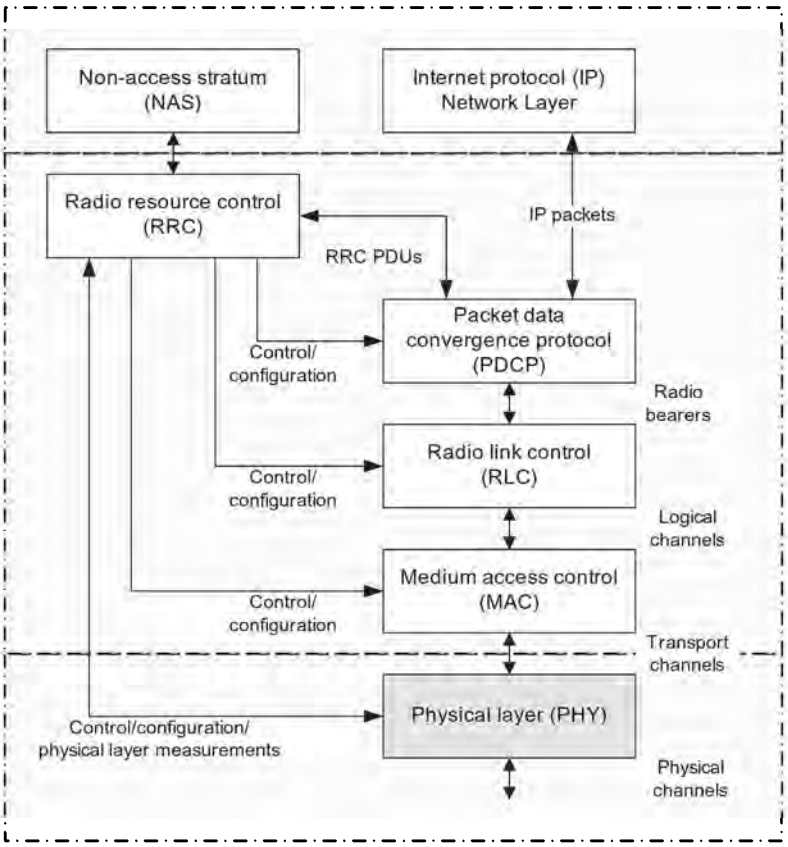


FIGURE 9.1

Physical Layer in LTE Protocol Stack

due to different design requirements and reliability and performance criteria, the procedures tend to be different. We will underline these differences and will describe the methods for selecting the best transmission format for each physical channel.

9.1 Overview of downlink physical layer processing

This section provides an overview of physical layer processing in LTE/LTE-Advanced downlink. The processing steps and the associated control signaling are depicted in Figure 9.2. The main criteria in the design of the LTE/LTE-Advanced physical layer were to increase the application throughput and capacity, reduce access latency, support higher user mobility, minimize intra-cell and inter-cell interference, improve reliability of control and data channel coverage especially

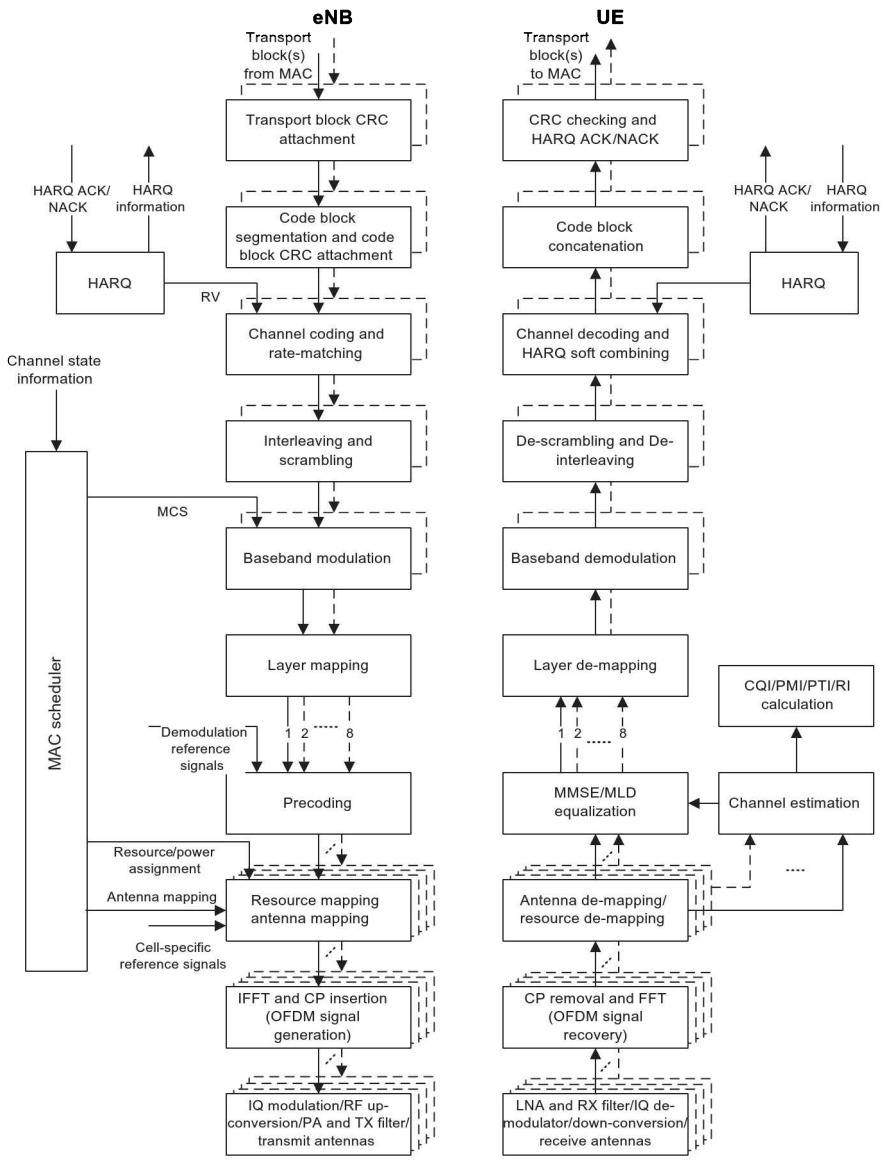


FIGURE 9.2

Downlink physical layer processing in LTE-Advanced [3,156].

at the cell-edge, and reduce the complexity and signaling overhead. The overall performance evaluations provided in Chapter 11 suggest that the desired goals have been successfully achieved. In the next sections of this chapter, we describe in detail the functional blocks and protocols of the downlink physical layer and their interactions based on the order that the information is processed as shown in Figure 9.2. It must be noted that although the overall procedure shown in Figure 9.2 is similar for LTE Rel-8/9 and LTE-Advanced Rel-10/11, the number of transmission layers, transmission modes, and multi-antenna processing are different. These differences will be identified in description of each function in the upcoming sections.

9.2 Characteristics of wireless channels

In a wireless communication system, a signal can travel from the transmitter to the receiver over multiple paths. This phenomenon is referred to as multipath propagation where signal attenuation varies on different paths. This effect, also known as multipath fading, can cause stochastic fluctuations in the received signal's magnitude, phase, and angle of arrival. The propagation over different paths is caused by scattering, reflection, diffraction, and refraction of the radio waves by stationary and moving objects, as well as the medium. It is obvious that different propagation mechanisms result in different channel and path loss models. As a result of wave propagation over multipath fading channels, the radio signal is attenuated due to mean path loss, as well as macroscopic and microscopic fading.

In an ideal free-space propagation model, the attenuation of RF signal energy between the transmitter and receiver follows the inverse-square law. The received power expressed in terms of transmitted power is attenuated proportional to the inverse of $L_s(d)$, which is called free space loss. When the receiving antenna is isotropic, the received signal power can be expressed as follows [19,48,49]:

$$P_{RX} = P_{TX} G_{TX} G_{RX} \left(\frac{\lambda}{4\pi d}\right)^2 = \frac{P_{TX} G_{TX} G_{RX}}{L_s(d)} \tag{9.1}$$

where P_{TX} and P_{RX} denote the transmitted and the received signal power, G_{TX} and G_{RX} denote the transmitting and receiving antenna gains, d is the distance between the transmitter and the receiver, and λ is the wavelength of the RF signal.

Macroscopic fading is caused by shadowing effects of buildings and natural obstructions, and is modeled by the local mean of a fast fading signal. The mean path loss $\overline{L}_p(d)$ as a function of distance d between the transmitter and receiver is proportional to the nth power of d relative to the reference distance d_0. In logarithmic scale, it can be expressed as follows:

$$\overline{L}_p(d) = L_s(d_0) + 10n \log(d/d_0) \quad \text{(dB)} \tag{9.2}$$

The reference distance d_0 corresponds to a point located in the far-field of the antenna typically 1 km for large cells, 100 m for micro-cells, and 1 m for indoor channels. In Eq. (9.2), $\overline{L}_p(d)$ is the mean path loss, which is typically $10n$ dB per decade attenuation for $d \gg d_0$. The value of n depends on the frequency, antenna heights, and propagation environment, and is equal to 2 in free space. Studies show that the path loss $L_p(d)$ is a random variable with log-normal distribution about the mean path loss $\overline{L}_p(d)$. Let $X \sim \mathbb{N}(0, \sigma^2)$ denote a zero-mean Gaussian random variable with standard deviation σ when measured in decibels, then:

$$L_p(d) = L_s(d_0) + 10n \log(d/d_0) + X \quad \text{(dB)} \tag{9.3}$$

The value of X is often derived empirically based on measurements. A typical value for σ is 8 dB. The parameters that statistically describe path loss due to large-scale fading (macroscopic fading) for an arbitrary location with a specific transmitting–receiving antenna separation include the reference distance d_0, the path loss exponent n, and the standard deviation σ of X [19,48,49].

Microscopic fading refers to the rapid fluctuations of the received signal in time and frequency, and is caused by scattering objects between the transmitting and receiving antennas. When the received RF signal is a superposition of independent scattered components plus a line-of-sight (LoS) component, the envelope of the received signal $r(t)$ has a Rician probability distribution function (PDF), and is referred to as Rician fading. As the magnitude of the LoS component approaches zero, the Rician PDF approaches a Rayleigh PDF. Thus:

$$f(r) = \frac{r(K+1)}{\sigma^2} \exp\left[-K - \frac{(K+1)r^2}{2\sigma^2}\right] I_0\left(\frac{2r}{\sigma}\sqrt{K(K+1)/2}\right) \quad r \geq 0 \tag{9.4}$$

where K and $I_0(r)$ denote the Rician factor and zero-order modified Bessel function of the first kind. In the absence of LoS path ($K = 0$), the Rician PDF reduces to Rayleigh distribution.

Time-varying fading due to scattering objects or transmitter/receiver motion results in Doppler spread. The time spreading effect of small-scale or microscopic fading is manifested in the time-domain as multipath delay spread, and in the frequency-domain as channel coherence bandwidth. Similarly, the time variation of the channel is characterized in the time-domain as channel coherence time, and in the frequency-domain as Doppler spread. In a fading channel, the relationship between maximum excess delay time τ_m and symbol time τ_s can be viewed in terms of two different degradation effects, i.e., frequency-selective fading and frequency non-selective or flat fading. A channel is said to exhibit frequency-selective fading if $\tau_m > \tau_s$. This condition occurs whenever the received multipath components of a symbol extend beyond the symbol's time duration. Such multipath dispersion of the signal results in inter-symbol interference (ISI) distortion. In the case of frequency-selective fading, mitigating the distortion is possible because many of the multipath components are separable by the receiver. A channel is said to

exhibit frequency non-selective or flat fading if $\tau_m < \tau_s$. In this case, all received multipath components of a symbol arrive within the symbol time duration; therefore, the components are not resolvable. In this case, there is no channel-induced ISI distortion, since the signal time spreading does not result in significant overlap among adjacent received symbols. There is still performance degradation, because the irresolvable phasor components can add up destructively to yield a substantial reduction in signal-to-noise ratio (SNR). Also, signals that are classified as exhibiting flat fading can sometimes experience frequency-selective distortion.

Figure 9.3 illustrates a multipath-intensity profile $\Lambda(\tau)$ versus delay τ where the term delay refers to the excess delay. It represents the signal's propagation delay that exceeds the delay of the first signal arriving at the receiver. For a typical wireless communication channel, the received signal usually consists of several discrete multipath components. The received signals are composed of a

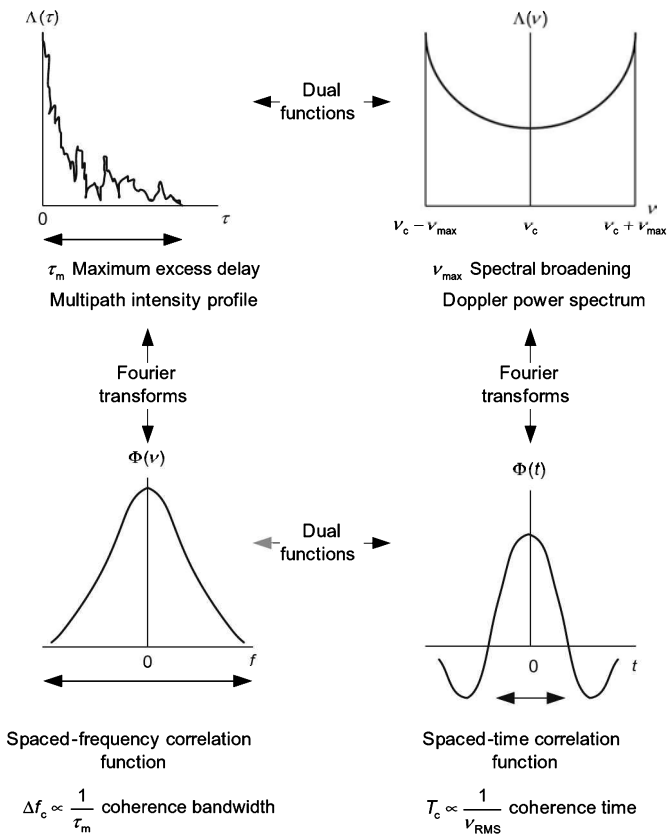


FIGURE 9.3

Illustration of the duality principle in time- and frequency-domains [19,48,49].

continuum of multipath components in some channels such as the tropospheric channel. In order to perform measurements of the multipath intensity profile, a wideband signal, i.e., a unit impulse or Dirac delta function, is used. For a single transmitted impulse, the time τ_m between the first and last received component is defined as the maximum excess delay during which the multipath signal power typically falls to some level 10–20 dB below that of the strongest component. Note that for an ideal system with zero excess delay, the function $\Lambda(\tau)$ would consist of an ideal impulse with weight equal to the total average received signal power. In the literature, the Fourier transform of $\Lambda(\tau)$ is referred to as the spaced-frequency correlation function $\Phi(\upsilon)$ [19]. The spaced-frequency correlation function $\Phi(\upsilon)$ is the channel's response to a pair of sinusoidal signals separated in frequency by υ. The coherence bandwidth Δf_c is a measure of the frequency range over which spectral components have a strong likelihood of amplitude correlation. In other words, a signal's spectral components over this range are affected by the channel in a similar manner. Note that Δf_c and τ_m are inversely proportional ($\Delta f_c \propto 1/\tau_m$). The maximum excess delay τ_m is not the best indicator of how any given wireless system will perform over a communication channel, because different channels with the same value of τ_m can exhibit different variations of signal intensity over the delay span. A useful measure of delay spread is most often characterized in terms of the root mean square (RMS) delay spread τ_{RMS} where [20]:

$$\tau_{RMS} = \sqrt{\frac{\int_0^{\tau_m} (\tau - \overline{\tau})^2 \Lambda(\tau) d\tau}{\int_0^{\tau_m} \Lambda(\tau) d\tau}} \tag{9.5}$$

in which the average multipath delay is calculated as follows:

$$\overline{\tau} = \frac{\int_0^{\tau_m} \tau \Lambda(\tau) d\tau}{\int_0^{\tau_m} \Lambda(\tau) d\tau} \tag{9.6}$$

An exact relationship between coherence bandwidth and delay spread does not exist and must be derived from signal analysis of actual signal dispersion measurements in specific channels. If coherence bandwidth is defined as the frequency interval over which the channel's complex-valued frequency transfer function has a correlation of at least 0.9, the coherence bandwidth is approximately $\Delta f_c \approx 1/(50\tau_{RMS})$. A common approximation of Δf_c corresponding to a frequency range over which the channel transfer function has a correlation of at least 0.5 is $\Delta f_c \approx 1/(5\tau_{RMS})$.

A channel is said to exhibit frequency-selective effects if $\Delta f_c < 1/\tau_s$, where the inverse symbol rate is approximately equal to the signal bandwidth W. In

practice, W may differ from $1/\tau_s$ due to filtering or data modulation. Frequency-selective fading effects arise whenever a signal's spectral components are not affected equally by the channel. This occurs whenever $\Delta f_c < W$. Frequency non-selective or flat fading degradation occurs whenever $\Delta f_c > W$. Hence, all of the signal's spectral components will be affected by the channel in a similar manner. Flat-fading does not introduce channel-induced ISI distortion, but performance degradation can still be expected due to loss in SNR whenever the signal experiences fading. In order to avoid channel-induced ISI distortion, the channel is required to exhibit flat-fading by ensuring that $\Delta f_c > W$. Therefore, the channel coherence bandwidth Δf_c sets an upper limit on the transmission rate that can be used without incorporating an equalizer in the receiver.

Figure 9.3 shows function $\Phi(t)$ which is known as the spaced-time correlation function, which is the autocorrelation function of the channel's response to a sinusoid. This function specifies the extent to which there is correlation between the channel's response to a sinusoid sent at time t_1 and the response to a similar sinusoid sent at time t_2, where $\Delta t = t_2 - t_1$. The coherence time is a measure of the expected time duration over which the channel's response is essentially invariant. To estimate $\Phi(t)$, a sinusoidal signal is transmitted through the channel and the autocorrelation function of the channel output is calculated. The function $\Phi(t)$ and the parameter T_c provide information about the speed of fading-channel variation. Note that for an ideal time-invariant channel, the channel's response would be highly correlated for all values of Δt and $\Phi(t)$ would be a constant function. If one ideally assumes uniformly distributed scattering around a mobile station with linearly polarized antennas, then the Doppler power spectrum (i.e., the inverse Fourier transform of spaced-time correlation function $\Lambda(\nu)$) has a U-shaped distribution, as shown in Figure 9.3. In a time-varying fading channel, the channel response to a pure sinusoidal tone spreads over a finite frequency range $\nu_c - \nu_{max} < \nu < \nu_c + \nu_{max}$, where ν_c and ν_{max} denote the frequency of the sinusoidal tone and the maximum Doppler spread, respectively. The RMS bandwidth of $\Lambda(\nu)$ is referred to as Doppler spread and is denoted by ν_{RMS} that can be calculated as follows:

$$\nu_{RMS} = \sqrt{\frac{\int_{\nu_c-\nu_{max}}^{\nu_c+\nu_{max}} (\nu-\overline{\nu})^2 \Lambda(\nu)\mathrm{d}\nu}{\int_{\nu_c-\nu_{max}}^{\nu_c+\nu_{max}} \Lambda(\nu)\mathrm{d}\nu}} \tag{9.7}$$

where:

$$\overline{\nu} = \frac{\int_{\nu_c-\nu_{max}}^{\nu_c+\nu_{max}} \nu\Lambda(\nu)\mathrm{d}\nu}{\int_{\nu_c-\nu_{max}}^{\nu_c+\nu_{max}} \Lambda(\nu)\mathrm{d}\nu} \tag{9.8}$$

The coherence time is typically defined as the time lag for which the signal autocorrelation coefficient reduces to 0.7. The coherence time is inversely proportional to Doppler spread $T_c \approx 1/\nu_{RMS}$. A common approximation for the value of coherence time as a function of Doppler spread is $T_c = 0.423/\nu_{RMS}$. It can be observed that the functions on the right-hand side of Figure 9.3 are dual of the functions on the left-hand side (duality principle).

The angle of spread refers to the spread in angle of arrival (AoA) of the multipath components at the receiver antenna array. At the transmitter, on the other hand, the angle of spread refers to the spread in the angle of departure (AoD) of the multipath components that leave the transmit antennas. If the angle spectrum function $\Theta(\theta)$ denotes the average power as a function of AoA, then the RMS angle spread can be calculated as follows:

$$\theta_{RMS} = \sqrt{\frac{\int_{-\pi}^{+\pi} (\theta - \bar{\theta})^2 \Theta(\theta) d\theta}{\int_{-\pi}^{+\pi} \Theta(\theta) d\theta}} \tag{9.9}$$

where:

$$\bar{\theta} = \frac{\int_{-\pi}^{+\pi} \theta \Theta(\theta) d\theta}{\int_{-\pi}^{+\pi} \Theta(\theta) d\theta} \tag{9.10}$$

The angle of spread causes space-selective fading, which manifests itself as variation of signal amplitude according to the location of antennas. The space-selective fading is characterized by the coherence distance D_c which is the spatial separation for which the autocorrelation coefficient of the spatial fading reduces to 0.7. The coherence distance is inversely proportional to the angle of spread $D_c \propto 1/\theta_{RMS}$.

