

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

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

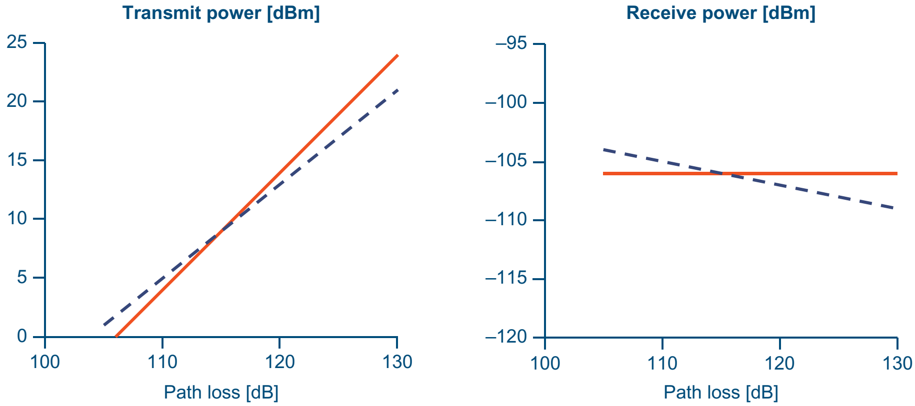

FIGURE 13.1

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

13.1.4 Power Control for SRS

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

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

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

13.1.5 Power Headroom

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

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

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

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

where the values for M and Δ_{MCS} correspond to the resource assignment and modulation-and-coding scheme used in the subframe to which the power-headroom report corresponds. It can be noted that the power headroom is not a measure of the difference between the maximum per-carrier transmit power and the actual carrier transmit power. Rather, comparing with expression (13.3) it can be seen that the power headroom is a measure of the difference between $P_{\mathrm{CMAX},c}$ and the transmit power that would have been used *assuming that there would have been no upper limit on the transmit power.* Thus, the power headroom can very well be negative. More exactly, a negative power headroom indicates that the per-carrier transmit power was limited by $P_{\mathrm{CMAX,c}}$ at the time of the power headroom reporting. As the network knows what modulation-and-coding scheme and resource size the terminal used for transmission in the subframe to which the power-headroom report corresponds, it can determine what are the valid combinations of modulation-and-coding scheme and resource size M, assuming that the downlink path loss PL_{DL} and the term δ have not changed substantially.

Type-1 power headroom can also be reported for subframes where there is no actual PUSCH transmission. In such cases, $10 \cdot \log_{10}(M)$ and Δ_{MCS} in the expression above are set to zero:

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

This can be seen as the power headroom assuming a default transmission configuration corresponding to the minimum possible resource assignment ($M = 1$) and the modulation-and-coding scheme associated with $\Delta_{\mathrm{MCS}} = 0\,\mathrm{dB}$.

Similarly, Type-2 power headroom reporting is defined as the difference between the maximum per-carrier transmit power and the sum of the PUSCH and PUCCH transmit power given by their corresponding power-control expressions (13.2) and (13.1) respectively, once again not taking into account any maximum per-carrier power when calculating the PUSCH and PUCCH transmit power.

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

13.2 SCHEDULING AND RATE ADAPTATION

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

For carrier aggregation, each component carrier is independently scheduled with individual scheduling assignments/grants and one DL-SCH/UL-SCH per scheduled component carrier. Semi-persistent scheduling is only supported on the primary component carriers, motivated by the fact that the main usage is for small payloads not requiring multiple component carriers.

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

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

The *uplink scheduler* serves a similar purpose, namely to dynamically control which terminals are to transmit on their UL-SCH (or UL-SCHs in the case of carrier aggregation) and on which uplink resources. The uplink scheduler is in complete control of the transport format the terminal will use, whereas the logical-channel multiplexing is controlled by the terminal according to a set of rules. Thus, uplink scheduling is *per terminal* and not per radio bearer. This is illustrated in the right part of Figure 13.2, where the scheduler controls the transport format and the terminal controls the logical-channel multiplexing.

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

13.2.1 Downlink Scheduling

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

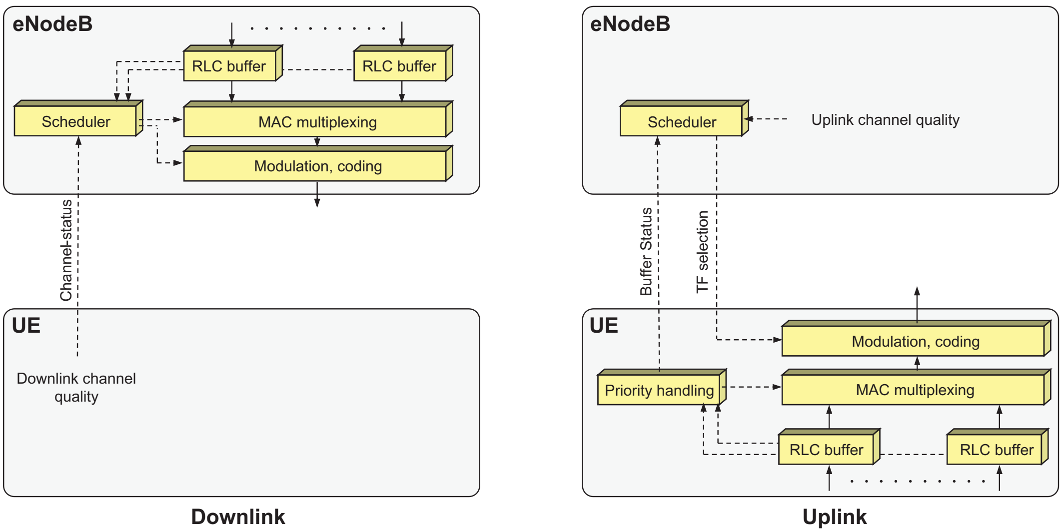

FIGURE 13.2

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

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

- *RLC*. Segmentation/concatenation of RLC SDUs is directly related to the instantaneous data rate. For low data rates, it may only be possible to deliver a part of an RLC SDU in a TTI, in which case segmentation is needed. Similarly, for high data rates, multiple RLC SDUs may need to be concatenated to form a sufficiently large transport block.
- *MAC*. Multiplexing of logical channels depends on the priorities between different streams. For example, radio resource control signaling, such as handover commands, typically has a higher priority than streaming data, which in turn has higher priority than a background file transfer. Thus, depending on the data rate and the amount of traffic of different priorities, the multiplexing of different logical channels is affected. Hybrid-ARQ retransmissions also need to be accounted for.
- *L1*. Coding, modulation and, if applicable, the number of transmission layers and the associated precoding matrix are obviously affected by the scheduling decision. The choices of these parameters are mainly determined by the radio conditions and the selected data rate – that is, the transport block size.

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

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

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

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

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

Information about the channel conditions at the terminal can be obtained in several ways. In principle, the eNodeB can use any information available, but typically the channel-state reports from the terminal, further described in Section 13.2.5, are used. However, additional sources of channel knowledge, for example exploiting channel reciprocity to estimate the downlink quality from uplink channel estimates in the case of TDD, can also be exploited by a particular scheduler implementation.

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

Downlink inter-cell interference coordination is also part of the implementation-specific scheduler strategy. A cell may signal to its neighboring cells the intention to transmit with a lower transmission power in the downlink on a set of resource blocks. This information can then be exploited by neighboring cells as a region of low interference where it is advantageous to schedule terminals at the cell edge, terminals that otherwise could not attain high data rates due to the interference level. Inter-cell interference handling is further discussed in Section 13.3.

13.2.2 Uplink Scheduling

The basic function of the *uplink scheduler* is similar to its downlink counterpart, namely to dynamically determine, for each 1 ms interval, which terminals are to transmit and on which uplink resources. As discussed before, the LTE uplink is primarily based on maintaining orthogonality between different uplink transmissions and the shared resource controlled by the eNodeB scheduler is time–frequency resource units. In addition to assigning the time–frequency resources to the terminal, the eNodeB scheduler is also responsible for controlling the transport format the terminal will use for each of the uplink component carriers. As the scheduler knows the transport format the terminal will use when it is transmitting, there is no need for outband control signaling from the terminal to the eNodeB. This is beneficial from a coverage perspective, taking into account that the cost per bit of transmitting outband control information can be significantly higher than the cost of data transmission, as the control signaling needs to be received with higher reliability. It also allows the scheduler to tightly control the uplink activity to maximize the resource usage compared to schemes where the terminal autonomously selects the data rate, as autonomous schemes typically require some margin in the scheduling decisions. A consequence of the scheduler being responsible for selection of the transport format is that accurate and detailed knowledge about the terminal situation with respect to buffer status and power availability is more accentuated in LTE compared to systems where the terminal autonomously controls the transmission parameters.

The basis for uplink scheduling is *scheduling grants*, containing the scheduling decision and providing the terminal information about the resources and the associated transport format to use for transmission of the UL-SCH on one component carrier. Only if the terminal has a valid grant is it allowed to transmit on the corresponding UL-SCH; autonomous transmissions are not possible without a corresponding grant. Dynamic grants are valid for one subframe – that is, for each subframe in which the terminal is to transmit on the UL-SCH, the scheduler issues a new grant. Uplink component carriers are scheduled independently; if the terminal is to transmit simultaneously on multiple component carriers, multiple scheduling grants are needed.

The terminal monitors a set of PDCCHs as described in Chapter 10 for uplink scheduling grants. Upon detection of a valid uplink grant, the terminal will transmit its UL-SCH according to the information in the grant. Obviously, the grant cannot relate to the same subframe it was received in as the uplink subframe has already started when the terminal has decoded the grant. The terminal also needs

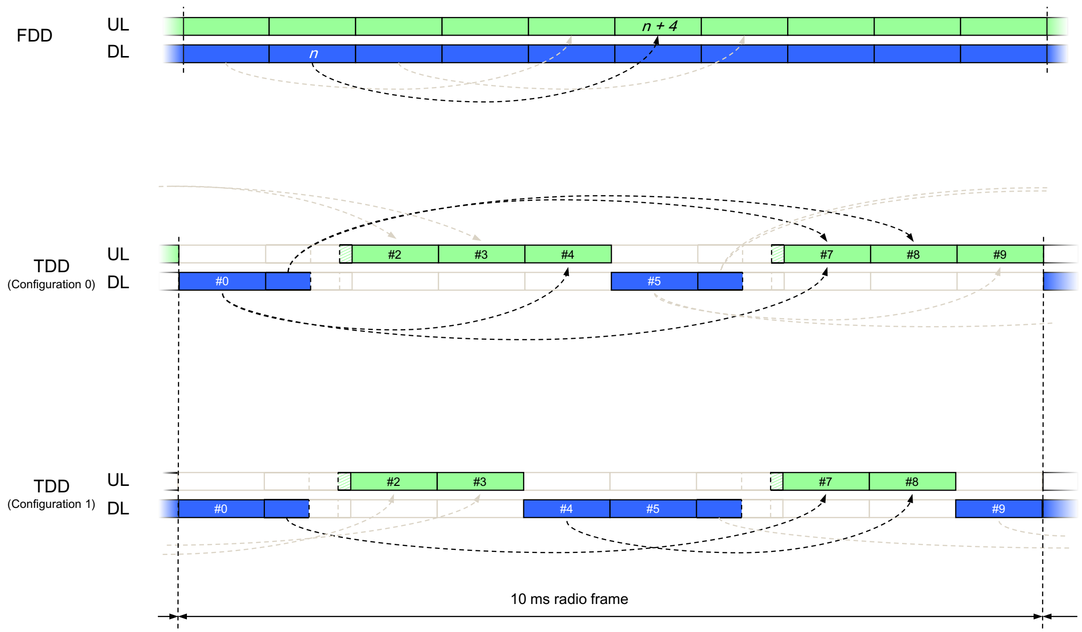

FIGURE 13.3

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

some time to prepare the data to transmit. Therefore, a grant received in subframe n affects the uplink transmission in a later subframe.

For FDD, the grant timing is straightforward. An uplink grant received in a subframe n triggers an uplink transmission in subframe $n + 4$, as illustrated in Figure 13.3. This is the same timing relation as used for uplink retransmission triggered by the PHICH, motivated by the possibility to override the PHICH by a dynamic scheduling grant, as described in Chapter 12.

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

For TDD configuration 0 there are more uplink subframes than downlink subframes, which calls for the possibility to schedule transmissions in multiple uplink subframes from a single downlink subframe. The same timing relation as for the other TDD configurations is used but with slight modifications. Recall from Chapter 10 that the grant transmitted in the downlink contains an uplink index consisting of two bits. For downlink–uplink configuration 0, the index field specifies which uplink subframe(s) a grant received in a downlink subframe applies to. For example, as illustrated in Figure 13.3, an uplink scheduling grant received in downlink subframe 0 applies to one or both of the uplink subframes 4 and 7, depending on which of the bits in the uplink index are set.