In Figure 9.3, a duality between multipath intensity function $\Lambda(\tau)$ and Doppler power spectrum $\Lambda(\nu)$ is identified. It means that the two functions exhibit similar behavior across the time-domain and frequency-domain. As the $\Lambda(\tau)$ function identifies expected power of the received signal as a function of delay, $\Lambda(\nu)$ identifies expected power of the received signal as a function of frequency. Similarly, the spaced-frequency correlation function $\Phi(f)$ and spaced-time correlation function $\Phi(t)$ are dual functions. It implies that $\Phi(f)$ represents channel correlation in frequency-domain and $\Phi(t)$ corresponds to channel correlation function in time-domain in a similar manner.

Having reviewed some underlying concepts and characteristics of the wireless channels, let us now focus on some practical aspects of communication (data transmission) over wireless channels, and in particular the use of the above properties in the formation of a channel vector/matrix. As mentioned earlier, a wireless channel exhibits time, frequency, and space-selective variations known as fading. This fading arises due to Doppler, delay, and angle spreads in the scattering environment. The

channel spreading can be observed by sending a single impulse in frequency or continuous-wave signal in time or angle (point source) through the channel and receiving a signal spread along the spectral, temporal, or spatial dimension, respectively. A frequency-flat solution, however, can be applied to a frequency-selective channel by decomposing the transmission band into multiple narrow, frequency-flat sub-bands. Specifically, we can apply the solution per sub-carrier in systems employing OFDM. In a rich scattering environment, a frequency-flat MIMO wireless channel can be modeled as a complex Gaussian random process, represented as a time-varying matrix. The channel at each time instance is a Gaussian random variable, specified by the mean and its covariance. A non-zero channel mean signifies the presence of a direct LoS, or a cluster of strong paths, and the channel envelop has the Rician statistics, while zero mean corresponds to the Rayleigh statistics. The channel covariance, on the other hand, encompasses the correlation among the antennas at the transmitter and the receiver. Assuming the channel is stationary, the channel temporal variation can be captured by the channel auto-covariance, measuring the correlation between two channel instances separated by a delay. At zero delay, the channel auto-covariance coincides with the channel covariance [129].

Considering the channel state information at the transmitter (CSIT) at the transmit time in the form of a channel estimate and the estimation error covariance derived from a channel measurement at an initial time and the channel statistics, the main source of irreducible error in channel estimation is the random time variation of the channel between the initial measurement and its use by the transmitter. Therefore, we assume that the initial channel measurement is error free. The error in the channel estimate depends only on the time delay and the channel time selectivity, or the Doppler spread.

Let $\mathbf{H}$ denote the $N_{\mathrm{RX}} \times N_{\mathrm{TX}}$ channel matrix in a system with N_{TX} transmit and N_{RX} receive antennas. The channel has statistical mean $\overline{\mathbf{H}} = \mathscr{E}[\mathbf{H}]$ and covariance $\mathbf{R}_{\mathbf{H}} = \mathscr{E}[\mathbf{h}\mathbf{h}^H] - \overline{\mathbf{h}}\,\overline{\mathbf{h}}^H$, where the lower-case letter denotes the vectorized version of the upper-case matrix. Let us further assume that we have an initial, accurate channel measurement $\mathbf{H}(0)$. The channel auto-covariance $\mathbf{R}(\tau)$ at time delay τ then indicates the correlation between the initial measurement $\mathbf{H}(0)$ and the current channel $\mathbf{H}(\tau)$, defined as $\mathbf{R}(\tau) = \mathscr{E}[\mathbf{h}(\tau)\mathbf{h}^H(0)] - \overline{\mathbf{h}}(\tau)\overline{\mathbf{h}}^H(0)$. Intuitively, when this correlation is strong (i.e., $\mathbf{R}(\tau)$ is large when measured in a suitable norm) then $\mathbf{H}(0)$ is useful for estimating $\mathbf{H}(\tau)$. The strongest correlation is when the delay is zero (i.e., if $\tau \to 0$, then $\mathbf{R}(\tau) \to \mathbf{R}(0)$). In a scalar system, $\mathbf{R}(\tau)$ and $\mathbf{R}(0)$ reduce to scalars $r(\tau)$ and $r(0)$, respectively, and they are related through $r(\tau) = \rho(\tau)r(0)$, where $|\rho(\tau)| \leq 1$ is the temporal correlation coefficient [129].

An important assumption about channel temporal homogeneity is that the temporal correlation-coefficient $\rho(\tau)$ between any pair of transmit-antenna and receive-antenna is identical. This assumption is based on the premise that the channel temporal statistics can be expected to be the same for all antenna pairs. It is now possible to separate the temporal correlation from the spatial correlation in the channel auto-covariance as $\mathbf{R}(\tau) = \rho(\tau)\mathbf{R}(0)$. The temporal correlation $\rho(\tau)$ is a function of the time delay τ and the channel Doppler spread. As an example, in

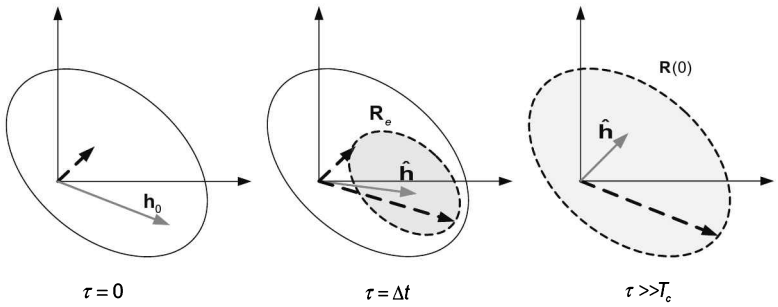


FIGURE 9.4

Illustration of a dynamic CSIT model [129].

Jake's model [21], $\rho(\tau) = J_0(2\pi\tau\nu)$, where ν is the channel Doppler spread and $J_0(\cdot)$ is the zero-order Bessel function of the first kind [35–37].

It can be shown that an estimate of the channel at time τ together with the estimation error covariance using minimum mean-squared error (MMSE) criterion results in the following $\hat{\mathbf{H}}(\tau) = \rho\mathbf{H}(0) - (1-\rho)\overline{\mathbf{H}}$ and $\mathbf{R}_e(\tau) = (1-\rho^2)\mathbf{R}(0)$. The two quantities $\hat{\mathbf{H}}(\tau)$ and $\mathbf{R}_e(\tau)$ function effectively as a new channel mean and a new channel covariance; thus they are referred to as the effective mean and the effective covariance, respectively, and constitute the CSIT. This CSIT ranges from perfect channel knowledge when $\rho = 1$, to pure stochastic when $\rho = 0$. Since the CSIT depends on the value of ρ which captures the channel time variation, it is called dynamic CSIT. The correlation factor ρ functions as a measure of CSIT quality. When $\rho = 1$, the channel estimate coincides with $\mathbf{H}(0)$ and is error free. As ρ approaches to zero, the effect of the initial channel measurement diminishes and the estimate asymptotically approaches the channel mean $\overline{\mathbf{H}}$. In parallel, the estimation error covariance $\mathbf{R}_e$ is zero when $\rho = 1$, and approaches zero as the correlation factor approaches zero. Figure 9.4 illustrates the CSIT variation as a function of the time delay τ. Several special cases of dynamic CSIT are of interest as follows: (1) perfect CSIT, in which the effective covariance is zero and the effective mean is the instantaneous channel matrix; (2) mean CSIT, in which the effective mean is non-zero and arbitrary, but the effective covariance is the identity matrix, corresponding to uncorrelated antennas; (3) covariance CSIT, in which the effective covariance matrix is non-identity and arbitrary, but the effective mean is zero, corresponding to Rayleigh fading. The general case in which both the mean and covariance matrices are arbitrary is referred to as statistical CSIT [129].

9.3 LTE/LTE-Advanced operating frequencies and band classes

The LTE/LTE-Advanced were designed to operate in the frequency bands defined in Table 9.1. The requirements were defined for 1.4, 3, 5, 10, 15, and 20 MHz

Table 9.1 E-UTRA Operating Frequencies, Band Classes, and Supported Bandwidths [1,2]

E-UTRA Operating Band (Band Class)	Uplink Operating Band eNB Receive/UE Transmit (MHz) $f_{UL\text{-}low} - f_{UL\text{-}high}$	$N_{Offset\text{-}DL}$	N_{DL}	Downlink Operating Band eNB Transmit/UE Receive (MHz) $f_{DL\text{-}low} - f_{DL\text{-}high}$	$N_{Offset\text{-}UL}$	N_{UL}	Duplex Scheme	Supported Bandwidths (MHz
1	1920–1980	0	0–599	2110–2170	18000	18000–18599	FDD	5, 10, 15, 20
2	1850–1910	600	600–1199	1930–1990	18600	18600–19199	FDD	1.4, 3, 5, 10, 15, 20
3	1710–1785	1200	1200–1949	1805–1880	19200	19200–19949	FDD	1.4, 3, 5, 10, 15, 20
4	1710–1755	1950	1950–2399	2110–2155	19950	19950–20399	FDD	1.4, 3, 5, 10, 15, 20
5	824–849	2400	2400–2649	869–894	20400	20400–20649	FDD	1.4, 3, 5, 10
6	830–840	2650	2650–2749	875–885	20650	20650–20749	FDD	5, 10
7	2500–2570	2750	2750–3449	2620–2690	20750	20750–21449	FDD	5, 10
8	880–915	3450	3450–3799	925–960	21450	21450–21799	FDD	1.4, 3, 5, 10
9	1749.9–1784.9	3800	3800–4149	1844.9–1879.9	21800	21800–22149	FDD	5, 10, 15, 20
10	1710–1770	4150	4150–4749	2110–2170	22150	22150–22749	FDD	5, 10, 15, 20
11	1427.9–1447.9	4750	4750–4949	1475.9–1495.9	22750	22750–22949	FDD	5, 10
12	699–716	5010	5010–5179	729–746	23010	23010–23179	FDD	1.4, 3, 5, 10
13	777–787	5180	5180–5279	746–756	23180	23180–23279	FDD	5, 10
14	788–798	5280	5280–5379	758–768	23280	23280–23379	FDD	5, 10
15	Reserved			Reserved			FDD	–
16	Reserved			Reserved			FDD	–
17	704–716	5730	5730–5849	734–746	23730	23730–23849	FDD	5, 10
18	815–830	5850	5850–5999	860–875	23850	23850–23999	FDD	5, 10, 15
19	830–845	6000	6000–6149	875–890	24000	24000–24149	FDD	5, 10, 15
20	832–862	6150	6150–6449	791–821	24150	24150–24449	FDD	5, 10, 15, 20

21	1447.9–1462.9	6450	6450–6599	1495.9–1510.9	24450	24450–24599	FDD	5, 10, 15
22	3410–3490	6600	6600–7399	3510–3590	24600	24600–25399	FDD	5, 10, 15, 20
23	2000–2020	7500	7500–7699	2180–2200	25500	25500–25699	FDD	1.4, 3, 5, 10
24	1626.5–1660.5	7700	7700–8039	1525–1559	25700	25700–26039	FDD	5, 10
25	1850–1915	8040	8040–8689	1930–1995	26040	26040–26689	FDD	1.4, 3, 5, 10, 15, 2
...				Not Specified				
33	1900–1920	36000	36000–36199	1900–1920	36000	36000–36199	TDD	5, 10, 15, 20
34	2010–2025	36200	36200–36349	2010–2025	36200	36200–36349	TDD	5, 10, 15
35	1850–1910	36350	36350–36949	1850–1910	36350	36350–36949	TDD	1.4, 3, 5, 10, 15, 2
36	1930–1990	36950	36950–37549	1930–1990	36950	36950–37549	TDD	1.4, 3, 5, 10, 15, 2
37	1910–1930	37550	37550–37749	1910–1930	37550	37550–37749	TDD	5, 10, 15, 20
38	2570–2620	37750	37750–38249	2570–2620	37750	37750–38249	TDD	5, 10, 15, 20
39	1880–1920	38250	38250–38649	1880–1920	38250	38250–38649	TDD	5, 10, 15, 20
40	2300–2400	38650	38650–39649	2300–2400	38650	38650–39649	TDD	5, 10, 15, 20
41	2496–2690	39650	39650 –41589	2496–2690	39650	39650 –41589	TDD	5, 10, 15, 20
42	3400–3600	41590	41590–43589	3400–3600	41590	41590–43589	TDD	5, 10, 15, 20
43	3600–3800	43590	43590–45589	3600–3800	43590	43590–45589	TDD	5, 10, 15, 20

bandwidth with a specific configuration in terms of the number of physical resource blocks (PRBs). Using carrier aggregation, a number of contiguous and/or non-contiguous frequency bands can be aggregated to create a virtually larger bandwidth. The channel raster is 100 kHz, which means the center frequency must be a multiple of 100 kHz [1]. To support transmission in paired and unpaired spectra, two duplexing schemes are supported, i.e., frequency division duplex (FDD), allowing both full and half-duplex terminal operation, as well as time division duplex (TDD). Table 9.2 gives the band classes where the LTE/LTE-Advanced systems can be deployed.

Figure 9.5 illustrates the relation between the channel bandwidth (BW_{Channel}) and the transmission bandwidth, i.e., the number of permissible PRBs (N_{RB}). The channel edges are defined as the lowest and highest frequencies of the carrier separated by the channel bandwidth; thus $f_C \pm (1/2)BW_{\text{Channel}}$. In the case of carrier aggregation, the aggregated channel bandwidth $BW_{\text{Channel-CA}}$ is defined as $BW_{\text{Channel-CA}} = f_{\text{edge-high}} - f_{\text{edge-low}}$. The lower band edge $f_{\text{edge-low}}$ and the upper band edge $f_{\text{edge-high}}$ of the aggregated channel bandwidth are used as the reference points in frequency for transmitter and receiver requirements, and are subsequently defined as follows: $f_{\text{edge-low}} = f_{\text{C-low}} - f_{\text{offset-low}}$ and $f_{\text{edge-high}} = f_{\text{C-high}} + f_{\text{offset-high}}$. The lower and upper frequency offsets depend on the transmission bandwidth configurations of the lowest and highest assigned edge component carrier, and are defined as $f_{\text{offset-low}} = 0.18N_{\text{RB-low}}/2 + BW_{\text{Guard-band}}$ and $f_{\text{offset-high}} = 0.18N_{\text{RB-high}}/2 + BW_{\text{Guard-band}}$, where $N_{\text{RB-low}}$ and $N_{\text{RB-high}}$ denote the transmission bandwidth configurations for the lowest and highest assigned component carrier, respectively, and $BW_{\text{Guard-band}}$ denotes the nominal guard band. Note that the 0.18 multiplier in the equations is the bandwidth of a PRB in MHz. The aggregated transmission bandwidth is defined as the number of the aggregated resource blocks within the entire assigned aggregated channel bandwidth, and is defined per carrier aggregation band class.

The frequency separation between carriers depends on the deployment scenario and the size of the frequency block available to the network operator, as well as the channel bandwidth. The nominal channel spacing between two adjacent E-UTRA carriers is defined as $\Delta f_{\text{spacing}} = (BW_{\text{Channel(1)}} + BW_{\text{Channel(2)}})/2$, where $BW_{\text{Channel(1)}}$ and $BW_{\text{Channel(2)}}$ parameters are the channel bandwidths of the two respective E-UTRA carriers. The channel spacing can be adjusted to optimize performance in a particular deployment scenario. For intra-band contiguous carrier aggregation bandwidth, the nominal channel spacing between two adjacent E-UTRA component carriers is specified as follows:

$$\Delta f_{\text{spacing-CA}} = 0.3 \left\lfloor \frac{BW_{\text{Channel(1)}} + BW_{\text{Channel(2)}} - 0.1\left|BW_{\text{Channel(1)}} - BW_{\text{Channel(2)}}\right|}{0.6} \right\rfloor \text{(MHz)} \tag{9.11}$$

The channel spacing for intra-band contiguous carrier aggregation can be adjusted to any integer multiple of 300 kHz (less than the nominal channel spacing)

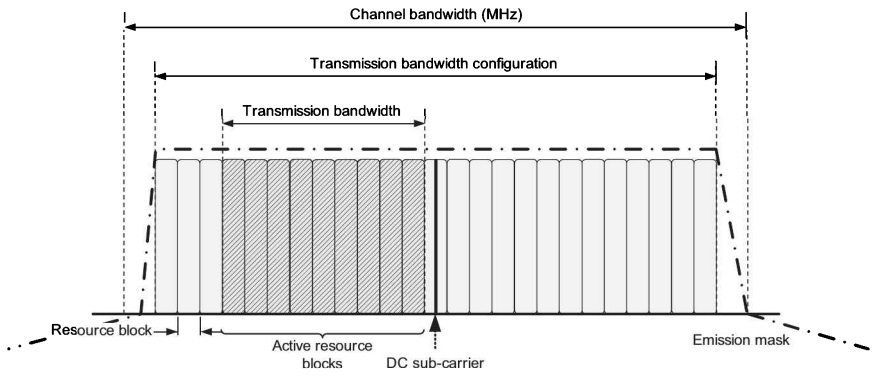


FIGURE 9.5

Relationship between channel bandwidth and transmission bandwidth in E-UTRA [1].

to optimize performance in a particular deployment scenario. The E-UTRA channel raster (including the carrier aggregation scenarios) is 100 kHz for all bands, which means that the carrier center frequency must be an integer multiple of 100 kHz.

The carrier frequency in the downlink/uplink directions is designated by the E-UTRA absolute radio frequency channel number (EARFCN) in the range of 0–65,535. The relation between EARFCN and the downlink carrier frequency (in MHz) is given by $f_{DL} = f_{DL-low} + 0.1(N_{DL} - N_{Offset-DL})$, where f_{DL-low}, N_{DL}, and $N_{Offset-DL}$ parameters are given in Table 9.1. Similarly, the relationship between EARFCN and the uplink carrier frequency (in MHz) for the uplink is defined as $f_{UL} = f_{UL-low} + 0.1(N_{UL} - N_{Offset-UL})$, where f_{UL-low}, N_{UL}, and $N_{Offset-UL}$ parameters are provided in Table 9.1. Note that N_{DL} and N_{UL} in the latter equations are the downlink and uplink EARFCN, respectively.

The channel numbers that designate carrier frequencies which are extremely close to the operating band edges are not used. This implies that the first 7, 15, 25, 50, 75, and 100 channel numbers at the lower operating band edge and the last 6, 14, 24, 49, 74, and 99 channel numbers at the upper operating band edge are not utilized for channel bandwidths of 1.4, 3, 5, 10, 15, and 20 MHz, respectively.

The fragmentation of the suitable and permissible frequency bands for deployment of LTE systems across the globe and the disparity of the operators' network deployments and legacy systems have made significant challenges against worldwide proliferation and accessibility of 4G networks (see Figure 9.6). Figure 9.6 depicts the distribution of the useable frequency bands for commercial LTE/LTE-Advanced deployments across the globe as of August 2012 [161]. In general, typical user equipment has to support multiple communication RF bands and standards (e.g., LTE/LTE-Advanced, UMTS, and 1xEV-DO) in order to allow operation across different regions and operator networks. Enabling the mobile terminals to cover all desired RF

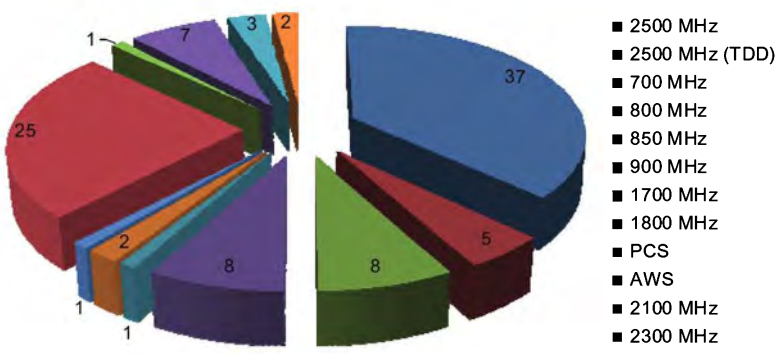


FIGURE 9.6

Distribution of frequency bands for E-UTRA commercial deployments across the world [161].

bands does not only result in an increased chip area due to more complex RF integrated circuit implementation, but also a larger phone form factor to accommodate external filters (see Figure 9.7 as an example of triple-band LTE transceiver). In the case of FDD systems, common handhelds use bulk acoustic wave duplex filters[1] to separate the receive signal from the transmit signal with a nominal frequency separation. These filters are not tunable, and thus require one duplexer per frequency constellation. The selectivity provided by the duplexer must reach a certain level such that third order inter-modulation frequency components are suppressed. These signals are generated by the transmit signal leaking into the receiver and an external interferer in a certain frequency band, thus degrading sensitivity of the entire receiver. Therefore, it is necessary to use either very expensive duplexers that could provide the desired selectivity, or to place an external (i.e., external to RFIC) low-noise amplifier followed by an additional surface acoustic wave filter[2] which is a more common solution. In general, it is not trivial to design extremely broadband RF filters, power or low-noise amplifiers, mixers, etc., to cover a number of contiguous and/or non-contiguous RF bands with optimal performance across the entire operating bands. Figure 9.8 shows the measured and simulated performance of a triple-fed

[1]Bulk acoustic wave filters are electro-mechanical devices typically operating in the frequency range of 2–16 GHz which can be implemented by ladder or lattice filters. Two main variants of bulk acoustic wave filter are thin film bulk acoustic resonator, or FBAR, and solid mounted bulk acoustic resonators [167].