Similarly to the downlink case, the uplink scheduler can exploit information about channel conditions, buffer status, and priorities of the different data flows, and, if some form of interference coordination is employed, the interference situation in neighboring cells. Channel-dependent scheduling, which typically is used for the downlink, can be used for the uplink as well. In the uplink, estimates of the channel quality can be obtained from the use of uplink channel sounding, as described in Chapter 11. For scenarios where the overhead from channel sounding is too costly, or when the variations in the channel are too rapid to be tracked, for example at high terminal speeds, uplink diversity can be used instead. The use of frequency hopping as discussed in Chapter 11 is one example of obtaining diversity in the uplink.

Inter-cell interference coordination can be used in the uplink for similar reasons as in the downlink by exchanging information between neighboring cells, as discussed in Section 13.3.

13.2.2.1 *Uplink Priority Handling*

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

The simplest multiplexing rule would be to serve logical channels in strict priority order. However, this may result in starvation of lower-priority channels; all resources would be given to the high-priority channel until its transmission buffer is empty. Typically, an operator would instead like to provide at least some throughput for low-priority services as well. Therefore, for each logical channel in an LTE terminal, a *prioritized data rate* is configured in addition to the priority value. The logical channels are then served in decreasing priority order up to their prioritized data rate, which avoids starvation as long as the scheduled data rate is at least as large as the sum of the prioritized data rates. Beyond the prioritized data rates, channels are served in strict priority order until the grant is fully exploited or the buffer is empty. This is illustrated in Figure 13.4.

13.2.2.2 *Scheduling Requests and Buffer Status Reports*

The scheduler needs knowledge about the amount of data awaiting transmission from the terminals to assign the proper amount of uplink resources. Obviously, there is no need to provide uplink resources to a terminal with no data to transmit as this would only result in the terminal performing padding to

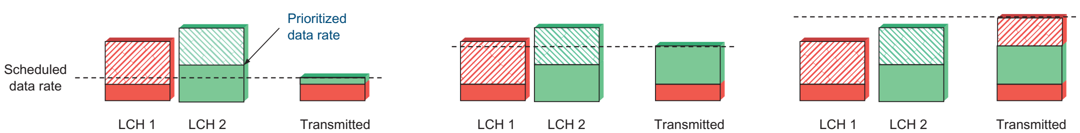

FIGURE 13.4

Prioritization of two logical channels for three different uplink grants.

fill up the granted resources. Hence, as a minimum, the scheduler needs to know whether the terminal has data to transmit and should be given a grant. This is known as a *scheduling request*.

A scheduling request is a simple flag, raised by the terminal to request uplink resources from the uplink scheduler. Since the terminal requesting resources by definition has no PUSCH resource, the scheduling request is transmitted on the PUCCH. Each terminal can be assigned a dedicated PUCCH scheduling request resource, occurring every *n*th subframe, as described in Chapter 11. With a dedicated scheduling-request mechanism, there is no need to provide the identity of the terminal requesting to be scheduled as the identity of the terminal is implicitly known from the resources upon which the request is transmitted. When data with higher priority than already existing in the transmit buffers arrives at the terminal and the terminal has no grant and hence cannot transmit the data, the terminal transmits a scheduling request at the next possible instant, as illustrated in Figure 13.5. Upon reception of the request, the scheduler can assign a grant to the terminal. If the terminal does not receive a scheduling grant until the next possible scheduling-request instant, then the scheduling request is repeated. There is only a single scheduling-request bit, irrespective of the number of uplink component carriers the terminal is capable of. In the case of carrier aggregation, the scheduling request is transmitted on the primary component carrier, in line with the general principle of PUCCH transmission on the primary component carrier only.

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

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

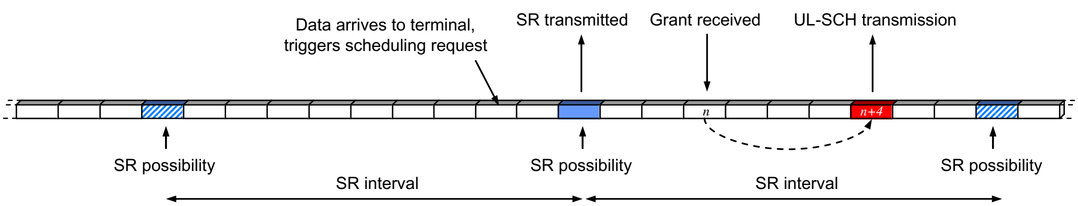

FIGURE 13.5

Scheduling-request transmission.

are suitable for a situation where there are a large number of terminals in the cell and the traffic intensity, and hence the scheduling intensity, is low. In situations with higher intensities, the collision rate between different terminals simultaneously requesting resources would be too high and lead to an inefficient design.

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

Terminals that already have a valid grant obviously do not need to request uplink resources. However, to allow the scheduler to determine the amount of resources to grant to each terminal in future subframes, information about the buffer situation and the power availability is useful, as discussed above. This information is provided to the scheduler as part of the uplink transmission through MAC control elements (see Chapter 8 for a discussion on MAC control elements and the general structure of a MAC header). The LCID field in one of the MAC subheaders is set to a reserved value indicating the presence of a buffer status report, as illustrated in Figure 13.6.

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

- Arrival of data with higher priority than currently in the transmission buffer – that is, data in a logical-channel group with higher priority than the one currently being transmitted – as this may impact the scheduling decision.
- Change of serving cell, in which case a buffer-status report is useful to provide the new serving cell with information about the situation in the terminal.
- Periodically as controlled by a timer.

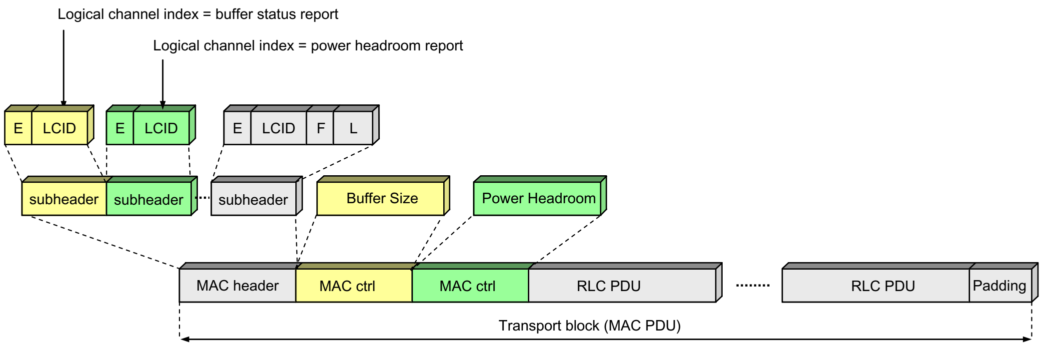

FIGURE 13.6

Signaling of buffer status and power-headroom reports.