[2]Surface acoustic wave filters are electro-mechanical devices commonly used in RF systems. The electrical signal is converted to a mechanical wave in a device constructed of a piezoelectric crystal or ceramic. This wave is delayed as it propagates across the device, before being converted back to an electrical signal by further electrodes. The delayed outputs are recombined to produce a direct analog implementation of a finite impulse response filter. This hybrid filtering technique has found applications in analog sampled filters. The surface acoustic wave filters are limited to frequencies up to 3 GHz [167].

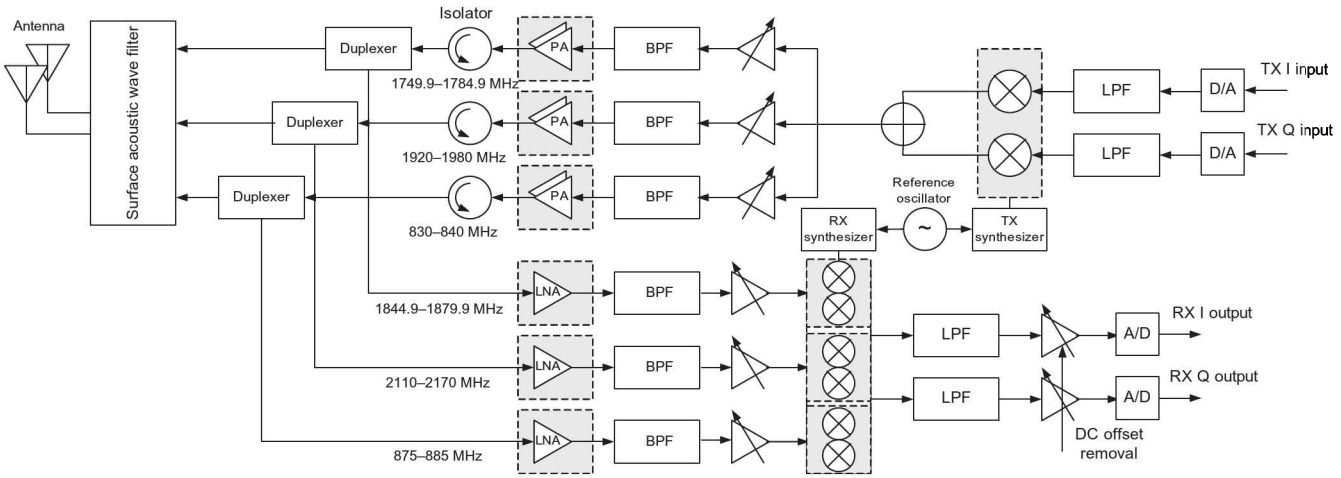


FIGURE 9.7

Example of a Triple-band LTE Transceiver [170].

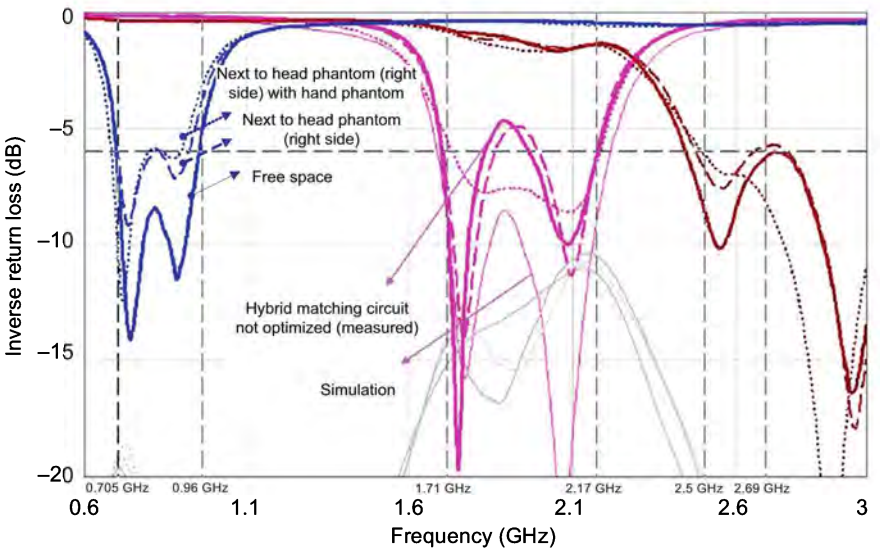


FIGURE 9.8

Simulated and measured inverse return loss of an example triple-fed antenna at 0.8, 1.9, 2.6 GHz [170].

antenna in terms of input return loss[3] (i.e., an indication of the quality of input impedance matching) under three different practical scenarios and in different frequency bands. It can be seen that satisfactory impedance matching is only achieved in a certain frequency region and not across the entire frequency bands, especially if the antennas are covered by hand or head phantoms. This example underlines the practical challenge of designing extremely wideband RF systems to fit into very small user-equipment's form factor in order to enable multi-band multi-mode LTE operation [170].

[3]Return loss is the loss of signal power resulting from the reflection caused at a discontinuity in a transmission media. This discontinuity can be a mismatch with the terminating load or with a device inserted in the line. It is usually expressed as the ratio of the incident power P_i and the reflected power P_r, i.e., RL (dB) = 10 log(P_i/P_r) in dB. The return loss is related to both the standing wave ratio (SWR) and reflection coefficient (Γ). Increasing the value of return loss corresponds to lower SWR. Return loss is a measure of how well the input/output impedances of devices or transmission lines are matched. A match is good, if the return loss value is high. A high return loss is desirable and results in a lower insertion loss. The ratio of the amplitude of the reflected wave V_r to the amplitude of the incident wave V_i is known as the reflection coefficient Γ, which is defined as $\Gamma = V_r/V_i$. When the source and load impedances are known values, the reflection coefficient is given by $\Gamma = (Z_L - Z_0)/(Z_L + Z_0)$ where Z_0 is the impedance of the transmission line toward the source and Z_L is the impedance toward the load. The return loss is the negative of the magnitude of the reflection coefficient in dB. Since power is proportional to the square of the voltage, the return loss is given by RL(dB) = $-$ 20 log(|Γ|) [169].

The performance of mobile device antennas is an important factor in the design and implementation of the device (in terms of practical form factor) as well as the ultimate communication quality that a user may expect from that device. This requires adopting an antenna design that yields high performance under realistic conditions in order to ensure satisfaction of LTE/LTE-Advanced system requirements. This challenge will be further intensified in the case of multi-band antenna array design which is required to fit into the small form factor of a mobile device and to provide satisfactory performance (near-field/far-field radiation patterns, gain, return loss, parasitic/spurious effects, etc.) over a large bandwidth.

9.4 Principles of OFDM

The orthogonal frequency division multiplexing (OFDM) is a form of multi-carrier modulation technique that was first introduced over 50 years ago. In a conventional serial data transmission system, the information bearing symbols are transmitted sequentially, with the frequency spectrum of each symbol occupying the entire available bandwidth. Figure 9.9 illustrates an unfiltered quadrature amplitude modulation (QAM) signal spectrum. It is in the form of $\sin(\pi f T_u)/\pi f T_u$ with zero crossing points at integer multiples of $1/T_\mathrm{u}$, where T_u is the QAM symbol period. The concept of OFDM is to transmit the data bits in parallel QAM modulated sub-carriers using frequency division multiplexing. The carrier spacing is carefully selected so that each sub-carrier is located on other sub-carriers' zero-crossing points in the frequency-domain. Although there are spectral overlaps among sub-carriers, they do not interfere with each other if they are sampled at the sub-carrier frequencies. In other words, they maintain spectral orthogonality. As shown in Figure 9.9, the OFDM signal in the frequency-domain is generated through aggregation of N_FFT parallel QAM-modulated sub-carriers, where adjacent sub-carriers, are separated by sub-carrier spacing $1/T_\mathrm{u}$. An example of OFDM signal spectrum as seen on a vector signal analyzer is shown in Figure 9.10.

Since an OFDM signal consists of many parallel QAM sub-carriers, the mathematical expression of the signal in the time-domain can be expressed as follows:

$$s(t) = \mathrm{Re}\left\{ e^{j\omega_c t} \sum_{k=-(N_\mathrm{FFT}-1)/2}^{(N_\mathrm{FFT}-1)/2} \alpha_k \, e^{j2\pi k(t-t_\mathrm{g})/T_\mathrm{u}} \right\} \quad mT_\mathrm{u} \le t \le (m+1)T_\mathrm{u} \qquad (9.12)$$

where $s(t)$ denotes the OFDM signal in the time-domain, α_k is the complex-valued data that is QAM-modulated and transmitted over sub-carrier k, N_FFT is the number of sub-carriers in the frequency-domain, ω_c is the RF carrier frequency, and T_g is the guard interval or the cyclic prefix (CP) length. For a large number of sub-carriers, direct generation and demodulation of the OFDM signal would require arrays of coherent sinusoidal generators which can become excessively complex and expensive. However, one can notice that the OFDM signal is actually the real part of the inverse discrete Fourier transform (IDFT) of the original complex-valued data

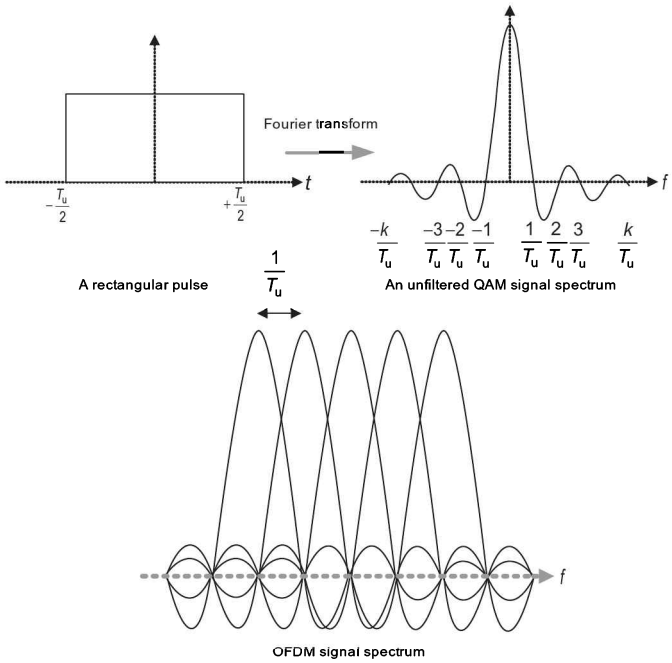


FIGURE 9.9

Illustration of OFDM concept [52].

symbols $\{\alpha_k\}|_{k=-(N_{\text{FFT}}-1)/2,\ldots,(N_{FFT}-1)/2}$. It can be seen that there are $N < N_{\text{FFT}}$ sub-carriers, each carrying the corresponding data α_k. The inverse of the sub-carrier spacing $\Delta f = 1/T_{\text{u}}$ is defined as the OFDM useful symbol duration T_{u}, which is N_{FFT} times longer than that of the original input data symbol duration.

Since IDFT is used in the OFDM modulator, the original data is defined in the frequency-domain, while the OFDM signal $s(t)$ is defined in the time-domain. The IDFT can be implemented via a computationally efficient fast Fourier transform (FFT) algorithm. The orthogonality of sub-carriers in OFDM can be maintained and individual sub-carriers can be completely separated and demodulated by an FFT at the receiver when there is no ISI introduced by the communication channel. In practice, linear distortions such as multipath delay cause ISI between OFDM symbols, resulting in loss of orthogonality and an effect that is similar to co-channel interference. However, when delay spread is small, i.e., within a fraction of the OFDM useful symbol length, the impact of ISI is insignificant, although it depends on the order of modulation implemented by the sub-carriers (see Figure 9.11). A simple solution to mitigate multipath delay is to increase the OFDM effective symbol duration such that it is much larger than the delay spread; however, when the delay spread is large, it requires a large number of sub-carriers and a large FFT size. Meanwhile, the system might become sensitive to Doppler shift and carrier frequency offset (CFO). An alternative approach to mitigate

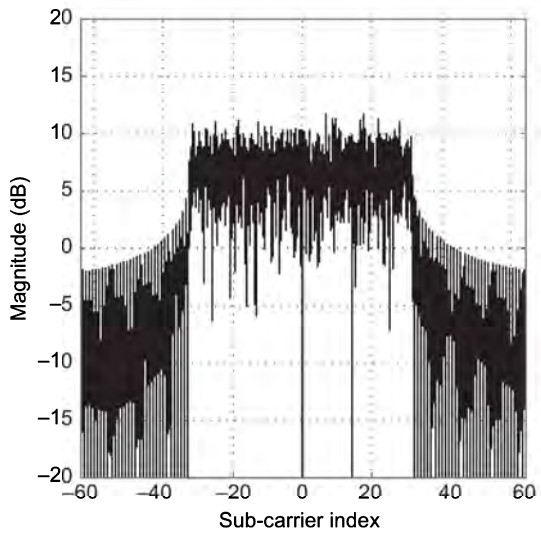


FIGURE 9.10

A snapshot of OFDM signal in the frequency-domain [158].

multipath distortion is to generate a cyclically extended guard interval, where each OFDM symbol is prefixed with a periodic extension of the signal itself, as shown in Figure 9.12 where the tail of the symbol is copied to the beginning of the symbol. The OFDM symbol duration is then defined as $T_s = T_u + T_g$, where T_g is the guard interval or cyclic prefix. When the guard interval is longer than the channel impulse response or the multipath delay, the ISI can be effectively eliminated. The ratio of the guard interval to useful OFDM symbol duration depends on the deployment scenario and the frequency band. Since the insertion of the guard intervals will reduce data throughput, T_g is usually selected less than $T_u/4$. The cyclic prefix should absorb most of the signal energy dispersed by the multipath channel. The entire ISI energy is contained within the cyclic prefix if its length is greater than that of the channel RMS delay spread ($T_g > \tau_{RMS}$). In general, it is sufficient to have most of the delay spread energy absorbed by the guard interval, considering the inherent robustness of large OFDM symbols to time dispersion.

The mapping of the modulated data symbol into multiple sub-carriers also allows an increase in the symbol duration. The symbol duration obtained through an OFDM scheme is much larger than that of a single carrier modulation technique with a similar transmission bandwidth. In general, when the channel delay spread exceeds the guard time, the energy contained in the ISI will be much smaller with respect to the useful OFDM symbol energy, as long as the symbol duration is much larger than the channel delay spread, i.e., $T_s \gg \tau_{RMS}$. Although large OFDM symbol duration is desirable to mitigate the ISI effects caused by time dispersion, large OFDM symbol duration can further reduce the ability to alleviate the effects of fast fading, particularly if the symbol period is large compared to the channel coherence

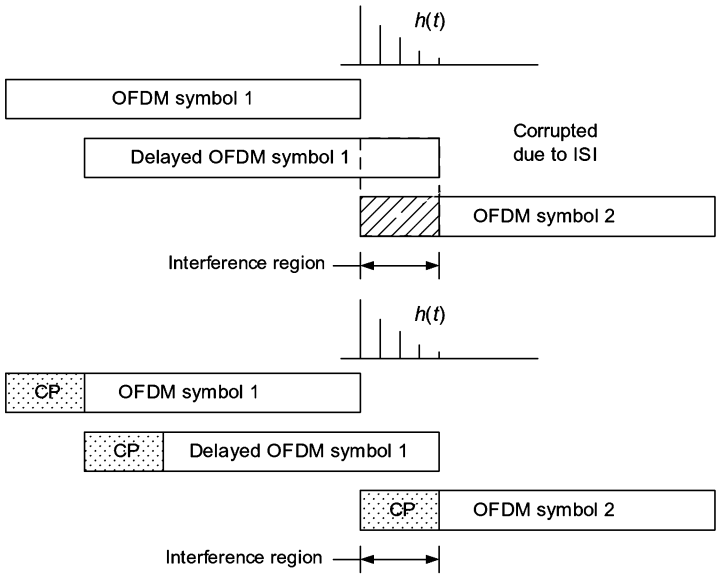


FIGURE 9.11

Illustration of the effect of guard interval for eliminating ISI ($h(t)$ is the hypothetical channel impulse response).

time, then the channel can no longer be considered as time-invariant over the OFDM symbol duration; therefore, this will introduce the inter-sub-carrier orthogonality loss. This can affect the performance in fast fading conditions. Hence, the symbol duration should be kept smaller than the minimum channel coherence time. Since the channel coherence time is inversely proportional to the maximum Doppler spread, the symbol duration T_s must, in general, be chosen such that $T_s \ll 1/\nu_{\text{RMS}}$. The large number of OFDM sub-carriers makes the bandwidth of the individual sub-carriers small relative to the overall signal bandwidth. With an adequate number of sub-carriers, the inter-carrier spacing is much smaller than the channel coherence bandwidth. Since the channel coherence bandwidth is inversely proportional to the channel delay spread τ_{RMS}, the sub-carrier separation is generally designed such that $\Delta f \ll 1/\tau_{\text{RMS}}$. In this case, the fading on each sub-carrier is flat and can be modeled as a complex-valued constant channel gain. The individual reception of the QAM symbols transmitted on each sub-carrier is therefore simplified to the case of a flat-fading channel. This enables a straightforward introduction of advanced MIMO schemes. Furthermore, in order to mitigate Doppler spread effects, the inter-carrier spacing should be much larger than the RMS Doppler spread $\Delta f \gg \nu_{\text{RMS}}$. Since the OFDM sampling frequency is typically larger than the actual signal bandwidth, only a subset of sub-carriers are used to carry QAM symbols. The remaining sub-carriers are left inactive prior to the IFFT, and are referred to as guard sub-carriers. The split between the active and the inactive

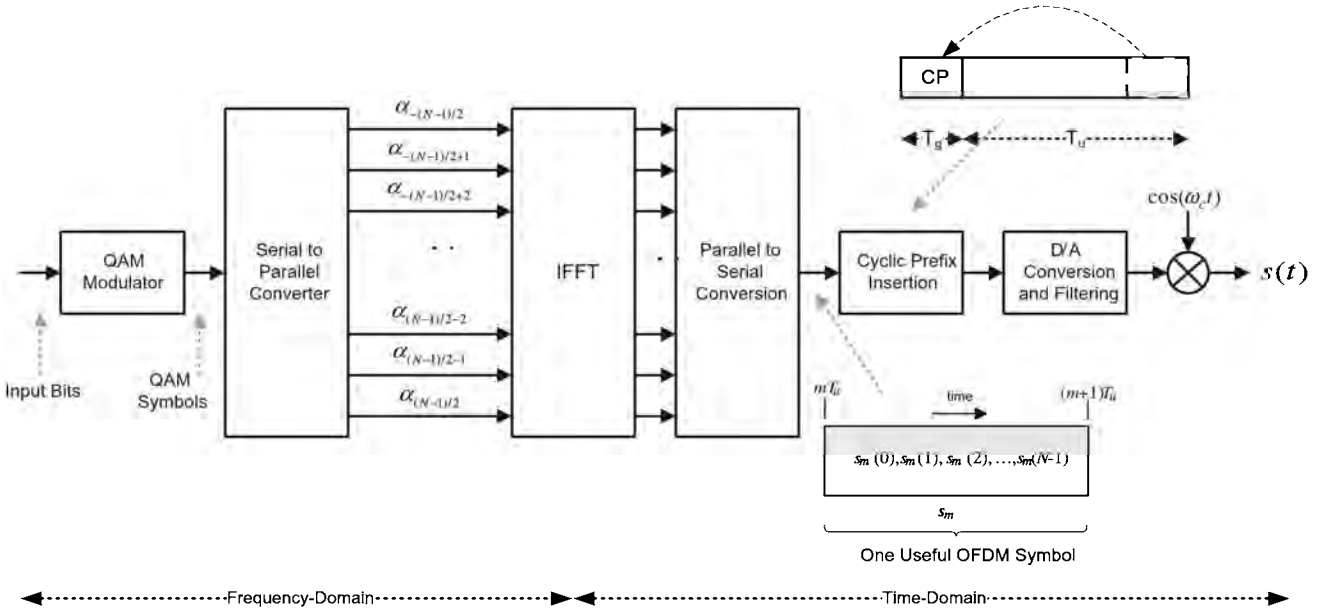


FIGURE 9.12

OFDM signal generation.

sub-carriers is determined based on the spectral sharing and regulatory constraints, such as the bandwidth allocation and the spectral mask.

An OFDM transmitter diagram is shown in Figure 9.12. The incoming bit stream is first QAM modulated to form the complex-valued QAM symbols. The QAM symbols are converted from serial to parallel with $N < N_{\text{FFT}}$ complex-valued numbers per block, where N_{FFT} is the size of FFT/IFFT operation. Each block is processed by an IFFT and the output of the IFFT forms an OFDM symbol, which is converted back to serial data for transmission. A guard interval or cyclic prefix is inserted between symbols to eliminate ISI effects caused by multipath distortion. The discrete symbols are filtered and converted to analog for RF up-conversion.

The reverse process is performed at the receiver. A one-tap equalizer is usually used for each sub-carrier to correct channel distortion. The tap coefficients are calculated based on channel information. As shown in Figures 9.9 and 9.10, one can understand that the OFDM signal spectrum is approximately rectangular. Since the OFDM modulator is an IFFT processor, its role is to convert data from the frequency-domain to the time-domain, and then transmit them in the time-domain over the channel. If a spectrum analyzer, which converts the input from the time-domain to the frequency-domain, is used to monitor an OFDM signal, what will be displayed is the original data.

When there is multipath distortion, a conventional single carrier wideband transmission system suffers from frequency selective fading. A complex adaptive equalizer must be used to equalize the in-band fading. The number of taps required for the equalizer is proportional to the symbol rate and the multipath delay. For an OFDM system, if the guard interval is larger than the multipath delay, the ISI can be eliminated and orthogonality can be maintained among sub-carriers. Since each OFDM sub-carrier occupies a very narrow spectrum, in the order of a few kHz, even under severe multipath distortion, they are only subject to flat fading. In other words, the OFDM converts a wideband frequency-selective fading channel to a series of narrowband frequency non-selective fading sub-channels by using the parallel multi-carrier transmission scheme (Figure 9.13). Since OFDM data sub-carriers are statistically independent and identically distributed, based on the central limit theorem, when the number of sub-carriers N_{FFT} is large, the OFDM signal distribution tends to be Gaussian.

In an OFDM system, the SNR is a measure of channel quality and is a key factor of link-level error assessment. There are different methods for calculation of SNR in single-antenna and multi-antenna transmission systems. For single-input/single-output systems, the SNR can be viewed as the received SNR, i.e., the received SNR before the detector. The post-processing SNR is often used for MIMO links, and represents the SNR after combining in the receiver and measures the likelihood that a coded message is decoded successfully. In link-level simulations, the SNR γ is typically calculated using the following method. Let vector $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{\text{TX}}})^H \in \mathbb{C}^{N_{\text{TX}} \times 1}$ denote the transmit signal, where $x_k \in \mathbb{C} \forall k = 1, 2, \ldots, N_{\text{TX}}$ is the complex-valued transmitted symbols from the kth transmit antenna. Note that N_{TX} is the number of transmit antennas. It can be shown that the total transmit signal power can be obtained as $\sigma_{\mathbf{x}}^2 = \text{trace}(\mathbf{R}_{\mathbf{x}})$,

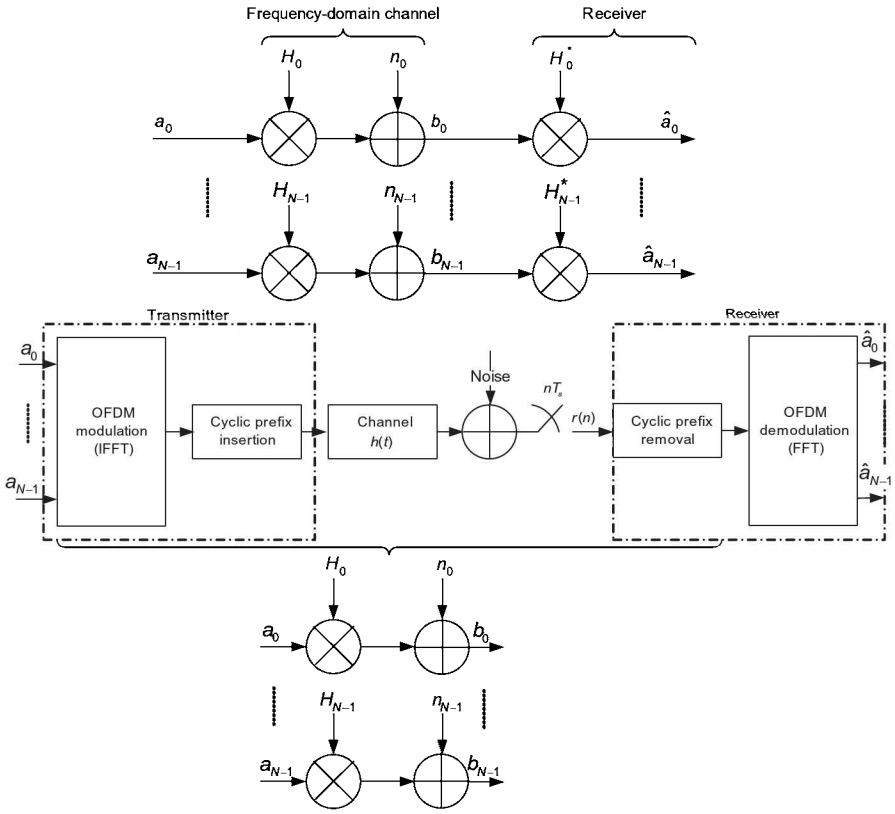


FIGURE 9.13

Single-tap equalization of the OFDM signal [14].

where $\mathbf{R_x}$ denotes the autocorrelation matrix of the transmitted signal. The transmit power from the kth antenna is given as $\sigma_{x_k}^2 = \mathscr{E}\{|x_k|^2\} = 1/N_{\text{TX}}$ (uniformly distributed power). If $\mathbf{H}$ represents the channel matrix with $||\mathbf{H}||_F^2 = N_{\text{TX}}N_{\text{RX}}$[4] in which N_{TX} and N_{RX} denote the number of transmit and receive antennas, respectively, the received signal vector $\mathbf{y}$ can be calculated as $\mathbf{y} = \mathbf{Hx} + \mathbf{v}$ where the complex-valued Gaussian-distributed noise vector $\mathbf{v} \sim \mathbb{Z}(\mathbf{0}, \sigma_{\mathbf{v}}^2\mathbf{I})$ denotes the noise vector with respect

[4] The p-norm of matrix $\mathbf{H}$ is defined as $\|\mathbf{H}\|_p = \left(\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}|h_{ij}|^p\right)^{1/p}$. In the special case when $p = 2$, the norm is called the Frobenius norm and for $p = \infty$ is called the maximum norm. The Frobenius norm or Hilbert–Schmidt norm of matrix $\mathbf{H}$ are similar, although the latter term is often reserved for the operators on a Hilbert space. In general, this norm can be defined in the following forms: $\|\mathbf{H}\|_p = \sqrt{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}|h_{ij}|^2} = \sqrt{\text{trace}(\mathbf{H}^H\mathbf{H})} = \sqrt{\sum_{i=0}^{\min(M,N)-1}\zeta_i^2}$ where $\mathbf{H}^H$ denotes the conjugate transpose of matrix $\mathbf{H}$ and ζ_i are the singular values of matrix $\mathbf{H}$. The Frobenius norm is further similar to the Euclidean norm on $\mathbb{R}^N$ and is obtained from an inner product on the space of all matrices. The Frobenius norm is sub-multiplicative and is very useful for numerical linear algebra [31,32,34].

to the size of the FFT N_{FFT} and the number of used sub-carriers before the detector N_{Used}. We define the complex-valued Gaussian-distributed noise vector $\mathbf{n} \sim \mathbb{Z}(\mathbf{0}, \sigma_{\mathbf{n}}^2 \mathbf{I})$ to be the noise after the FFT operation. The receive SNR before the detector is given as $\gamma_{\text{pre-FFT}} = \frac{||\mathbf{Hx}||_F^2}{N_{\text{RX}} \sigma_{\mathbf{v}}^2} = \sigma_{\mathbf{v}}^{-2}$ whereas the SNR after the FFT operation is given as $\gamma_{\text{post-FFT}} = \frac{||\mathbf{Hx}||_F^2}{N_{\text{RX}} \sigma_{\mathbf{n}}^2} = \sigma_{\mathbf{n}}^{-2}$. It can be seen that the ratio of pre-FFT and post-FFT SNRs $\frac{\gamma_{\text{pre-FFT}}}{\gamma_{\text{post-FFT}}} = \frac{\sigma_{\mathbf{v}}^2}{\sigma_{\mathbf{n}}^2} = \frac{N_{\text{FFT}}}{N_{\text{Used}}}$; $N_{\text{Used}} \leq N_{\text{FFT}}$ is always positive, implying that the FFT operation suppresses the noise and enhances the SNR [142].

9.4.1 Peak-to-average power ratio of OFDM signals

The peak-to-average power ratio (PAPR) for a single-carrier modulation signal depends on its constellation and the pulse shaping filter roll-off factor. For a Gaussian distributed OFDM signal, the cumulative distribution function (CDF)[5] of PAPR for 99.0%, 99.9%, and 99.99% are approximately 8.3, 10.3, and 11.8 dB, respectively. Since the OFDM signal has a high PAPR, it could be clipped in the transmitter power amplifier (PA), because of its limited dynamic range or non-linearity. Higher output back-off is required to prevent performance degradation and inter-modulation products leaking into adjacent channels. Therefore, RF PAs should be operated in a very large linear region. Otherwise, the signal peaks leak into the non-linear region of the PA causing signal distortion. This signal distortion introduces inter-modulation among the sub-carriers and out-of-band emission. Thus, the PAs should be operated with large power back-offs. On the other hand, this leads to very inefficient amplification and expensive transmitters. Thus, it is highly desirable to reduce the PAPR. In addition to inefficient operation of the PA, a high PAPR requires a larger dynamic range for the receiver analog-to-digital (A/D) converter. To reduce the PAPR, several techniques have been proposed and used such as clipping,[6] channel coding, temporal windowing, tone reservation,[7] and tone injection. However,

[5]The CDF of the real-valued random variable X is defined as $x \rightarrow F_X(x) = P(X \leq x), \forall x \in \mathbb{R}$, where the right-hand side represents the probability that random variable X takes on a real value less than or equal to x. The CDF of X can be defined in terms of the probability density function $f(x)$ as $F(x) = \int_{-\infty}^{x} f(x)\text{d}x$. The complementary CDF (CCDF), on the other hand, is defined as $P(X > x) = 1 - F_X(x)$ [30].

[6]Since the OFDM signal has a high PAPR, it may be clipped in the transmitter power amplifier, because of its limited dynamic range or non-linearity. Higher output back-off is required to prevent BER degradation and intermodulation products spilling into adjacent channels. However, clipping of an OFDM signal has a similar effect to impulse interference, against which an OFDM system is inherently robust. Computer simulations show that for a coded OFDM system, clipping of 0.5% of the time results in a BER degradation of 0.2 dB. At 0.1% clipping, the degradation is less than 0.1 dB [52].

[7]In the tone reservation method, the transmitter and the receiver reserve a subset of tones or sub-carriers for generating PAPR reduction signals. Those reserved tones are not used for data transmission [53].

most of these methods are unable to achieve simultaneously a large reduction in PAPR with low complexity, low coding overhead, without performance degradation, and without transmitter-receiver symbol handshake. The PAPR ξ of the OFDM signal, defined in Eq. (9.12), is calculated as follows:

$$\xi = \left. \frac{\max |s(t)|^2}{\mathscr{E}\{|s(t)|^2\}} \right|_{mT_u \le t \le (m+1)T_u} \tag{9.13}$$

In the above equation, $\mathscr{E}\{.\}$ denotes the expectation operator, and m is an integer. From the central-limit theorem, for large values of N_{FFT}, the real and imaginary values of $s(t)$ (from Eq. (9.12) and without the $\mathrm{Re}\{\cdot\}$ operator) would have Gaussian distribution. Consequently, the amplitude of the OFDM signal has a Rayleigh distribution with zero mean and a variance of N_{FFT} times the variance of one complex sinusoid. Assuming the samples to be mutually uncorrelated, the CDF for the peak power per OFDM symbol is given by:

$$P(\xi > \gamma) = [1 - (1 - e^{-\gamma})^{N_{FFT}}] \tag{9.14}$$

From the above equation, it can be seen that a large PAPR occurs only infrequently due to relatively large values of N_{FFT} used in practice.

Figure 9.14 shows the CDF of OFDM PAPR for BPSK and QPSK modulation assuming a 40 MHz channel bandwidth and $N_{FFT} = 128$. We further applied a 90° phase rotation to the sub-carriers in the upper 20 MHz of the

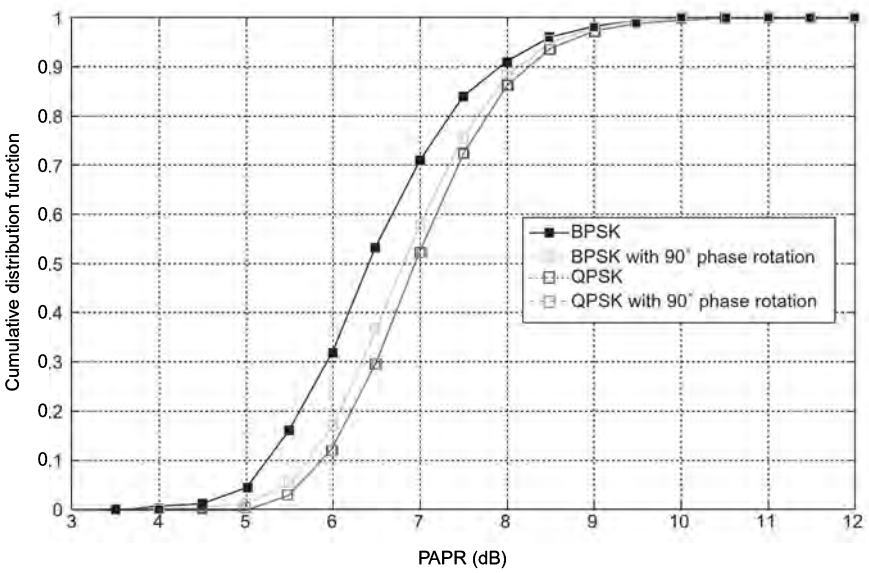


FIGURE 9.14

CDF of OFDM PAPR with BPSK/QPSK modulation and $N_{FFT} = 128$ [159].

channel, and investigated the effect on the PAPR reduction [159]. It is shown that large PAPR values are less likely to occur with large FFT sizes, as suggested by Eq. (9.14).

9.4.2 Effect of non-linearity on OFDM signals

Power amplifiers generally have a non-linear amplitude response, where the output power is saturated for large input powers. Most applications require operation in the linear region of the PA response where the output power is a linear function of the input. The larger the linear operation region, or alternatively the higher the saturation point, the more expensive the PA. Therefore, it is imperative to reduce the PAPR of the OFDM signal before processing through the PA. Figures 9.15 and 9.16 illustrate examples of PA AM/AM response and QAM signal constellation fuzziness when the OFDM signal with large PAPR is processed through a PA with a low saturation point (Figure 9.16). In this example, 64QAM modulation, 256-point IFFT with $T_{cp} = (1/4)T_u$, and PA saturation point of 6 dB are assumed.

The modulation accuracy or the permissible signal constellation fuzziness is often measured in terms of an error vector magnitude (EVM) metric. The EVM may be defined as the square root of the ratio of the mean error vector power to

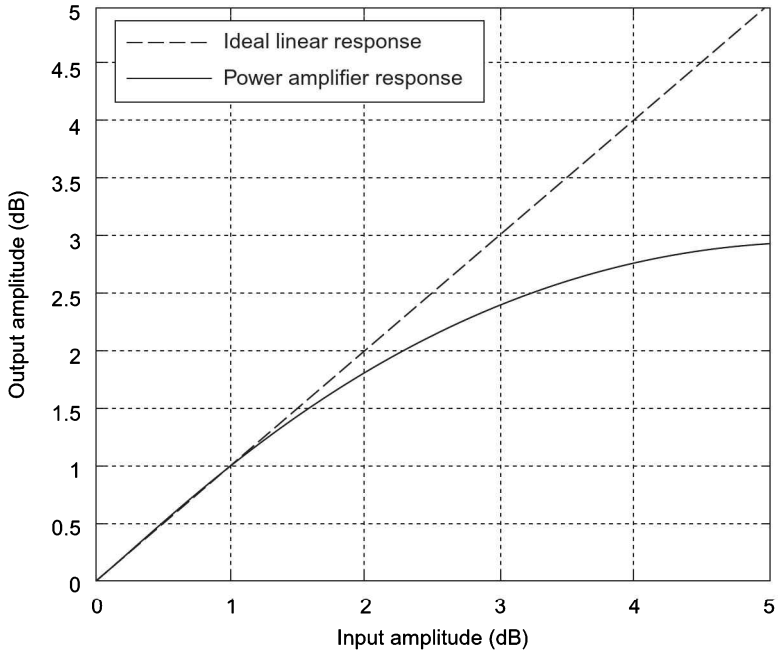


FIGURE 9.15

Example AM/AM response of a power amplifier (saturation point = 6 dB) [141].

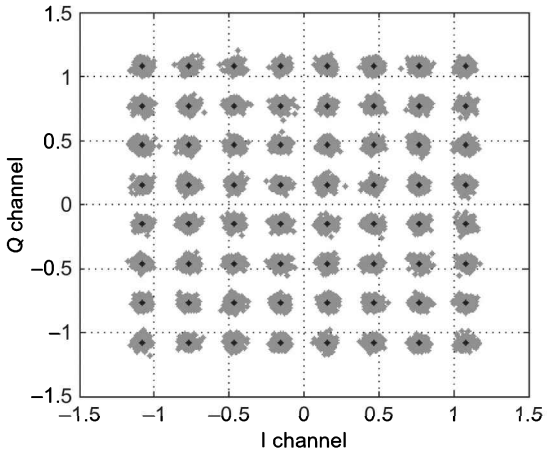


FIGURE 9.16

64QAM signal constellation before and after power amplification [141].

the mean reference signal power expressed as a percentage, i.e., the EVM can be expressed as $\mathrm{EVM} = \sqrt{\mathscr{E}|\mathbf{e}|^2/\mathscr{E}|\mathbf{s}|^2}$ in mathematical form. In other words, the EVM defines the average constellation error with respect to the farthest constellation point (i.e., the distance between the reference signal and measured signal points in the $I-Q$ plane), and is calculated as follows [1]:

$$\mathrm{EVM} = \sqrt{\frac{\sum_{i=0}^{N_{\mathrm{samples}}-1} (\Delta I_i^2 + \Delta Q_i^2)}{N_{\mathrm{samples}} r_{\mathrm{max}}^2}} \tag{9.15}$$

where N_{samples} is the number of symbols used in the measurement period, ΔI_i and ΔQ_i refer to the in-phase (I) and quadrature (Q) components of the ith error vector, and r_{max} denotes the maximum constellation amplitude. The EVM is measured over the continuous portion of a burst occupying at least one-fourth of the transmission frame at maximum power settings. The permissible EVM can be estimated from the transmitter implementation margin, if the error vector is considered noise, which is added to the channel noise. The implementation margin is the excess power needed to maintain the carrier-to-noise ratio (C/N) intact, when going from the ideal to the realistic transmitter. The EVM cannot be measured at the antenna connector, but should be measured by an ideal receiver with a certain carrier recovery loop bandwidth specified as a percentage of the symbol rate. The measured EVM includes the effects of the transmitter filter accuracy, D/A converter, modulator imbalances, untracked phase noise, and PA non-linearity.