- Instead of padding. If the amount of padding required to match the scheduled transport block size is larger than a buffer-status report, a buffer-status report is inserted. Clearly it is better to exploit the available payload for useful scheduling information instead of padding if possible.

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

Information about the power headroom is fed back from the terminals to the eNodeB in a similar way as the buffer-status reports – that is, only when the terminal is scheduled to transmit on the UL-SCH. Type-1 reports are provided for all component carriers simultaneously, while Type-2 reports are provided for the primary component carrier only.

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

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

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

13.2.3 Semi-Persistent Scheduling

The basis for uplink and downlinks scheduling is dynamic scheduling, as described in Sections 13.2.1 and 13.2.2. dynamic scheduling with a new scheduling decision taken in each subframe allows for full flexibility in terms of the resources used and can handle large variations in the amount of data to transmit at the cost of the scheduling decision being sent on a PDCCH in each subframe. In many situations, the overhead in terms of control signaling on the PDCCH is well motivated and relatively small compared to the payload on DL-SCH/UL-SCH. However, some services, most notably voice-over IP, are characterized by regularly occurring transmission of relatively small payloads. To reduce the control signaling overhead for those services, LTE provides semi-persistent scheduling in addition to dynamic scheduling.

With semi-persistent scheduling, the terminal is provided with the scheduling decision on the PDCCH, together with an indication that this applies to every *n*th subframe until further notice. Hence, control signaling is only used once and the overhead is reduced, as illustrated in Figure 13.7. The periodicity of semi-persistently scheduled transmissions – that is, the value of *n* – is configured by RRC signaling in advance, while activation (and deactivation) is done using the PDCCH using the semi-persistent C-RNTI.[3] For example, for voice-over IP the scheduler can configure a periodicity

[3]Each terminal has two identities, the "normal" C-RNTI for dynamic scheduling and the semi-persistent C-RNTI for semi-persistent scheduling.

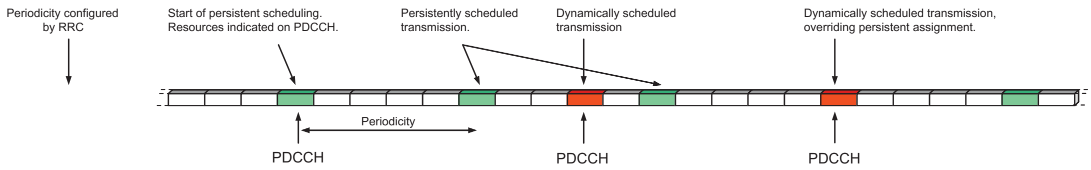

FIGURE 13.7

Example of semi-persistent scheduling.

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

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

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

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

13.2.4 Scheduling for Half-Duplex FDD

Half-duplex FDD implies that a single terminal cannot receive and transmit at the same time while the eNodeB still operates in full duplex. In LTE, half-duplex FDD is implemented as a scheduler constraint, implying it is up to the scheduler to ensure that a single terminal is not scheduled simultaneous in uplink and downlink. Hence, from a terminal perspective, subframes are dynamically used for uplink or downlink. Briefly, the basic principle for half-duplex FDD is that a terminal is receiving in the downlink unless it has been explicitly instructed to transmit in the uplink (either UL-SCH transmission or hybrid-ARQ acknowledgements triggered by a downlink transmission). The timing and structure for control signaling are identical between half- and-full duplex FDD terminals.

An alternative design approach would be to base half-duplex FDD on the TDD control signaling structure and timing, with a semi-static configuration of subframes to either downlink or uplink. However, this would complicate supporting a mixture of half- and-full duplex terminals in the same cell as the timing of the control signaling would differ. It would also imply a waste of uplink spectrum resources. All terminals need to be able to receive subframes 0 and 5, as those subframes are used for system information and synchronization signals. Hence, if a fixed uplink–downlink allocation were to be used, no uplink transmissions could take place in those two subframes, resulting in a

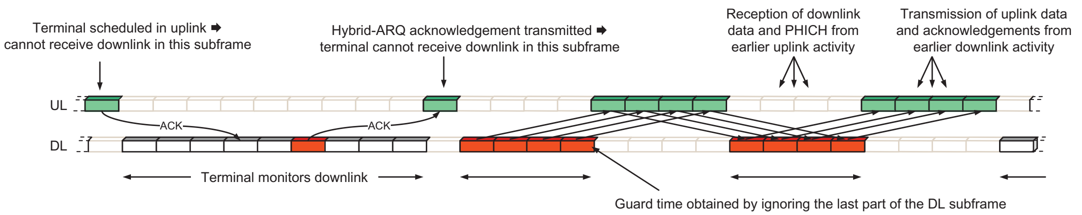

FIGURE 13.8

Example of half-duplex FDD terminal operation.

loss in uplink spectral efficiency of 20%. Clearly this is not attractive and led to the choice of implementing half-duplex FDD as a scheduling strategy instead.

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

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

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

13.2.5 Channel-State Reporting

As mentioned several times, the possibility for downlink channel-dependent scheduling – that is, selecting the downlink transmission configuration and related parameters depending on the instantaneous downlink channel conditions – is a key feature of LTE. An important part of the support for downlink channel-dependent scheduling is *channel-state reports* provided by terminals to the network, reports on which the latter can base its scheduling decisions.

The channel-state reports consist of one or several pieces of information:

- *Rank indication* (RI), providing a recommendation on the transmission rank to use or, expressed differently, the number of layers that should preferably be used for downlink transmission to the terminal. RI only needs to be reported by terminals that are configured to be in one of the spatial-multiplexing transmission modes. There is at most one RI reported, valid across the full bandwidth – that is, the RI is frequency non-selective. Frequency-dependent transmission rank would be impossible to utilize since all layers are transmitted on the same set of resource blocks in LTE.
- *Precoder matrix indication* (PMI), indicating which of the precoder matrices (see Chapter 10) should preferably be used for the downlink transmission. The reported precoder matrix is determined assuming the number of layers indicated by the RI. The precoder recommendation may be frequency selective, implying that the terminal may recommend different precoders for different parts of the downlink spectrum. Furthermore, the network can restrict the set of matrices from which the terminal should select the recommended precoder, so-called codebook subset restriction, to avoid reporting precoders that are not useful in the antenna setup used.
- *Channel-quality indication* (CQI), representing the highest modulation-and-coding scheme that, if used, would mean PDSCCH transmissions (using the recommended RI and PMI) were received with a block-error rate of at most 10%. The reason to use CQI as a feedback quantity instead of, for example, the signal-to-noise ratio, is to account for different receiver implementation in the terminal. Also, basing the feedback reports on CQI instead of signal-to-noise ratio also simplifies the testing of terminals; a terminal delivering data with more than 10% block-error probability when using the modulation-and-coding scheme indicated by the CQI would fail the test. As will be discussed further below, multiple CQI reports, each representing the channel quality in a certain part of the downlink spectrum, can be part of a channel-state report.