As mentioned earlier, the EVM is a measure of the difference between the reference waveform and the measured waveform. In practice, before calculating the EVM, the measured waveform is corrected by the sample timing offset and RF frequency offset. Then the IQ origin offset is removed from the measured waveform before calculating the EVM. The measured waveform is further modified by selecting the absolute phase and absolute amplitude of the transmit chain. The EVM value is defined after the front-end IDFT as the square root of the ratio of the mean error vector power to the mean reference signal power expressed as a percentage value.

The basic EVM measurement interval in the time-domain depends on the type of physical channel. The measurement interval is equal to the length of one preamble sequence for the physical random access channel (PRACH), and is equal to the length of one slot for the physical uplink control channel (PUCCH) and physical uplink shared channel (PUSCH) in the time-domain. When the PUSCH or PUCCH transmission slot is shortened due to multiplexing with Sounding Reference Signal (SRS), the EVM measurement interval is accordingly reduced by one symbol. The PUSCH or PUCCH EVM measurement interval is also reduced when the mean power, modulation, or allocation between slots is expected to change. In the case of PUSCH transmission, the measurement interval is reduced by a time interval equal to the sum of 5 μs and the applicable exclusion period adjacent to the point where the power change is expected to occur. The PUSCH exclusion period is applied to the signal obtained after the front-end IDFT. In the case of PUCCH transmission with power change, the PUCCH EVM measurement interval is reduced by one symbol adjacent to the point where the power change is expected to occur [1].

9.4.3 Effect of carrier frequency offset on OFDM signals

As mentioned earlier, an OFDM system transmits information as a series of OFDM symbols. The time-domain samples $x_m(n)$ of the mth OFDM symbol are generated by performing IDFT on the information symbols $X_m(k)|_{k=0,1,\ldots,N_{\text{FFT}}-1}$, as follows [132]:

$$x_m(n) = \frac{1}{N_{\text{FFT}}} \sum_{k=0}^{N_{\text{FFT}}-1} X_m(k) e^{j2\pi k(n-N_{\text{CP}})/N_{\text{FFT}}}; \quad \forall 0 \le n \le N_{\text{FFT}} + N_{\text{CP}} - 1 \qquad (9.16)$$

where N_{FFT} and N_{CP} are the number of data samples and cyclic prefix samples, respectively. The OFDM symbol $x_m(n)$ is transmitted through a channel $h_m(n)$ and is perturbed by a Gaussian noise $z_m(n)$. The channel $h_m(n)$ is assumed to be block stationary, i.e., it is time invariant over each OFDM symbol. With this assumption, the output $y_m(n)$ of the channel can be represented as $y_m(n) = h_m(n)^* x_m(n) + z_m(n)$, where $h_m(n)^* x_m(n) = \sum_{k=-\infty}^{+\infty} h_m(k) x_m(k-n)$ and $z_m(n)$ is a zero-mean additive white Gaussian noise (AWGN) with variance σ_z^2.

Since the channel impulse response $h_m(n)$ is assumed to be block stationary, the channel response does not change over each OFDM symbol; however, the channel response $h_m(n)$ may vary across different OFDM symbols, thus it is a function of the OFDM symbol index m.

When the receiver oscillator is not perfectly synchronized to the transmitter oscillator, there can be a CFO $\Delta f_{\text{CFO}} = f_{\text{TX}} - f_{\text{RX}}$ between the transmitter carrier frequency f_{TX} and the receiver carrier frequency f_{RX}. Furthermore, there may be a phase offset θ_0 between the transmitter carrier and the receiver carrier. The mth received symbol $y_m(n)$ can be represented as $y_m(n) = e^{j\{2\pi\Delta f_{\text{CFO}}[n+m(N_{\text{FFT}}+N_{\text{CP}})]T_s+\theta_0\}}[h_m(n)^* x_m(n)] + z_m(n)$, where T_s is the sampling period. The CFO Δf_{CFO} can be represented relative to the sub-carrier bandwidth $(1/N_{\text{FFT}}T_s)$ by defining the relative frequency offset $\delta_{\text{CFO}} = \Delta f_{\text{CFO}} N_{\text{FFT}} T_s$. The CFO attenuates the desired signal and introduces the ICI, thus decreasing the SNR. The SNR of the kth sub-carrier can be expressed as $\text{SNR}_k(\delta_{\text{CFO}}) = (\phi^2_{N_{\text{FFT}}}(\delta_{\text{CFO}})P_h\sigma_x^2/(1 - \phi^2_{N_{\text{FFT}}}(\delta_{\text{CFO}}))P_h\sigma_x^2 + \sigma_z^2)$, where $\phi_{N_{\text{FFT}}}(\delta_{\text{CFO}}) = (\sin(\pi\delta_{\text{CFO}})/N_{\text{FFT}}\sin(\pi\delta_{\text{CFO}}/N_{\text{FFT}}))$ in order to demonstrate the dependence of the SNR on the frequency offset [132]. In the latter equation, P_h, σ_x^2, and σ_z^2 denote the total average power of channel impulse response, variance of the signal, and variance of the additive noise, respectively. The sub-carrier index k is dropped since the SNR is the same for all sub-carriers. From this SNR expression, it is clearly seen that the effect of the frequency offset is to decrease the signal power by $\phi^2_{N_{\text{FFT}}}(\delta_{\text{CFO}})$ and to convert the decreased power to interference power. The SNR depends not only on the frequency offset δ_{CFO}, but also on the number of sub-carriers; however, as N_{FFT} increases $\phi^2_{N_{\text{FFT}}}(\delta_{\text{CFO}})$ converges to $\text{sinc}^2(\delta_{\text{CFO}})$. Therefore, the SNR converges to $\text{SNR}(\delta_{\text{CFO}}) = (\text{sinc}^2(\delta_{\text{CFO}})P_h\sigma_x^2/(1 - \text{sinc}^2(\delta_{\text{CFO}}))P_h\sigma_x^2 + \sigma_z^2)$ as N_{FFT} becomes increasingly large [132]. The effect of CFO on OFDM signal in the frequency-domain, which appears as inter-carrier interference, for different values of δ_{CFO} is shown in Figure 9.17. Figure 9.18 illustrates the variation of $\text{SNR}(\delta_{\text{CFO}})$ as a function of δ_{CFO} with $\gamma = P_h\sigma_x^2/\sigma_z^2$ as a parameter. In the above equations, the power of inter-carrier interference as a function of relative CFO is defined as $P_{\text{ICI}}(\delta_{\text{CFO}}) = (1 - \text{sinc}^2(\delta_{\text{CFO}}))P_h\sigma_x^2$. In practice, the sub-carrier spacing is not the same among different tones due to mismatched oscillators (i.e., frequency offset), Doppler shift, and timing synchronization errors, resulting in inter-carrier interference and loss of orthogonality. It can be seen that the ICI increases with the increase of the OFDM symbol duration (or alternatively decrease of sub-carrier spacing) and the frequency offset. The effects of timing offset are typically less than that of the frequency offset, provided that the cyclic prefix is sufficiently large.

It can be shown that the ICI power can be calculated as a function of the generic Doppler power spectrum $\Lambda(\nu)$ as follows [14]:

$$P_{\text{ICI}} = \int_{-\nu_{\max}}^{\nu_{\max}} \Lambda(\nu)(1 - \text{sinc}^2(T_s\nu))\text{d}\nu \tag{9.17}$$

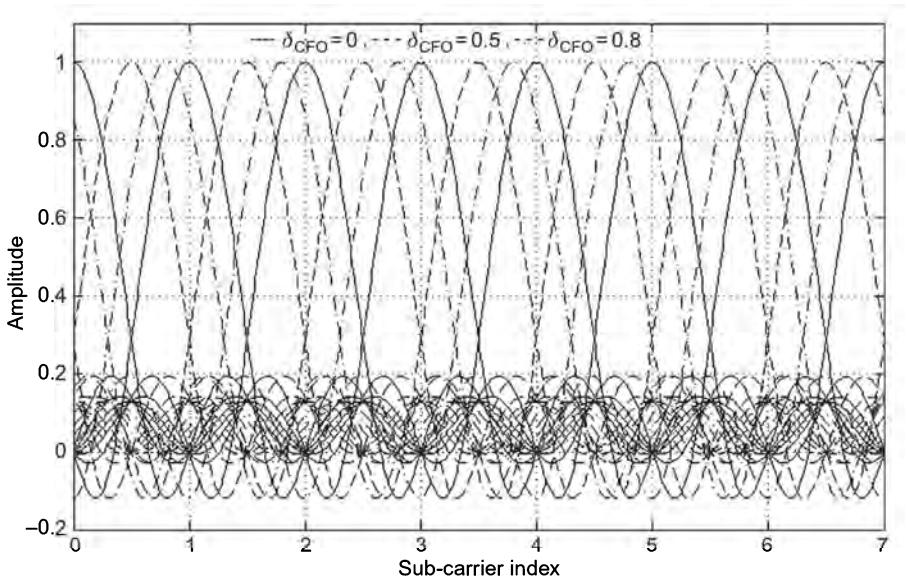


FIGURE 9.17

The effect of carrier frequency offset on OFDM signal for various δ_{CFO} values [159].

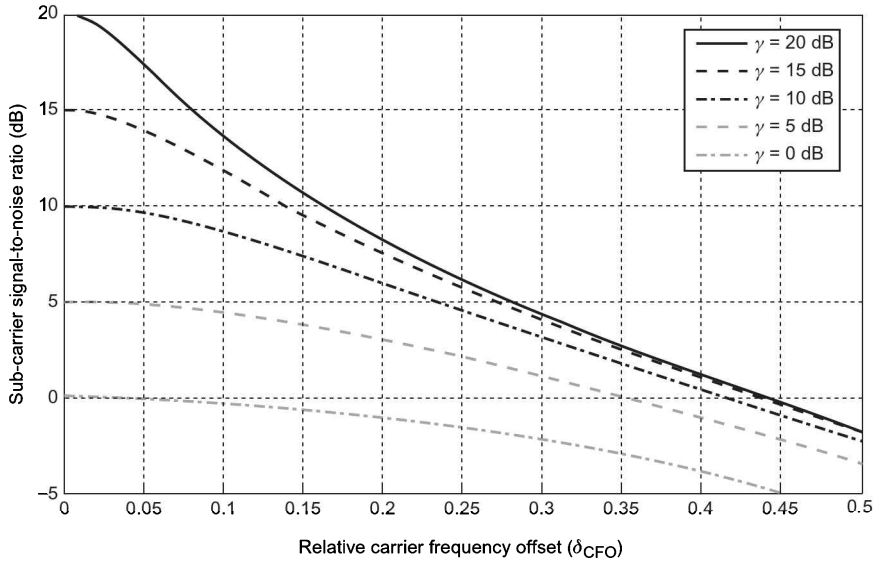


FIGURE 9.18

Variation of the SNR as a function of relative carrier frequency offset δ_{CFO} [159].

where ν_{max} denotes the maximum Doppler frequency. We further assume that the transmitted signal power is normalized. It can be noted that the ICI generated as a result of CFO is a special case of the above equation when $\Lambda(f) = \delta(f - f_{CFO})$, in which $\delta(f)$ represents the Dirac delta function. Using the classic Jakes' model of Doppler spread where the spaced-time correlation function is defined as $\Phi(t) = J_0(2\pi\nu t)$ in which $J_0(x)$ denotes the zero-order Bessel function of the first kind [131], the ICI power can be written as follows:

$$P_{\mathrm{ICI_{Jakes}}} = 1 - 2\int_0^1 (1-f)J_0(2\pi\nu_{max}T_s f)\mathrm{d}f \tag{9.18}$$

which approximately gives an upper bound on the ICI power due to Doppler spread. Comparison of the power of ICI generated by CFO and Doppler spread suggests that the ICI impairment due to the former is higher than the latter.

The effect of CFO on EVM for an OFDM system with $E_b/N_0 = 30$ dB, $N_{FFT} = 64$ has also been investigated and is shown in Figure 9.19. It is shown that the EVM increases with increasing values of relative CFO even at such relatively high SNR.

Figure 9.20 shows the bit error rate sensitivity of an OFDM system under AWGN channel to CFO. The increasing values of relative CFO degrade the BER performance of the system.

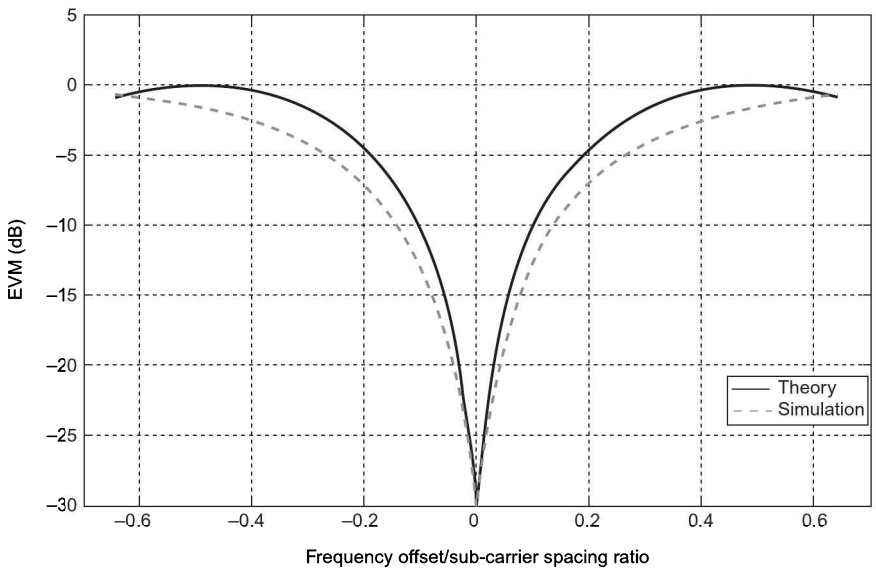


FIGURE 9.19

Error vector magnitude versus frequency offset in OFDM ($E_b/N_0 = 30$ dB, $N_{FFT} = 64$) [159].

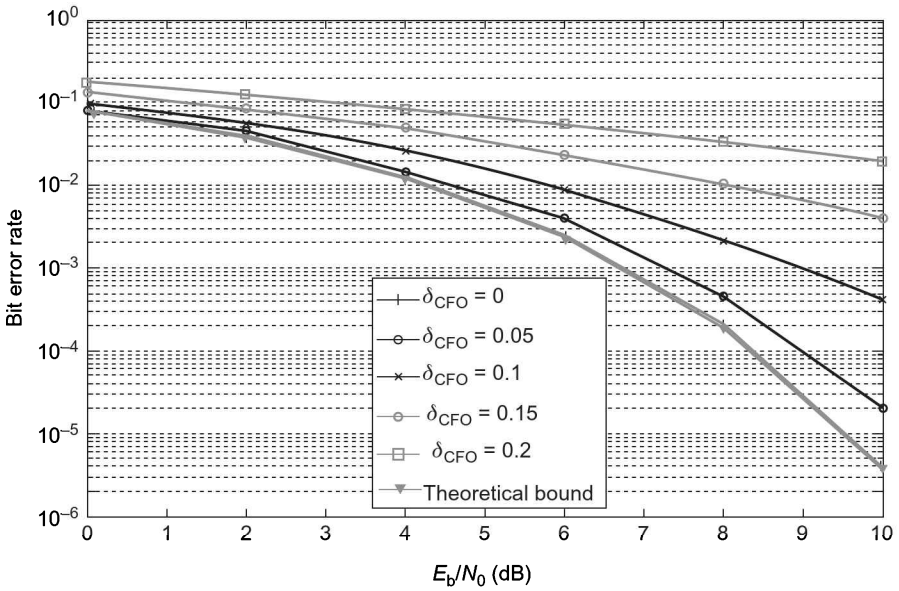


FIGURE 9.20

BER sensitivity to relative CFO δ_{CFO} in an OFDM system in AWGN channel [141].

9.4.4 Effect of oscillator phase noise on OFDM signals

Oscillators are used in typical radio circuits to drive the RF mixers which are utilized for the up-conversion/down-conversion of the band-pass signals. Ideally, the spectrum of the oscillator is expected to have an impulse at the frequency of oscillation with no frequency components elsewhere. However, the spectrum of a practical oscillator's output does exhibit spectrum variation around the oscillation frequency due to phase noise. The impact of local oscillator phase noise on the performance of an OFDM system has been extensively studied in the literature [159]. It has been shown that phase noise may have significant effects on OFDM signals with small sub-carrier spacing (i.e., large OFDM symbol duration in time). Long symbol duration is required for implementing a long guard interval that can mitigate long multipath delay in single frequency operation without excessive reduction of data throughput. The studies suggest that phase noise in OFDM systems can result in two effects: a common sub-carrier phase rotation on all the sub-carriers; and a thermal-noise-like sub-carrier de-orthogonality.

The common phase error, i.e., constellation rotation, on the demodulated sub-carriers is caused by the phase noise spectrum from DC up to the frequency of sub-carrier spacing. This low-pass effect is due to the long integration time of the OFDM symbol duration. This phase error can in principle be corrected by using pilots within the same symbol (in-band pilots). The phase error causes sub-carrier

constellation blurring rather than rotation. It results from the phase noise spectrum contained within the system bandwidth. This part of the phase noise is more crucial, since it cannot be easily corrected. The SNR degradation caused by the common phase error can be quantified as $S/N_{\text{phase-rotation}} = 1/(I(\alpha)\beta\,\Delta f)$, where Δf is the sub-carrier spacing, β denotes the upper bound of the phase noise spectral mask, α is the ratio of the equivalent spectrum mask noise bandwidth and sub-carrier spacing, and $I(\alpha) = \int_{-\alpha}^{+\alpha} \text{sinc}(\pi x)\text{d}x$ with $I(0.5) = 0.774$, $I(1) = 0.903$, and $I(\infty) = 0.774$. It can be seen that, when $\alpha > 1$, the common phase error decreases as the sub-carrier spacing decreases [52].

To mathematically model the effect of the phase noise, let us consider the noisy output of an oscillator which contains phase noise $\phi(t)$ as follows $v(t) = A_c \cos(\omega_c t + \phi(t))$. Let us further assume that the stochastic variation of the phase can be modeled as the output of a system with a step function impulse response and input perturbation $n(t)$ as follows: $\phi(t) = \int_{-\infty}^{t} n(\tau)\text{d}\tau$. Based on the above assumption, the single-sided power spectral density of the phase can be written as follows: $S_\phi(f) = S_n(f)/(2\pi f)^2$, in which $S_n(f)$ denotes the noise power spectral density function. As an example, if $S_n(f)$ is modeled as white noise, then $S_\phi(f) \approx f^{-2}$, and if $S_n(f)$ is modeled as flicker noise, then $S_\phi(f) \approx f^{-3}$. Considering that the power spectral density of the phase is difficult to observe, one may alternatively look at the power spectral density of the oscillator's noisy output $v(t)$. It can be shown that the power spectral density of $v(t)$ can be calculated as follows [159]:

$$S_v(f) = \sum_{k=-\infty}^{k=+\infty} \left(\frac{a_k^2 + b_k^2}{2}\right) \frac{\beta k^2 f_c^2}{\beta^2 \pi^2 k^4 f_c^4 + (f - k f_c)^2} \tag{9.19}$$

where $\{a_k\}$ and $\{b_k\}$ denote the Fourier series coefficients of $v(t)$ and β is a constant. Given that we are only interested in evaluating $S_v(f)$ at f_c, the above equation can be simplified as follows:

$$S_v(f) = \left(\frac{a_1^2 + b_1^2}{2}\right) \frac{\beta f_c^2}{\beta^2 \pi^2 f_c^4 + (f - f_c)^2} \tag{9.20}$$

The above function is a Lorentz distribution.[8] We now define function $\Omega(f)$ as the ratio of noise power in the 1 Hz bandwidth at offset f from center frequency to carrier power which is expressed in dBc/Hz. As theoretically expected, having a higher phase noise in the signal does not increase the total power. A signal with higher phase noise will have smaller power near f_c and will have a broader spectrum around the center frequency. Conversely, a signal with lower phase noise

[8]The Cauchy distribution is a continuous probability distribution which is also known as the Lorentz distribution, Cauchy–Lorentz distribution, or Lorentzian function. It describes the distribution of a random variable that is the ratio of two independent standard normal random variables and has the probability density function defined as $f(x; 0, 1) = 1/[\pi(1 + x^2)]$ [30].

has a sharper peak at the center frequency with less deviation. Therefore, $\Omega(f)$ can be expressed as follows:

$$\Omega(f) = \frac{1}{\pi}\frac{\gamma}{\gamma^2 + (f - f_c)^2} \quad \gamma = \beta\pi f_c^2 \tag{9.21}$$

It can be shown that $\int_{-\infty}^{+\infty} \Omega(f)\mathrm{d}f = 1$ [159]. Figure 9.21 illustrates the ratio of the noise power to carrier power at $f_c = 500$ Hz for different values of β. It can be seen from the figure, for higher values of β, the spectrum becomes wider with smaller magnitude of the main lobe of the spectrum. Note that a wider main lobe does not increase the total power of the carrier.

In summary, an OFDM system has the following advantages [52]:

- It is sufficiently flexible in meeting various design requirements such as low complexity, bandwidth efficiency and scalability, spectrum shaping, and low sensitivity to various impairments.
- OFDM is well suited to multi-antenna transmission schemes.
- It requires no adaptation to instantaneous channel response for low-order modulations. It is robust to impulse interference and channel variations. Channel estimation is required for high-order QAM modulations.
- As a multi-tone transmission technique, OFDM is less sensitive to shift in sampling time in comparison to serial transmission techniques.

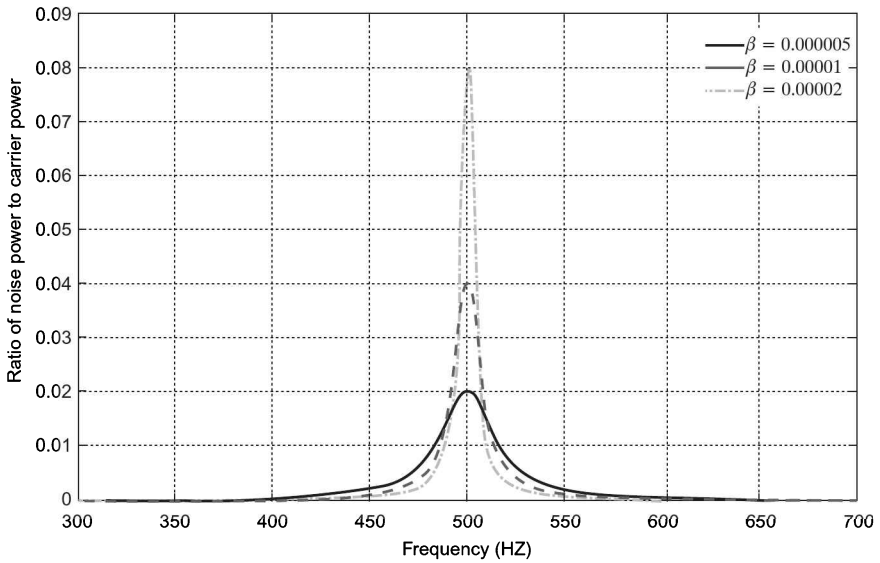


FIGURE 9.21

Plot of the noise power to carrier power $\Omega(f)$ at $f_c = 500$ Hz for different values of β [159].

- The bandwidth efficiency of an OFDM system approaches the Nyquist rate as the FFT size increases (this, however, increases system complexity and is subject to an upper limit due to channel variations and CFO).
- For fixed bandwidth efficiency, the complexity (multiplications per symbol) of an FFT-based OFDM system grows logarithmically with the increase of the channel multipath delay spread, in comparison with a linear increase of complexity for equalizers for a single carrier system.
- A properly coded and interleaved OFDM system can exceed the performance of many other practical systems. Higher transmitter output back-off is required in comparison to a single carrier system, because of the high PAPR of an OFDM system.
- As a parallel transmission technique, OFDM is more sensitive to CFO and tone interference than that of a single carrier system.

9.5 Multiple access schemes

Orthogonal frequency division multiple access (OFDMA) is the multi-user variant of the OFDM scheme where multiple access is achieved by assigning subsets of sub-carriers to different users, allowing simultaneous data transmission from several users. In OFDMA, the radio resources are two-dimensional (2D) regions over time (an integer number of OFDM symbols) and frequency (a number of contiguous or non-contiguous sub-carriers). The difference between OFDM and OFDMA is illustrated in Figure 9.22. Similar to OFDM, OFDMA employs multiple closely spaced sub-carriers, but the sub-carriers are divided into groups of sub-carriers where each group is called a resource block. The grouping of sub-carriers into groups of resource blocks is referred to as sub-channelization. The sub-carriers that form a resource block do not need to be physically adjacent. In the downlink, a resource block may be allocated to different users. In the uplink, a user may be assigned to one or more resource blocks.

Sub-channelization defines sub-channels that can be allocated to mobile stations depending on their channel conditions and service requirements. Using sub-channelization, within the same time slot (i.e., an integer number of OFDM symbols), an OFDMA system can allocate more transmit power to user devices with lower SNR, and less power to user devices with higher SNR. Sub-channelization also enables the base station to allocate higher power to sub-channels assigned to indoor mobile terminals, resulting in better indoor coverage.

There are various multiple access schemes that have been used in cellular systems in the past few years which allow the network to share the available radio resources (i.e., time, frequency, code, and space) among a number of users in the cell in the downlink and uplink directions. Figure 9.23 illustrates the concept of resource sharing in various multiple access schemes. As mentioned earlier, OFDMA has been a promising multiple-access scheme that has recently been

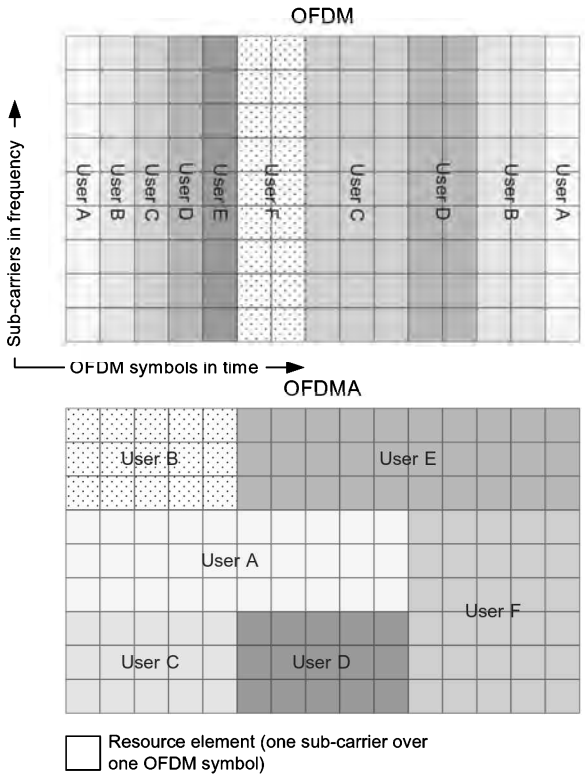


FIGURE 9.22

Illustration of OFDM and OFDMA concepts.

used in mobile broadband radio access technologies. The E-UTRA uses OFDMA and SC-FDMA as the multiple access schemes in the downlink and uplink directions, respectively. The OFDMA/SC-FDMA parameters for the E-UTRA have been chosen to facilitate coexistence with the UTRA on the same platform and sharing the same temporal sampling frequency of 30.72 MHz. The E-UTRA OFDMA/SC-FDMA parameters are given in Table 9.2. The following parameters are compatible with UMTS sampling frequency/chip rate of 3.84 Mcps and frame timing.

The OFDMA/SC-FDMA parameters in E-UTRA satisfy general design criteria for OFDM systems as follows:

$$\begin{aligned} &T_{CP} \geq \tau_{RMS} && \text{(to prevent ISI due to delay spread)} \\ &\Delta f \gg \nu_{max} && \text{(to mitigate ICI due to Doppler spread)} \\ &T_{CP}\ \Delta f \ll 1 && \text{(for spectral efficiency improvement)} \end{aligned} \tag{9.22}$$

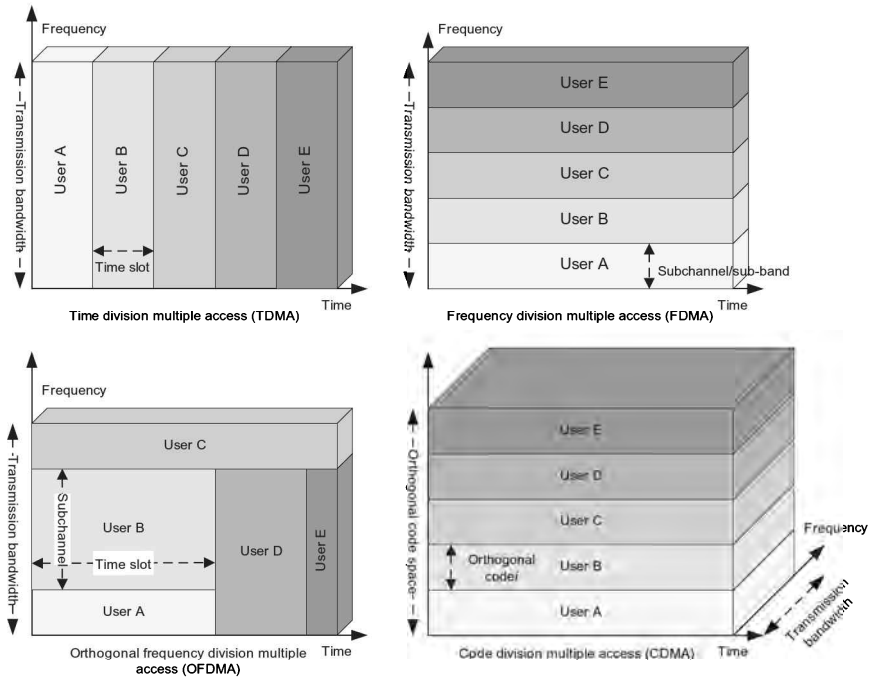


FIGURE 9.23

An illustration of the concept of different multiple access schemes.

In OFDMA, an OFDM symbol is constructed of sub-carriers, the number of which are determined by the FFT size. There are several sub-carrier types: (1) data sub-carriers are used for data transmission; (2) pilot or reference-signal sub-carriers are utilized for channel estimation and coherent detection; and (3) null sub-carriers are not used for pilot/data transmission. The null sub-carriers, including the DC sub-carrier, are used for guard bands. The number of used (or occupied) sub-carriers is always less than the FFT/IFFT size. The guard bands are used to allow spectrum sharing and to reduce the adjacent channel interference and out-of-band emissions. The sampling frequency is selected to be greater than or equal to the channel bandwidth. The number of time samples in a radio frame is always an integer, and to further simplify the design of analog transmit filter the sampling frequency is scaled by a factor greater than one (e.g., the sampling frequency for 20 MHz bandwidth is 30.72 MHz). As given in Table 9.2, the value of the over-sampling factor depends on the channel bandwidth. The E-UTRA supports a large number of channel bandwidths. The channel bandwidths that are integer multiples of 5 MHz are more important. The bandwidths that are not given

Table 9.2 E-UTRA OFDMA/SC-FDMA Parameters [3]

Parameter		Value					
Channel bandwidth (MHz)		1.4	3	5	10	15	20
Number of resource blocks $N_{RB}^{DL/UL}$		6	15	25	50	75	100
Number of occupied sub-carriers		72	180	300	600	900	1200
IFFT/FFT size		128	256	512	1024	1536	2048
Sub-carrier spacing Δf (kHz)		15 (7.5)					
Sampling frequency (MHz)		1.92	3.84	7.68	15.36	23.04	30.72
Samples/slot		960	1920	3840	7680	11520	15360
CP size T_{CP}	Normal CP ($\Delta f = 15$ kHz)	5.21 μs (first symbol of the slot) 4.69 μs (other symbols of the slot) $N_s = 7$ symbols/slot $N_{CP}(l) = \begin{cases} 160 & l = 0 \\ 144 & l = 1, 2, \ldots, 6 \end{cases}$					
	Extended CP ($\Delta f = 15$ kHz)	16.67 μs $N_s = 6$ symbols/slot $N_{CP}(l) = 512 \quad l = 0, 1, \ldots, 5$					
	Extended CP ($\Delta f = 7.5$ kHz)	33.33 μs $N_s = 3$ symbols/slot $N_{CP}(l) = 1024 \quad l = 0, 1, 2$					

in Table 9.2 are supported via tone dropping, i.e., dropping the sub-carriers at the edges of the frequency band based on 10 and 20 MHz systems (e.g., for 6 MHz channel bandwidth, the parameter of 10 MHz is used and the edge sub-carriers are dropped to fit the spectrum block).

9.6 Duplex modes

One of the key elements of any radio communication system is the way in which radio communications are maintained in both directions. For cellular systems, it is necessary to enable transmission of data in both directions simultaneously, and this places a number of constraints on the schemes that may be used to control the transmission flow. As a result, the duplex scheme forms a very basic part of the overall specification for a cellular or any radio communication system. The term duplex refers to the bidirectional communications between two devices. When unpaired spectrum, or alternatively the same RF carrier is used for downlink and uplink communications, the transmit/receive functions are time-multiplexed.

When paired spectrum, or alternatively two RF carriers are used for downlink and uplink communications, the transmit/receive functions are frequency-multiplexed. The FDD is a duplex scheme in which uplink and downlink

transmissions occur simultaneously using different frequencies. The downlink and uplink frequencies are separated by a sufficiently large frequency offset (see Table 9.1 for various E-UTRA bands). For the FDD scheme to properly operate, it is necessary that the frequency separation, i.e., channel separation between the transmission and reception frequencies, be sufficient to prevent the receiver being excessively affected by the transmitter signal. Receiver blocking is an important issue with FDD schemes, and often highly selective filters may be required. For cellular systems using FDD, filters are required within the base station and also the user terminal to ensure sufficient isolation of the transmitter signal without desensitizing the receiver. While implementation cost is not a significant constraint for the base stations, placing a filter into the user terminals is more of an issue. The use of an FDD system does enable simultaneous transmission and reception of signals. However, two channels are required and this may not always use the available spectrum efficiently. The spectrum used for FDD systems is allocated by the regulatory authorities. Since there is a frequency separation between the uplink and downlink frequencies, it is not normally possible to reallocate spectrum to change the balance between the capacity of the uplink and downlink directions, if there are changing capacity requirements for each direction.

The TDD is a duplex scheme where uplink and downlink transmissions occur at different times, but may share the same frequency. In other words, the downlink and uplink transmissions are multiplexed in time and are not concurrent. While FDD transmissions require a large frequency separation between the transmit and receive frequencies, TDD schemes require a guard time or guard interval between transmission and reception. This must be sufficient to allow the signals traveling from the remote transmitter to arrive before a transmission is started and the receiver is shut down. Although this delay is relatively short, when switching between transmission and reception many times a second, even a small guard time can reduce the efficiency of the system, because a percentage of the time must be used for the guard interval. For small-sized cells, e.g., up to a mile or so, the guard interval is normally small and acceptable. For large cell sizes, it may become an issue.

The guard interval required for TDD will comprise two main elements: (1) a time allowance for the propagation delay for any transmission from a remote transmitter to arrive at the receiver. This will depend upon the distances involved (i.e., cell radius). (2) A time allowance for the transceiver to switch from receive to transmit mode. The switching times can vary considerably depending on the implementation, but can take a few microseconds. As a result, TDD is not normally suitable for use over very large cell sizes as the guard time increases and the spectral efficiency decreases.

It is often found that traffic in both directions is not balanced. Typically there is more data transmitted in the downlink direction of a cellular system. This means that, ideally, the capacity should be greater in the downlink direction. Using a TDD system, it is possible to change the capacity in either direction relatively easily by changing the number of time slots allocated to each direction. Often this is dynamically configurable so it can be altered to meet the demand. A further aspect

to be noted with TDD transmission is that of latency. Since the data may not be immediately transmitted, a result of the time multiplexing between transmit and receive circuitry, there may be a small delay between the time that the data is generated until it is actually transmitted in a TDD system. Typically this may be a few milliseconds depending upon the frame timing. Both TDD and FDD have their advantages, and each can be used in different deployment scenarios. Before deciding on a particular type of duplex scheme, it is necessary to analyze the advantages and disadvantages of each duplex mode. Table 9.3 summarizes the relative advantages/disadvantages of TDD and FDD systems.

The E-UTRA supports TDD and FDD schemes with a great extent of commonality in the baseband processing. In order to reduce the implementation complexity and cost of FDD terminals, and further to increase the reuse of baseband functional elements, a half-duplex FDD (H-FDD) operation is supported where the downlink and uplink transmissions are not simultaneous but occur in two different frequencies. As shown in Figure 9.24, classic H-FDD operation does not efficiently utilize the radio resources on the downlink and uplink RF carriers. The complementary grouping and scheduling of users would allow efficient use of downlink and uplink resources in an H-FDD operation. The various duplex schemes are illustrated in Figure 9.24. For H-FDD operation, a guard period (GP) is virtually created by the UE by not receiving the last OFDM symbol(s) of a downlink subframe immediately preceding an uplink subframe in which the UE is active. The length of the GP is the sum of the maximum round-trip propagation delay in the cell, transmit-to-receive, and receive-to-transmit switching delay at the UE.

In E-UTRA, the differences between TDD and FDD schemes are mainly in the physical layer, and therefore are transparent to the higher layers. As a result, there are no significant operational differences between the two modes in the system architecture. At the physical layer, the fundamental design goal was to achieve as much commonality between the two modes as possible. The key design differences between the two stem from the need to support various TDD UL/DL allocations and provide coexistence with other TDD systems. In this regard, several additional features are exclusively developed for E-UTRA TDD. For instance, the FDD and TDD modes differ in the time placement of the downlink synchronization signals. In FDD, the primary and secondary synchronization signals are contiguously placed within one subframe, while in TDD, the two signals are placed in different subframes, separated by two OFDM symbols. However, this separation of the primary and secondary synchronization signals does not affect the performance of synchronization channels relative to FDD systems.

The system-level simulations also show that the sector throughput and cell-edge throughput of E-UTRA TDD and FDD mode are quite similar for best effort traffic when comparing a 10 + 10 MHz FDD system to a 20 MHz TDD system. The spectral efficiency of the TDD mode is slightly degraded due to the

Table 9.3 Comparison of TDD and FDD Attributes

Attribute	TDD	FDD
Paired spectrum	Does not require paired spectrum as both transmit and receive occur on the same channel.	Requires paired spectrum with sufficient frequency separation to allow simultaneous transmission and reception.
Hardware cost	Lower cost as no duplexer[a] is needed to isolate the transmitter and receiver.	Duplexer is needed and the implementation cost is higher
Channel reciprocity	Channel propagation is the same in both directions which enables estimation of the downlink channel from the uplink channel.	Channel characteristics are different in both directions due to use of different frequencies.
UL/DL asymmetry	It is possible to dynamically change the UL and DL ratio based on the traffic volume in each direction.	UL/DL bandwidths are determined by frequency allocation designated by the regulatory authorities. It is therefore not possible to dynamically adapt to the traffic volume.
Guard period/ frequency separation	Guard period required to ensure uplink and downlink transmissions do not interfere. Large guard period will limit the capacity. Larger guard period is normally required if distances are increased to accommodate larger propagation delays. Note that a guard band in frequency-domain is required to suppress interference to adjacent bands.	Frequency separation is required to provide sufficient isolation between uplink and downlink. However, large frequency separation does not impact the capacity. Note that a guard band in frequency-domain is required to suppress interference to adjacent bands.
Discontinuous transmission	Discontinuous transmission is required to allow both uplink and downlink transmissions. This can degrade the performance of the RF power amplifier in the transmitter.	Continuous transmission is possible.
Switching point synchronization	Base stations are required to be synchronized with respect to the uplink and downlink transmission times. If neighboring base stations use different uplink and downlink assignments and share the same channel, then interference may occur between cells.	Not applicable.

[a] *A diplexer is a passive device that implements frequency domain multiplexing. Two ports are multiplexed onto a third port. The signals on each input port occupy non-overlapping frequency bands. Consequently, the signals on the input ports can coexist on the output port without interfering with each other. On the other hand, a duplexer is a device that allows bidirectional communication over a single channel. In radar and radio communications systems, the duplexer isolates the receiver from the transmitter while permitting them to share a common antenna. Most radio repeater systems include a duplexer. Duplexers are designed for operation in the frequency band used by the receiver and transmitter, and must be capable of handling the output power of the transmitter. They must provide sufficient isolation between transmitter and receiver to prevent receiver desensitization [34].*

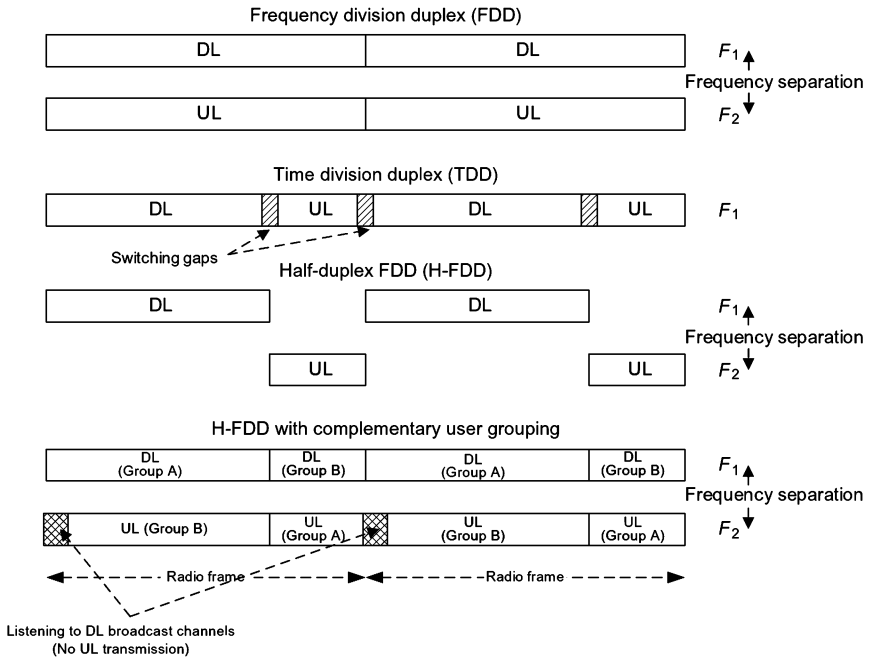


FIGURE 9.24

Illustration of the concept of different duplexing schemes.

additional overhead corresponding to the GP. Another advantage of TDD systems is the flexibility they may offer to network operators in terms of adjusting the DL/UL ratio. This feature allows operators to configure the DL/UL ratio to suit the user traffic in their network.