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

Although referred to as channel-state reports, what a terminal delivers to the network are not explicit reports of the downlink channel state. Rather, what the terminal delivers are *recommendations* on the transmission rank and precoding matrix to use, together with an indication of the highest possible modulation-and-coding scheme that the network preferably should not exceed. Information about the actual modulation scheme and coding rate used for DL-SCH transmission as well as the set of resource blocks used for the transmission is always included in the downlink scheduling assignment. Hence, the eNodeB is free to follow the CSI report or to select transmission parameters on its own.

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

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

- The network may follow the latest terminal recommendation, in which case the eNodeB only has to confirm (a one-bit indicator in the downlink scheduling assignment) that the precoder configuration recommended by the terminal is used for the downlink transmission. On receiving such a confirmation, the terminal will use its recommended configuration when demodulating and decoding the corresponding DL-SCH transmission. Since the PMI computed in the terminal can be frequency selective, an eNodeB following the precoding matrix recommended by the terminal may have to apply different precoding matrices for different (sets of) resource blocks.
- The network may select a different precoder, information about which then needs to be explicitly included in the downlink scheduling assignment. The terminal then uses this configuration when demodulating and decoding the DL-SCH. To reduce the amount of downlink signaling, only a single precoding matrix can be signaled in the scheduling assignment, implying that, if the network overrides the recommendation, then the precoding is frequency non-selective. The network may also choose to override the transmission rank only, in which case the terminal assumes that a subset of the columns in each of the recommended precoder matrices is used.

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

- Aperiodic channel-state reports are delivered when explicitly requested by the network by means of the channel-state-request flag included in uplink scheduling grants (see Section 10.4.5). An aperiodic channel-state report is always delivered using the PUSCH – that is, on a dynamically assigned resource.
- Periodic channel-state reports are configured by the network to be delivered with a certain periodicity, possibly as often as once every 2 ms, on a semi-statically configured PUCCH resource. However, similar to hybrid-ARQ acknowledgements normally delivered on PUCCH, channel-state reports are "re-routed" to the PUSCH[4] if the terminal has a valid uplink grant and is anyway to transmit on the PUSCH.

Aperiodic and periodic reports, despite both providing estimates on the channel conditions, are quite different in terms of their detailed contents and the usage. In general, aperiodic reports are larger and more detailed than their periodic counterparts. There are several reasons for this. First, the PUSCH, upon which the aperiodic report is transmitted, is capable of a larger payload, and hence a more detailed report, than the PUCCH used for the periodic reports. Furthermore, as aperiodic reports are transmitted on a per-need basis only, the overhead from these reports are less of an issue compared to periodic reports. Finally, if the network requests a report it is likely that it will transmit a large amount of data to the terminal, which makes the overhead from the report less of an issue compared to a periodic report that is transmitted irrespective of whether the terminal in question will be scheduled in the near future or not. Hence, as the structure and usage of aperiodic and periodic reports is different, they are described separately below, starting with aperiodic reports.

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

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

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

The different aperiodic reporting modes are summarized in Table 13.1.

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

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

- Wideband reports, reflecting the average channel quality across the entire cell bandwidth with a single CQI value. If PMI reporting is enabled, a single PMI valid across the full bandwidth is reported.
- UE-selected reports. Although named in the same way as for aperiodic reports, the principle for UE-selected periodic reports is different. The total bandwidth (of a component carrier) is divided into one to four *bandwidth parts*, with the number of bandwidth parts obtained form the cell bandwidth. For each bandwidth part, the terminal selects the best sub-band within that part. The

Table 13.1 Possible Aperiodic Reporting Modes for Different Transmission Modes

Transmission Mode		**Reporting Mode**							
		Wideband CQI		**Frequency-Selective CQI**					
				UE-Selected Sub-Bands			**Conf. Sub-Bands**		
		1-0: No PMI	**1-1: Wideband PMI**	**1-2: Selective PMI**	**2-0: No PMI**	**2-1: Wideband PMI**	**2-2: Selective PMI**	**3-0: No PMI**	**3-1: Wideband PMI**
1	Single antenna, CRS				•			•	
2	Transmit diversity				•			•	
3	Open-loop spatial mux.				•			•	
4	Closed-loop spatial mux.			•			•		•
5	Multi-user MIMO								•
6	Codebook-based beam-forming			•			•		•
7	Single-layer trans., DM-RS				•			•	
8	Dual-layer trans., DM-RS			•	•		•	•	•
9	Multi-layer trans., DM-RS			•	•		•	•	•

sub-band size ranges from four to eight resource blocks. Since the supported payload size of the PUCCH is limited, the reporting cycles through the bandwidth parts and in one subframe report the wideband CQI and PMI (if enabled) for that bandwidth part, as well as the best sub-band and the CQI for that sub-band. The RI (if enabled) is reported in a separate subframe.

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

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

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

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

For aperiodic reporting, the two-bit[5] CSI request in the downlink control signaling allows for three different types of CSI reports to be requested (the fourth bit combination represents no CSI request). Of these three alternatives, one is used to trigger a CSI reports for the downlink component carrier associated with the uplink component carrier for which the scheduling grant relates to. The remaining alternatives point to one of two configurable combinations of component carriers for which the CSI report should be generated. Thus, as an example, for a terminal capable of two downlink component carriers, aperiodic reports can, with the proper configuration, be requested for the primary component carrier, the secondary component carrier, or both.

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

13.2.6 Discontinuous Reception (DRX) and Component Carrier Deactivation

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

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

Table 13.2 Possible Periodic Reporting Modes for Different Transmission Modes

Transmission Mode		**Reporting Mode**							
		Wideband CQI		**Frequency-Selective CQI**					
				UE-Selected Sub-Bands			**Conf. Sub-Bands**		
		1-0: No PMI	**1-1: Wideband PMI**	**1-2: Selective PMI**	**2-0: No PMI**	**2-1: Wideband PMI**	**2-2: Selective PMI**	**3-0: No PMI**	**3-1: Wideband PMI**
1	Single antenna, CRS	•			•				
2	Transmit diversity	•			•				
3	Open-loop spatial mux.	•			•				
4	Closed-loop spatial mux.		•			•			
5	Multi-user MIMO		•			•			
6	Codebook-based beam-forming		•			•			
7	Single-layer trans., DM-RS	•			•				
8	Dual-layer trans., DM-RS	•	•		•	•			
9	Multi-layer trans., DM-RS	•	•		•	•			

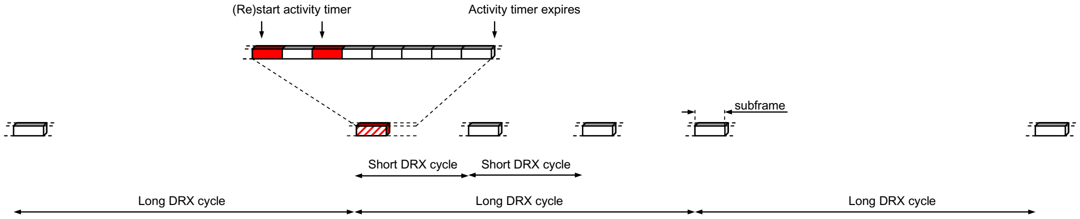

FIGURE 13.9

Illustration of DRX operation.

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

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

Retransmissions take place regardless of the DRX cycle. Thus, the terminal receives and transmits hybrid-ARQ acknowledgements as normal in response to data transmission. In the uplink, this also includes retransmissions in the subframes given by the synchronous hybrid-ARQ timing relation. In the downlink, where asynchronous hybrid ARQ is used, the retransmission time is not fixed in the specifications. To handle this, the terminal monitors the downlink for retransmissions in a configurable time window after the previous transmission.

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

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

Activation and deactivation of downlink component carriers are done through MAC control elements. A command to activate a component carrier takes effect eight subframes after receipt – that is, if the MAC control element is received in subframe n, then the additional component carriers are activated starting from subframe $n + 8$. There is also a timer-based mechanism for deactivation such that a terminal may, after a configurable time with no activity on a certain component carrier, deactivate that component carrier. The primary component carrier is always active as it must be possible for the network to communicate with the terminal.

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

13.3 INTER-CELL INTERFERENCE COORDINATION

Like all modern mobile-communication technologies, LTE can be deployed with one-cell frequency reuse. Fundamentally, this implies that the LTE transmission structure has been designed so that reliable transmission is also possible for the low signal-to-interference ratios (SIR) that may occur in a reuse-one deployment when the same time–frequency resource issued in neighboring cells (SIR as low as -5 dB or even somewhat lower in the worst case[6]). This is especially true for transmission of critical information such as system information and L1/L2 control signaling. It should be noted that data transmission on DL-SCH can always be made sufficiently reliable by selecting a sufficiently low instantaneous data rate in combination with hybrid-ARQ retransmissions.

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

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

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

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

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

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

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

The above discussion implicitly assumed downlink transmission with terminals at the cell edge being interfered by downlink transmissions from other cells. The concept of inter-cell interference coordination is equally applicable to the uplink, although the interference situation in this case is somewhat different.

For the uplink, the interference level experienced by a certain link does not depend on where the transmitting terminal is located, but rather on the location of the *interfering* terminals, with interfering terminals closer to the cell border causing more interference to neighboring cells. The location of the transmitting terminal is still important though, as a terminal closer to the cell site can raise its transmission power to compensate for high interference coming from terminals in neighboring cells, something that may not be possible for cell-edge terminals. Thus, the fundamental goal of uplink inter-cell interference is the same as for the downlink – that is, to coordinate the, in this case, uplink scheduling between cells to avoid simultaneous transmissions from terminals at the cell border in neighboring cells causing severe interference to each other.

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

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

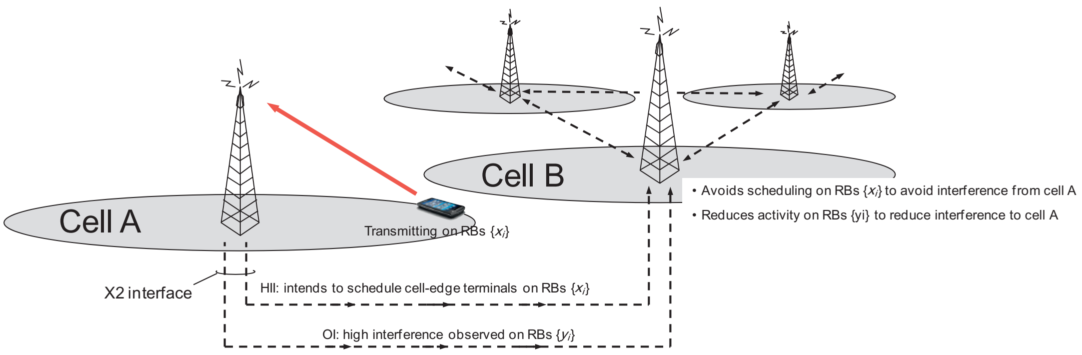

FIGURE 13.10

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

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

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

The *High Interference Indicator* provides information about the set of resource blocks within which an eNodeB is likely to schedule transmissions from cell-edge terminals – that is, resource blocks on which a neighboring cell can expect higher interference. Although nothing is explicitly specified on how an eNodeB should react to the HII (or any other ICIC-related X2 signaling) received from a neighboring eNodeB, a reasonable action for the receiving eNodeB would be to try to avoid scheduling its own cell-edge terminals on the same resource blocks, thereby reducing the uplink interference to cell-edge transmissions in its own cell as well as in the cell from which the HII was received. The HII can thus be seen as a *proactive* tool for ICIC, trying to prevent the occurrence of *too-low-SIR* situations.

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

For the downlink, *the Relative Narrowband Transmit Power* (RNTP) is defined to support ICIC operation (see Figure 13.11). The RNTP is similar to the HII in the sense that it provides information, for each resource block, whether or not the relative transmit power of that resource block is to exceed a certain level. Similar to the HII, a neighboring cell can use the information provided by the received RNTP when scheduling its own terminals, especially terminals on the cell edge that are more likely to be interfered by the neighboring cell.