A common misperception regarding TDD systems is that the uplink coverage of the TDD systems is inferior relative to their FDD counterpart. The link-budget analysis for both systems in the outdoor and indoor environments with compatible system configuration and transmitter and receiver settings shows that the TDD system matches the uplink coverage of the FDD counterpart [10]. In fact, the E-UTRA TDD and FDD modes utilize the same slot/subframe structure and resource block size, thus the same amount of power is transmitted during the same time interval, implying that the coverage will be the same. What has been fueling this misconception is the single-user uplink data rate in an unloaded network at the cell-edge, where an FDD system will show a cell-edge performance that is twice that of the TDD system. It must be noted that the link budget analysis focuses only on the cell-edge single user performance when that user terminal is at the maximum limit of its uplink power, hence it can only be given access to five resource blocks (one resource block every 1 ms for 5 ms, or half of the length of

the 10 ms E-UTRA radio frame with DL:UL = 1:1 ratio) to meet the 200 mW maximum power requirement. In comparison, in an FDD system at the cell-edge, one single device will be able to send 200 mW continuously for 10 ms, assuming that the scheduler can give the device access to 10 resource blocks of 1 ms with no waiting time. However, in real-life scenarios, the cells have more than one user and the performance of FDD and TDD systems will be very similar, as the FDD device will not be able to access the full-frame 10 uplink resource blocks similar to an unloaded network. In that case, because the TDD channel is twice as wide as the FDD channel, the user is likely to be given access to as many resource blocks as an FDD system despite the half-uplink time in the frame compared to an FDD frame. For network operators, an E-UTRA TDD will provide the same coverage and very similar sector throughput relative to an E-UTRA FDD system. The TDD with its adaptable DL:UL ratio would allow an optimized and efficient use of the uplink and downlink bandwidths as the operators are able to match the asymmetry of data traffic and to maximize spectrum utilization and efficiency in almost all scenarios [163].

9.7 Frame structure

Downlink and uplink transmissions are structured into radio frames with 10 ms duration. The E-UTRA supports two frame structures, namely, type 1 and type 2 for FDD and TDD duplex schemes, respectively. The radio frames are numbered using a system frame number (SFN). The SFN is defined using a string of 10 bits which provide an integer value from 0 to 1023. Therefore, the SFN repeats once every 10.24 s, i.e., 1024 frames. The eight most significant bits of the SFN are broadcast within the master information block (MIB) of the system information on the broadcast channel (BCH). The remaining bits of SFN are deduced from the four frame cycles used to transmit the MIB (i.e., the first radio frame corresponds to "00" and the fourth radio frame corresponds to "11"). As given in Table 9.2, using the normal cyclic prefix, there are seven OFDM symbols within each slot. In the case of multi-cell aggregation, the same frame structure is used across all serving cells. Transmissions in multiple cells can be aggregated up to four secondary cells (in this context cell refers to RF carriers) in addition to the primary cell.

To provide enhanced multicast and broadcast services, E-UTRA has the capability to transmit multicast and broadcast content over a single frequency network (MBSFN), where a time-synchronized common waveform is transmitted from multiple eNBs, allowing macro-diversity combining of multi-cell transmissions at the UE. The cyclic prefix is utilized to cover the difference in the propagation delays, which makes the MBSFN transmission appear to the UE as a transmission from a single large cell. In LTE, transmission on a dedicated carrier for MBSFN

with the possibility to use a longer CP with a sub-carrier bandwidth of 7.5 kHz is supported, as well as transmission of MBSFN on a carrier with both MBMS and unicast transmissions using time division multiplexing [7].

9.7.1 Frame structure type 1

Frame structure type 1 is applicable to both full-duplex FDD and H-FDD schemes. In frame structure type 1, as illustrated in Figure 9.25, each 10 ms radio frame is divided into 10 equally-sized subframes ($T_{\text{frame}} = 307,200T_s = 10T_{\text{subframe}}$). Each subframe consists of two equally sized and consecutive 0.5 ms slots ($T_{\text{slot}} = 15360T_s$). Each slot in turn consists of a number of OFDM symbols which can be either seven (normal cyclic prefix) or six (extended cyclic prefix). As shown in Figure 9.25, the useful symbol time is $T_u = 2048T_s = 66.7\ \mu s$. For a normal cyclic prefix, the first symbol has a cyclic prefix lengths of $T_{CP} = 160T_s = 5.2\ \mu s$. The remaining six OFDM symbols have a cyclic prefix of length $T_{CP} = 144T_s = 4.7\ \mu s$. The reason for different cyclic prefix lengths of the first symbol is to make the overall slot length in terms of time units divisible by 15,360. For the extended mode, the cyclic prefix is $T_{CP} = 512T_s = 16.67\ \mu s$. The length of the cyclic prefix is longer than the delay spread of a few microseconds typically encountered in practice. The normal cyclic prefix is used in urban cells and high data rate applications, while the extended cyclic prefix is used in special cases like multi-cell broadcast and in very large cells (e.g., rural areas, low data rate

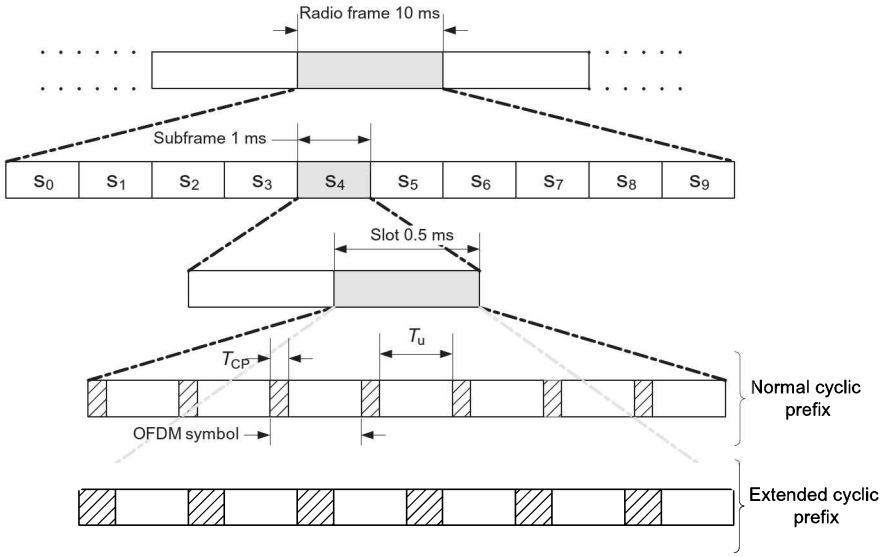


FIGURE 9.25

Frame structure type 1 with different cyclic prefix sizes [3,10].

applications). The cyclic prefix reduces the physical layer capacity by approximately 7.5% in the case of a normal cyclic prefix.

One way to reduce the relative overhead due to cyclic prefix insertion is to reduce the sub-carrier spacing with a corresponding increase in the OFDM symbol duration. However, this will increase the sensitivity of the OFDM transmission to frequency deviations resulting from fast channel variations and large Doppler spread, as well as different types of frequency errors due to RF circuitry. For FDD systems, 10 subframes are separately available for downlink and uplink transmissions in each 10 ms interval. Uplink and downlink transmissions are separated in the frequency-domain. In H-FDD operation, the UE cannot transmit and receive at the same time, while there are no such restrictions in full-duplex FDD.

The transmission time interval (TTI) of the E-UTRA downlink/uplink is 1 ms. The TTI is defined as the duration of the transmission of the physical layer encoded packets over the radio air-interface. In other words, the TTI refers to the length of an independently decodable transmission on the air-link. The data on a transport channel is organized into transport blocks. In each TTI, at most one transport block of variable size is transmitted over the radio air-interface to or from a terminal in the absence of spatial multiplexing. In the case of spatial multiplexing, up to two transport blocks per TTI are transmitted [3].

In the frequency-domain, the number of sub-carriers N_{FFT} ranges from 128 to 2048, depending on the channel bandwidth, with 512 and 1024 for 5 and 10 MHz, respectively, being most commonly used in practice. The sub-carrier spacing is 15 kHz. The sampling rate is chosen as $f_s = 15,000N_{\text{FFT}}$, resulting in a sampling rate that is an integer multiple or sub-multiple of the WCDMA chip rate of 3.84 Mcps. The OFDM parameters have been chosen such that FFT sizes and sampling rates are easily scalable for all operation bandwidths, while at the same time facilitating the implementation of multi-mode terminals with a common reference clock. Not all the sub-carriers are modulated or used. The DC sub-carrier is not used, as well as a number of sub-carriers on either side of the frequency band; thereby, approximately 10% of sub-carriers are used as guard sub-carriers.

9.7.2 Frame structure type 2

Frame structure type 2 is applied to a TDD duplex scheme where the downlink and uplink transmissions are time-multiplexed over the extent of a radio frame (Figure 9.26). Each radio frame of length 10 ms ($T_{\text{frame}} = 307,200T_s$) comprises two half-frames, where each half-frame consists of five subframes of length 1 ms. The supported downlink/uplink ratios (i.e., the ratio of downlink and uplink subframes in a frame) are listed in Table 9.4 and shown in Figure 9.27, where for each subframe in a radio frame, "DL" denotes the subframe reserved for downlink transmissions, "UL" denotes the subframe designated to uplink transmissions, and "S" denotes a special subframe comprising three fields downlink pilot time slot (DwPTS), GP, and uplink pilot time slot (UpPTS). The length of DwPTS and

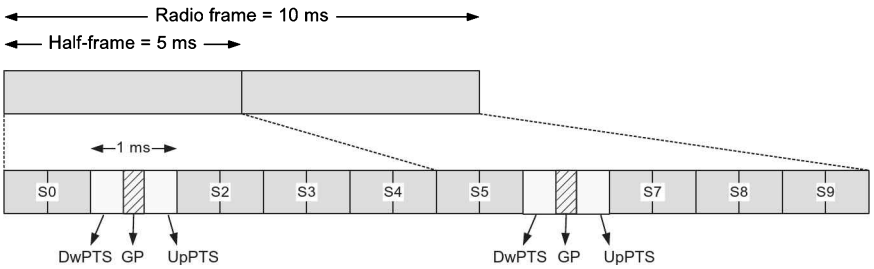


FIGURE 9.26

Frame structure type 2 [3].

Table 9.4 E-UTRA TDD Permissible DL/UL Ratios [3]

E-UTRA TDD Frame Configuration	Switching-Point Periodicity (ms)	Subframe Number									
		0	1	2	3	4	5	6	7	8	9
0	5	DL	S	UL	UL	UL	DL	S	UL	UL	UL
1	5	DL	S	UL	UL	DL	DL	S	UL	UL	DL
2	5	DL	S	UL	DL	DL	DL	S	UL	DL	DL
3	10	DL	S	UL	UL	UL	DL	DL	DL	DL	DL
4	10	DL	S	UL	UL	DL	DL	DL	DL	DL	DL
5	10	DL	S	UL	DL	DL	DL	DL	DL	DL	DL
6	5	DL	S	UL	UL	UL	DL	S	UL	UL	DL

UpPTS are given in Table 9.5, assuming that the total length of DwPTS, GP, and UpPTS is always 1 ms. Each subframe i is defined as two slots $2i$ and $2i + 1$ of length 0.5 ms.

The E-UTRA TDD supports downlink/uplink ratios with both 5 and 10 ms downlink/uplink switching-point periodicity. In the case of 5 ms switching-point periodicity, there are two special subframes within one radio frame. Subframes 0 and 5 and DwPTS are always reserved for downlink transmission due to essential downlink physical channel mapping. The UpPTS interval and the subframe immediately following the special subframe are always reserved for uplink transmission. In case multiple cells are aggregated, and in order to minimize inter-cell interference, the UE may assume the same downlink/uplink ratio across all the cells and that the GPs of the special subframes in different cells have a minimum overlap of $1456T_s$. For the switching from uplink to downlink transmission, no special subframe is provisioned, but the GP includes the sum of switching times from DL to UL and UL to DL. The switching from UL to DL is achieved by provisioning an appropriate timing advance at the UE.

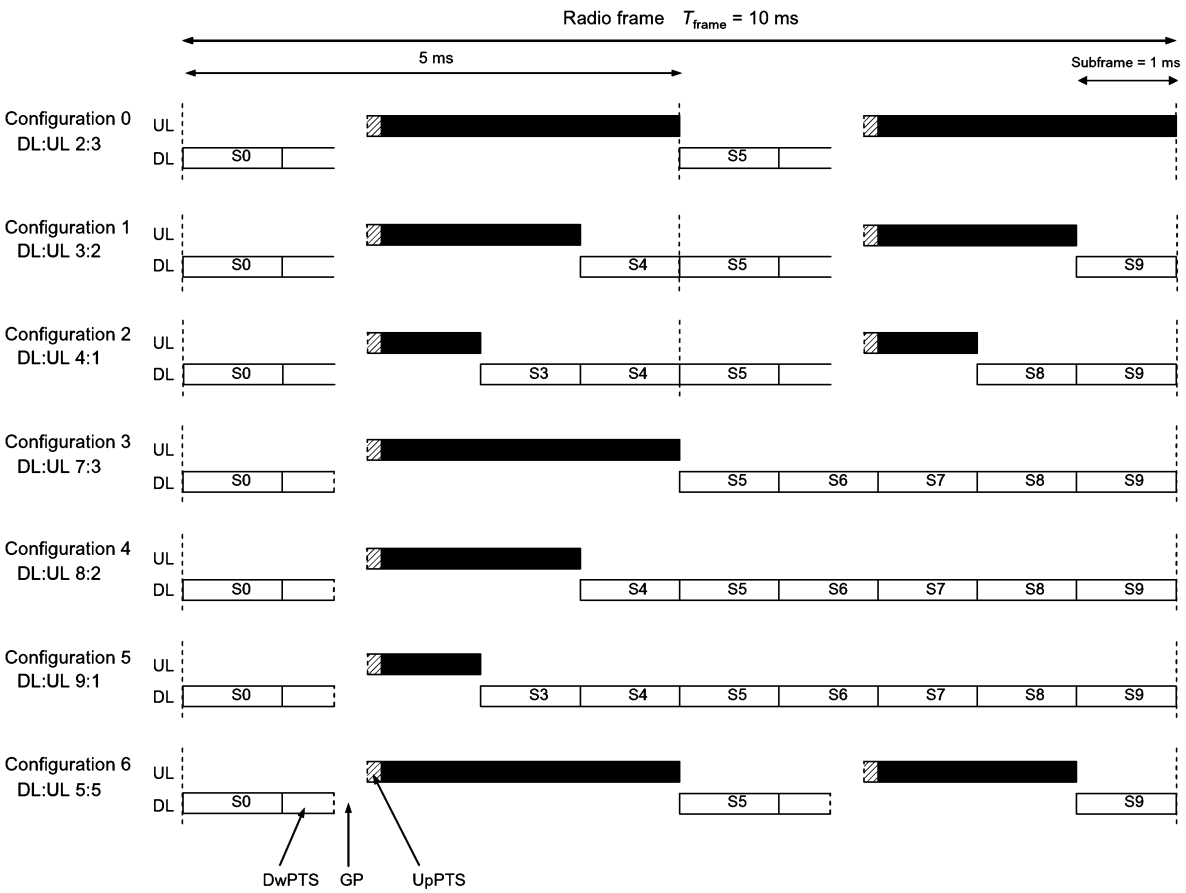


FIGURE 9.27

Illustration of various frame structure type-2 configurations [10].

Table 9.5 Various Special Subframe Configurations (DwPTS/GP/UpPTS) [3]

Special Subframe Configuration	Normal Cyclic Prefix in Downlink				Extended Cyclic Prefix in Downlink			
	DwPTS (Duration/ Number of OFDM Symbols)	GP (Number of OFDM Symbols)	UpPTS		DwPTS (Duration/ Number of OFDM Symbols)	GP (Number of OFDM Symbols)	UpPTS	
			Normal Cyclic Prefix in Uplink (Duration/ Number of OFDM Symbols)	Extended Cyclic Prefix in Uplink (Duration/ Number of OFDM Symbols)			Normal Cyclic Prefix in Uplink (Duration/ Number of OFDM Symbols)	Extended Cyclic Prefix in Uplink (Duration/ Number of OFDM Symbols)
0	$6592T_s$ 3	10			$7680T_s$ 3	8		
1	$19760T_s$ 9	4			$20480T_s$ 8	3		
2	$21952T_s$ 10	3	$2192T_s$ 1	$2560T_s$ 1	$23040T_s$ 9	2	$2192T_s$ 1	$2560T_s$ 1
3	$24144T_s$ 11	2			$25600T_s$ 10	1		
4	$26336T_s$ 12	1			$7680T_s$ 3	7		
5	$6592T_s$ 3	9			$20480T_s$ 8	2		
6	$19760T_s$ 9	3			$23040T_s$ 9	1	$4384T_s$ 2	$5120T_s$ 2
7	$21952T_s$ 10	2	$4384T_s$ 2	$5120T_s$ 2	$12800T_s$			
8	$24144T_s$ 11	1			– –	–	– –	– –
9	$13168T_s$ 6	6			– –	–	– –	– –

Within a radio frame, an E-UTRA TDD system switches multiple times between downlink and uplink transmissions. In this process, different signal propagation times between eNB and various UEs must be taken into consideration in order to prevent interference with other subframes. The timing advance feature prevents conflicts as a result of switching from uplink to downlink transmission. The UEs are informed by the eNB as to when they must start their uplink transmission. The greater the distance between the eNB and the UE, the earlier the UE starts transmitting. This helps to ensure that all signals reach the eNB with proper timing. When switching from the downlink to the uplink, a GP is inserted between the DwPTS and the UpPTS fields. The duration of the GP depends on the signal propagation round-trip time from the eNB to the UE, as well as on the time the UE requires to switch from receiving to transmitting states. The duration of the GP is configured by the network based on the cell size. Because the overall length of the special subframe remains constant, and the GP length varies based on cell size, the lengths of the DwPTS and UpPTS have to be adjusted. There are ten different special subframe configurations specified for E-UTRA TDD, as given in Table 9.5. The length of the GP determines the maximum supportable cell size. The E-UTRA should support cell sizes of up to 100 km, requiring a GP of approximately 666.7 μs. This is possible by choosing configuration 0 in Table 9.5 for the special subframe.

While the GP separates uplink and downlink transmissions, the other special fields are used for data/control signaling transmission. The DwPTS field carries synchronization and user data, as well as the downlink control channel for transmitting scheduling and control information. The UpPTS field is used for transmitting PRACH and SRS.

An example of coexistence with a legacy TD-SCDMA system [171] is shown in Figure 9.28, which illustrates switching point alignment between the two systems. It must be noted that in order to reduce the number of special subframe patterns in LTE TDD, only few legacy TD-SCDMA frame configurations are supported; therefore, the support is limited to the most common deployment scenarios, e.g., TD-SCDMA DL:UL = 5:2 and DL:UL = 4:3 configurations [163]. Despite the unequal size of the slots in TD-SCDMA and LTE TDD, the use of an

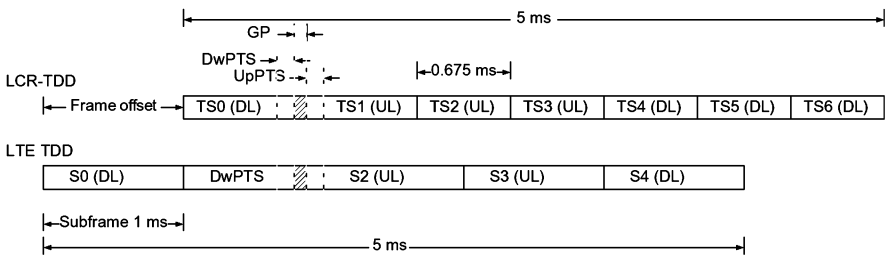


FIGURE 9.28

Example of LTE TDD coexistence with a TD-SCDMA system [163].

appropriate frame offset (as shown in Figure 9.28) allows alignment of the special subframes to the possible extent such that uplink and downlink transmissions of the two systems are synchronized [163].