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

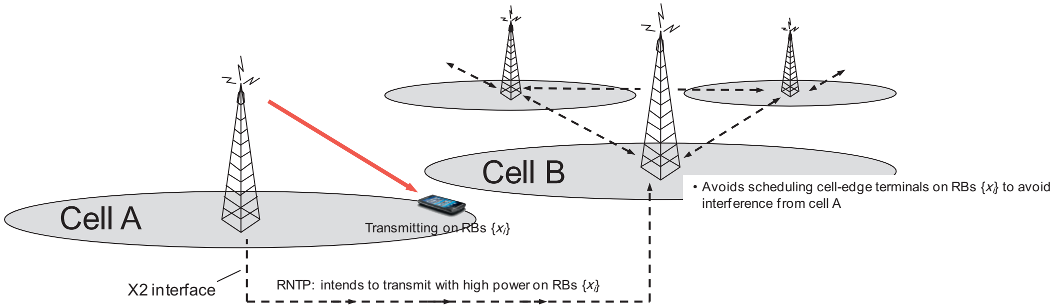

FIGURE 13.11

Illustration of downlink ICIC based on RNTP X2 signaling.

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

13.4 HETEROGENEOUS NETWORK DEPLOYMENTS

The continuous increase in traffic within mobile-broadband systems and an equally continuous increase in terms of the data rates requested by end-users will impact how cellular networks are deployed in the future. In general, providing very high system capacity (traffic per m^2) and very high per-user data rates will require a densification of the radio-access network – that is, the deployment of additional network nodes. By increasing the number of cells, the traffic per m^2 can be increased without requiring a corresponding increase in the traffic that needs to be supported per network node. Also, by increasing the number of network nodes, the base-station-to-terminal distances will, in general, be shorter, implying a link-budget improvement and a corresponding improvement in achievable data rates.

A general densification of the macro-cell layer – that is, reducing the coverage area of each cell and increasing the total number of macro-cell sites[8] – as illustrated in the upper part of Figure 13.12, is a path that has already been taken by many operators. As an example, in many major cities the distance between macro-cell sites is often less than a few hundred meters in many cases.

An alternative or complement to a uniform densification of the macro-cell layer is to deploy additional lower-power nodes under the coverage area of a macro cell, as illustrated in the lower part of Figure 13.12. In such a *heterogeneous* or *multi-layered* network deployment, the underlaid *pico-cell* layer does not need to provide full-area coverage. Rather, pico sites can be deployed to increase capacity and achievable data rates where needed. Outside of the pico-layer coverage, terminals would access the network by means of the overlaid macro cell.

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

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

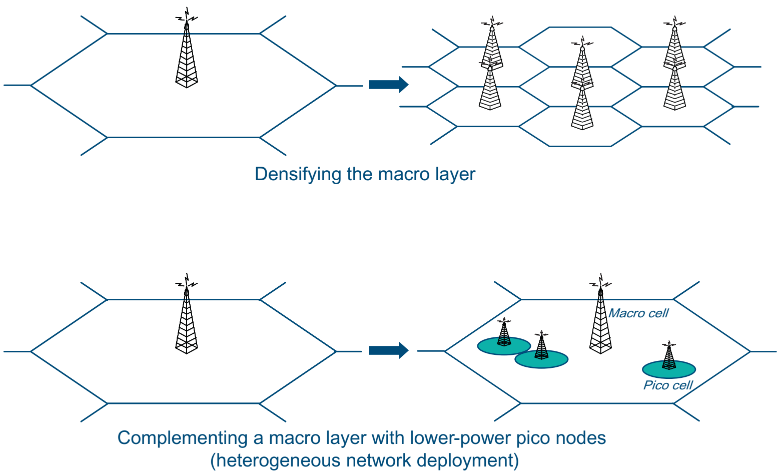

FIGURE 13.12

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

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

13.4.1 Interference Handling in a Heterogeneous Deployment

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

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

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

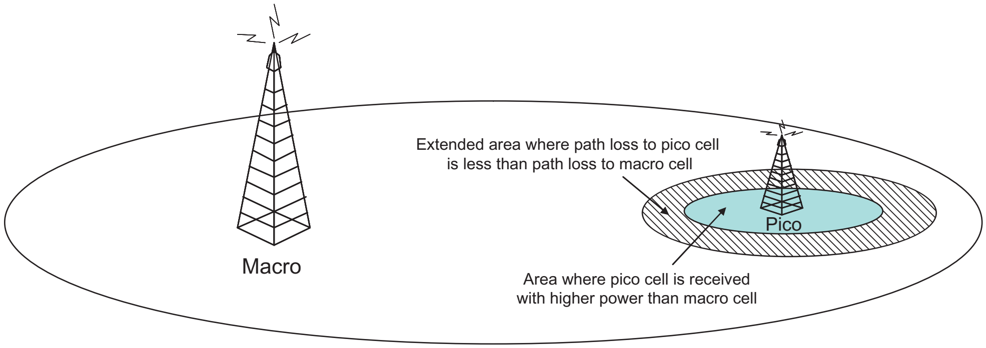

FIGURE 13.13

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

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

The characteristics of the inter-layer interference in a heterogeneous network deployment will also depend on the exact cell-selection strategy being used. Conventional cell selection is typically based on terminal measurements of the received power of some downlink signal, more specifically the cell-specific reference signals in the case of LTE. However, in a heterogeneous network deployment with cells with substantially different transmit power, including different power of the reference signals, selecting the cell that is received with the highest power implies that the terminal may often select the higher-power macro cell even if the path loss to a pico cell is significantly smaller. This will obviously not be optimal from an uplink coverage and capacity point of view.

It should also be noted that, even in terms of downlink system efficiency, it may not be optimal to select the cell with the highest received power in a heterogeneous network deployment. Although the high-power macro cell is received with higher power, this is at least partly due to the higher macro-cell transmit power. In that case, transmission from the macro cell is associated with a higher "cost" in terms of interference to other cells. Expressed alternatively, a transmission from the macro cell will prohibit the use of the same physical resource in *any* of the underlaid pico cells.

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

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

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

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

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

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

Less obvious is how to handle interference due to the macro-cell transmissions that cannot be dynamically scheduled, such as the L1/L2 control signaling (PDCCH, PCFICH, and PHICH). Within a cell layer, for example between two macro cells, interference between such transmissions is not a critical issue as LTE, including its control channels, has been designed to allow for one-cell frequency reuse and a corresponding SIR down to and even below −5 dB. However, in a heterogeneous network deployment with extensive range expansion, the downlink interference from the macro cell can be more than 15 dB above the signal received from the pico cell, corresponding to an SIR below −15 dB. In such low SIR it will not be possible to correctly decode even the very robust control channels. Instead, means are needed to avoid macro-to-pico interference for the L1/L2 control channels.

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

In terms of data (PDSCH) transmission, both carriers are available in both cell layers and interference between the layers is handled by "conventional" inter-cell interference coordination, as discussed above. As already mentioned, such interference coordination can be more or less dynamic depending on the time scale on which the macro cell and its underlaid pico cells can be coordinated. Also, the possibility for carrier aggregation allows for both carriers – that is, the total available spectrum – to be assigned for transmission to a single terminal. Thus, at least for carrier-aggregation-capable terminals, there is no spectrum fragmentation in terms of data (PDSCH) transmission.

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

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

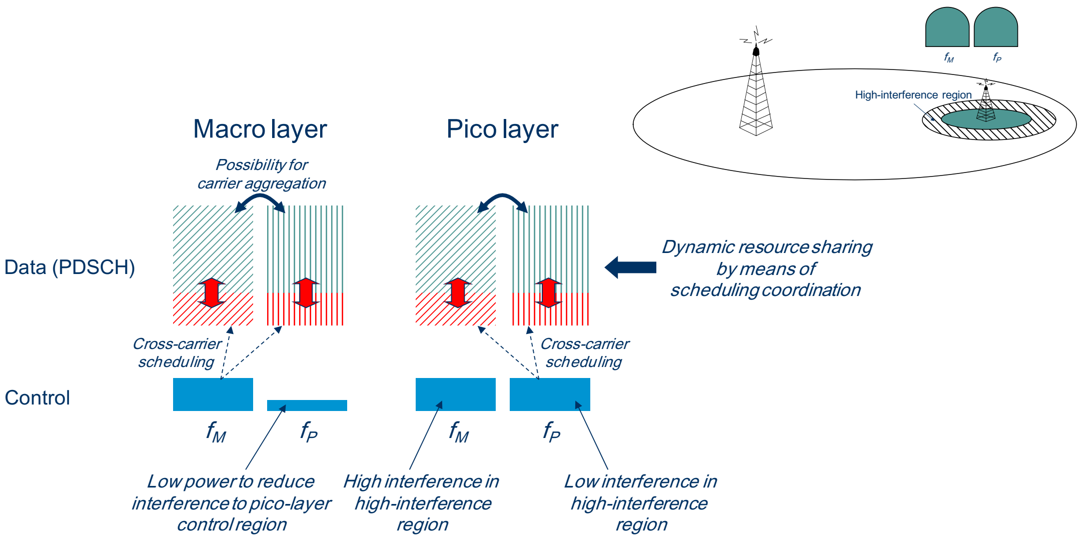

FIGURE 13.14

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

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

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

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

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

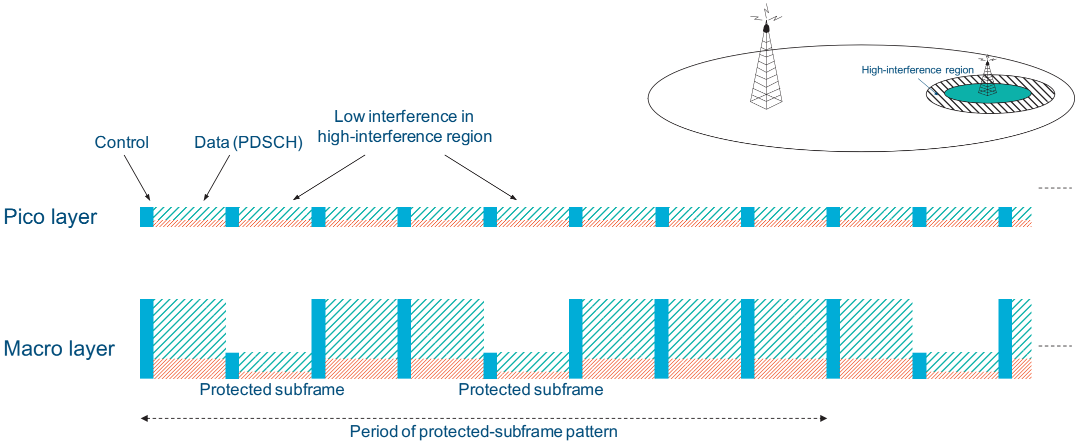

FIGURE 13.15

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

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

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

It should be noted that the macro cell must not necessarily completely avoid control-signaling transmission in the protected subframes. In particular, it could be beneficial to retain the possibility for a limited amount of control signaling related to uplink transmissions, for example a limited amount of uplink scheduling grants and/or PHICH transmission, in order not to cause too much impact on the uplink scheduling. As long as the macro-cell control-signaling transmissions are limited and only occupy a small fraction of the overall control region, the interference to terminals in the high-interference region of the pico cell could be kept at an acceptable level. However, the signaling of protected-subframe patterns is also defined so that impact on the uplink scheduling is minimized even if no uplink scheduling grants and PHICH can be transmitted in protected subframes. This is achieved by having the protected-subframe patterns matched to the eight-subframe timing of the uplink hybrid-ARQ protocol. It should be noted that this implies that the pattern is not aligned to the 10 ms frame but only to a 40 ms four-frame structure for FDD. For TDD the periodicity also depends on the uplink-downlink configuration.

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

It should also be noted that, in some cases, a pico cell could be located at the border between, and suffer interference from, two macro cells. If the macro cells have different configured and only partly overlapping sets of protected subframes, the pico-cell scheduling as well as the configuration of the CSI-measurement sets need to take the structure of the protected sets of both macro cells into account.

13.4.2 Interference Coordination in the Case of Home-eNodeB

In the case of home-eNodeB with CSG there are additional interference issues due to the fact that a terminal can be very close to the home-eNodeB and still have to communicate with the overlaid macro cell. Such a terminal may then be severely interfered on the downlink by any home-eNodeB transmission and may also cause severe uplink interference to the home-eNodeB. In principle, this can be solved by the same means as above – that is, by relying on interference coordination between the scheduling in the home-eNodeB layer and an overlaid macro, possibly extended by the carrier-aggregation approach.

A key difference in this case, though, is that the interference avoidance must be two-way – that is, one must not only avoid interference from the macro cell to home-eNodeB terminals in the high-interference outer region of the home-eNodeB coverage area, but also home-eNodeB interference to terminals close to the home-eNodeB but not being part of the home-eNodeB CSG.

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