9.7.3 MBSFN subframe structure

In E-UTRA frame structure, each downlink subframe and the DwPTS fields in the case of TDD is divided into a control region, consisting of the first OFDM symbols of the subframe, and a data region, comprising the remaining OFDM symbols of the subframe. A subset of the downlink subframes in a radio frame on a carrier supporting physical downlink shared channel (PDSCH) transmission can be configured as MBSFN subframes by higher layers. Each MBSFN subframe is divided into a non-MBSFN region and an MBSFN region. The non-MBSFN region spans the first one or two OFDM symbols in an MBSFN subframe. The MBSFN region in an MBSFN subframe is defined as the OFDM symbols not used for the non-MBSFN region.

The MBSFN subframes were originally designed to support MBSFN transmissions. However, they have also been found useful in other areas such as E-UTRA relaying functionality or the new carrier types. The MBSFN subframes are viewed as a generic tool, not necessarily related to MBSFN transmissions, which can provide time–frequency resources for enabling non-backward compatible functionalities relative to LTE Rel-8. An MBSFN subframe, as illustrated in Figure 9.29, consists of a control region of length one or two OFDM symbols, which is identical to normal subframes, followed by an MBSFN data region whose content will depend on the usage of the MBSFN subframe. The reason for maintaining the same control region in the MBSFN subframes is in consideration for normal

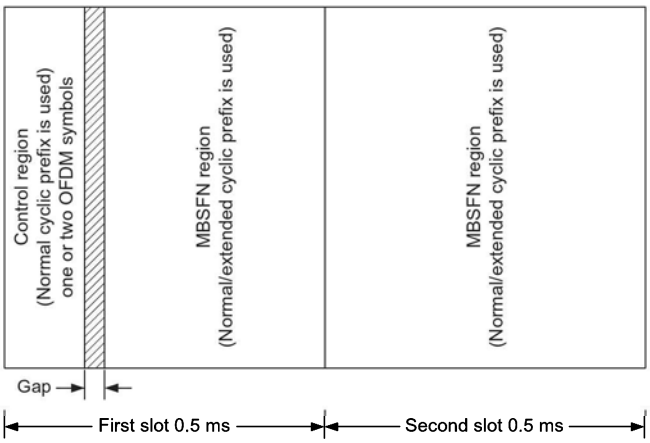


FIGURE 9.29

MBSFN subframe structure [15].

operation of LTE Rel-8 UEs, and in order to transmit control signaling necessary for uplink transmissions. All user terminals will be capable of receiving the control region of an MBSFN subframe. This is the main reason for which MBSFN subframes have been found useful for introducing new and non-backward compatible types of signaling and transmission modes in the new releases of LTE, without compromising the backward compatibility with earlier releases. The user terminals that are not capable of receiving transmissions in the MBSFN data region will simply ignore those transmissions.

Information about the set of subframes that are configured as MBSFN subframes in a cell is provided as part of the system information. In principle, an arbitrary pattern of MBSFN subframes can be configured with the pattern repeating after 40 ms. However, since physical signals and other essential information for system operation need to be transmitted in an orderly fashion to the UEs to find and attach to a cell, subframes where such information is mapped cannot be configured as MBSFN subframes. Therefore, subframes 0, 4, 5, and 9 for FDD and subframes 0, 1, 5, and 6 for TDD cannot be configured as MBSFN subframes, leaving the remaining six subframes as candidates for MBSFN subframes.

9.8 Resource structure, allocation, mapping

The minimum time–frequency resource unit used for downlink transmission is a resource element, which is defined as one sub-carrier over one OFDM symbol. For both TDD and FDD duplex schemes, a group of 12 sub-carriers contiguous in frequency over one slot in time form a PRB, i.e., a two-dimensional region corresponding to one slot in the time-domain and 180 kHz (i.e.,12×15 kHz) in the frequency-domain. The transmissions are allocated in units of PRB. More specifically, the data or control signaling in each slot is transmitted via one or several resource elements of $12 \times N_{RB}^{DL}$ sub-carriers over N_s^{DL} OFDM symbols. The value of the parameter N_{RB}^{DL} depends on the downlink transmission bandwidth configured for the cell, and is given as $\min(N_{RB}^{DL}) \leq N_{RB}^{DL} \leq \max(N_{RB}^{DL})$, where $\min(N_{RB}^{DL}) = 6$ for 1.4 MHz RF channels, and $\max(N_{RB}^{DL}) = 110$ for 20 MHz channels. The set of permissible values for N_{RB}^{DL} are given in Table 9.2. The number of OFDM symbols in a slot N_s^{DL} depends on the cyclic prefix length and the sub-carrier spacing configuration as given in Table 9.2.

There are a number of physical channels and physical signals that are mapped to predetermined (or configurable) locations in time and frequency over a radio frame that are used to assist proper operation of the system and the mobile terminals. The downlink physical channels correspond to a set of resource elements carrying data/control signaling information originating from higher protocol layers. The following downlink physical channels are defined in LTE/LTE-Advanced:

- Physical broadcast channel (PBCH): It carries the MIB portion of the system information. It is mapped to six resource blocks which are centered around

DC sub-carrier in subframe 0 and utilizes the resource elements which are not reserved for transmission of reference signals or other downlink control signaling. The BCH transport block is mapped to four subframes within a 40 ms interval.

- Physical control format indicator channel (PCFICH): It carries the size of PDCCH and is mapped to the first OFDM symbol in each downlink or special subframe. It contains the information on the number of OFDM symbols used for the downlink control region. The exact location of PCFICH is determined by the cell identity and system bandwidth.
- Physical hybrid-ARQ indicator channel (PHICH): It carries HARQ feedback corresponding to uplink data transmissions.
- Physical downlink control channel (PDCCH): It informs the UE(s) about transport format, resource allocation, HARQ feedback, and uplink grants. It is mapped to the first $L = 1, 2, \text{or } 3$ OFDM symbols in each downlink subframe. The number of the OFDM symbols for PDCCH is obtained via PCFICH. Multiple PDCCHs may be monitored by a UE. It also carries uplink power control commands.
- Physical downlink shared channel (PDSCH): It carriers downlink data transmission for different users as well as upper layer signaling.
- Physical multicast channel (PMCH): It is used to broadcast MBMS services.

In addition to the above physical channels, a set of downlink physical signals corresponding to a predetermined set of resource elements are specified and used by the physical layer which do not carry information originating from higher layers but are extremely important for proper LTE/LTE-Advanced system operation. The downlink physical signals include a variety of reference signals (also known as pilots which allow the UE to estimate the communication channel and coherently detect/decode data and control information that can be either cell-specific or UE-specific) and synchronization signals that provide timing and frequency synchronization, as well as physical identity of the cell to UEs attempting to select/reselect the cell. Figure 9.30 depicts the location of the downlink physical channels and signals in time and frequency within a type 1 frame structure.

9.8.1 Physical and virtual resource blocks

Each resource element is uniquely identified by the index pair (k, l) in a slot where $k = 0, \ldots, 12 \times N_{\text{RB}}^{\text{DL}} - 1$ and $l = 0, \ldots, N_{\text{s}}^{\text{DL}} - 1$ are the indices corresponding to frequency and time dimensions, respectively. Throughout this book, we assume that LTE/LTE-Advanced sub-carrier spacing is 15 kHz and there are 12 sub-carriers in each physical and virtual resource block (VRB) per slot. The structure of resource blocks in the downlink is shown in Figure 9.31. LTE/LTE-Advanced defines two types of resource blocks: physical and VRBs where in the latter, the resource blocks over a slot are permuted across frequency dimension to

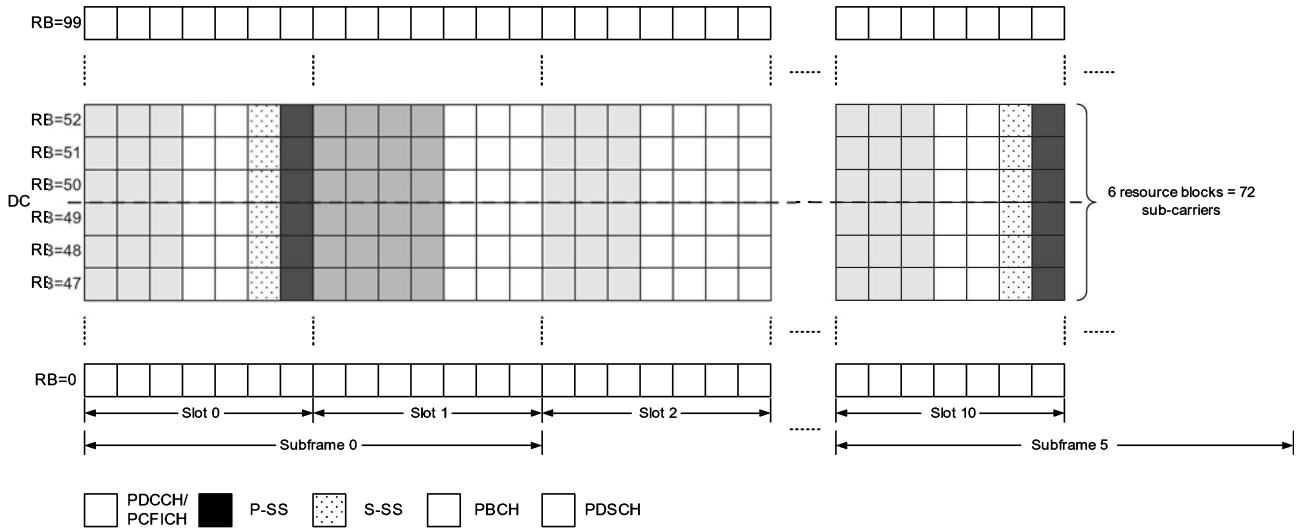


FIGURE 9.30

Physical resource structure and channel mapping (FDD, $N_s^{DL} = 7$, $N_{RB}^{DL} = 100$) [158].

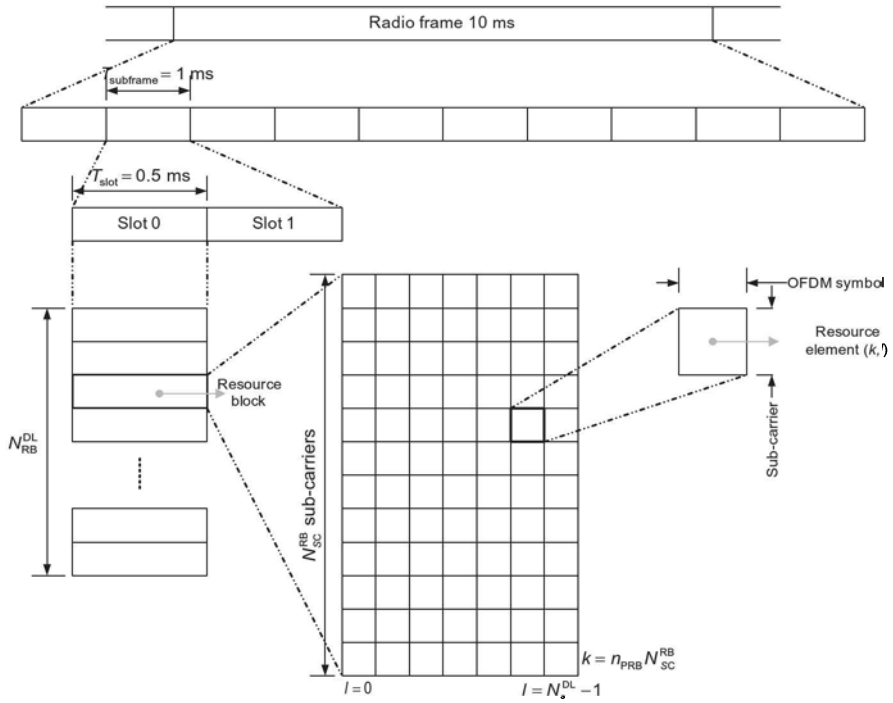


FIGURE 9.31

Resource block structure in time and frequency dimensions [3,162].

take advantage of frequency diversity. Another unit of resource allocation in the downlink is resource block group (RBG), which is a group of resource blocks. Allocating resources in RBG quanta reduces signaling overhead by reducing the number of bits required to uniquely address and identify each RBG relative to individual resource block identification/addressing. The size and number of RBGs depend on the channel bandwidth.

A PRB is defined as N_s^{DL} consecutive OFDM symbols in the time-domain and 12 (or 24 in the case of 7.5 kHz sub-carrier spacing) consecutive sub-carriers in the frequency-domain. A PRB consists of $12 \times N_s^{DL}$ resource elements corresponding to one slot in the time-domain and 180 kHz in the frequency-domain. The PRBs are numbered from 0 to $N_{RB}^{DL} - 1$ in the frequency-domain (see Figure 9.31). The relation between the PRB number n_{PRB} in the frequency-domain and resource elements (k, l) in a slot is given by $n_{PRB} = \lfloor k/12 \rfloor$.

A VRB has the same size as a PRB. There are two types of VRBs: VRBs of localized type and VRBs of distributed type. For each type of VRB, a pair of VRBs over two slots in a subframe is assigned together by a single VRB number n_{VRB}. VRBs of localized type are mapped directly to PRBs such that VRB index

n_{VRB} is identical to PRB index n_{PRB}. The VRBs are numbered from 0 to $N_{VRB}^{DL} - 1$, where $N_{VRB}^{DL} = N_{RB}^{DL}$. In the localized allocations, there is a one-to-one mapping between virtual and PRBs, whereas in the distributed case of resource allocation, the VRB numbers are mapped to PRB numbers according to a certain rule and inter-slot hopping is further utilized. In the latter, the first part of a VRB pair is mapped to one PRB and the other part of the VRB pair is mapped to another PRB which is at a predefined gap distance (which causes the inter-slot hopping) in order to achieve frequency diversity gain. This mechanism is especially important for small resource block allocations due to inherently less frequency diversity.

The distributed VRBs are mapped to PRBs such that consecutive VRBs are not mapped to sequential PRBs in the frequency-domain. A single VRB pair may be further split in the frequency-domain. Allocation of distributed resources includes a mapping from logically consecutive VRB pairs to physically non-consecutive PRB pairs to provide frequency diversity among VRB pairs. This is performed by a virtual block-interleaver or permutation scheme which operates on resource block pairs. The allocation of distributed resources further includes splitting of each resource-block pair such that the two resource blocks of the resource-block pair are transmitted with a certain frequency gap which provides additional frequency diversity for a single VRB pair. The latter can be interpreted as the introduction of frequency hopping on a slot basis. We will see later in this section that the associated PDCCH downlink control information (DCI) indicates whether allocated VRBs are of a localized or distributed type in the case of type 2 resource allocations. Thus, it is possible to dynamically switch between distributed and localized allocation, and also mix distributed and localized transmission for different terminals within the same subframe. The exact size of the frequency gap in Figure 9.32 depends on the overall downlink system bandwidth as given in Table 9.6. The gaps have been chosen based on the following criteria: (1) the gap size should be in the order of half of the total bandwidth (in terms of the number of resource blocks) in order to provide sufficient frequency diversity gain; and (2) the gap size should be an integer multiple of P^2, where parameter P is the size of a RBG for resource allocation types 0 and 1. The reason for this constraint is to ensure that distributed transmission as described above and transmissions based on downlink allocation types 0 and 1 can coexist on the same subframe. Due to the constraint that the gap size should be an integer multiple of P^2, the gap size N_{gap} may deviate from precisely half of the system bandwidth when the number of resource blocks is an odd number. In those cases, only a fraction of available resource blocks within the system bandwidth can be used for distributed transmission. An example of distributed resource allocation is shown in Figure 9.32.

The permissible values of gap size N_{gap} are given in Table 9.6. Note that for $6 \leq N_{RB}^{DL} \leq 49$, only one gap value $N_{gap} = N_{gap}^{(1)}$ is defined, whereas for $50 \leq N_{RB}^{DL} \leq 110$, two gap values $N_{gap}^{(1)}$ and $N_{gap}^{(2)}$ are defined where the choice of the gap value is signaled as part of the downlink scheduling assignment. VRBs of

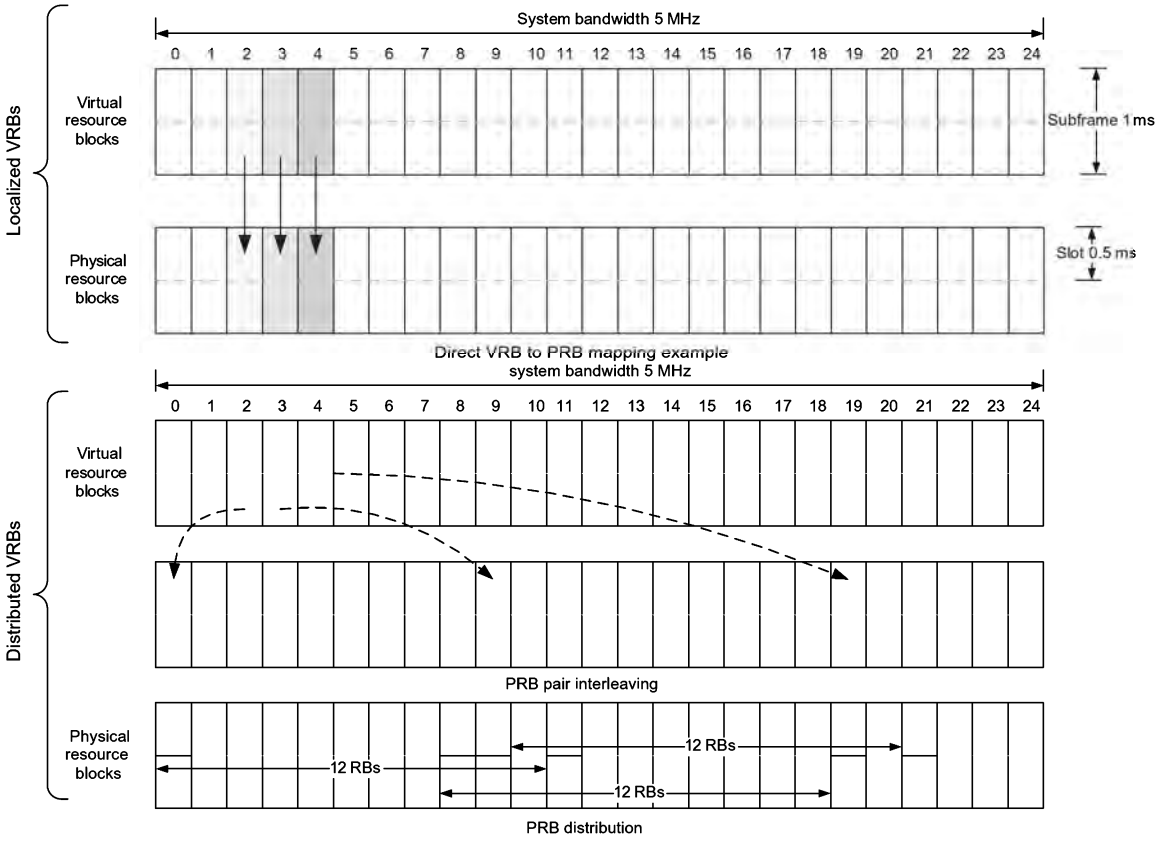


FIGURE 9.32

Illustration of localized and distributed VRB to PRB mapping ($N_{\text{VRB}}^{\text{DL}} = N_{\text{RB}}^{\text{DL}} = 25$) [15].

Table 9.6 Distributed VRB Gap Size as a Function of System Bandwidth [3,5]

Total Number of Resource Blocks N_{RB}^{DL}	Gap Size N_{gap}		Resource Block Group Size P
	$N_{gap}^{(1)}$	$N_{gap}^{(2)}$	
6–10	$\lceil N_{RB}^{DL}/2 \rceil$	–	1
11	4	–	2
12–19	8	–	2
20–26	12	–	2
27–44	18	–	3
45–49	27	–	3
50–63	27	9	3
64–79	32	16	4
80–110	48	16	4

distributed type are numbered from 0 to $N_{VRB}^{DL} - 1$, where $N_{VRB}^{DL} = 2\min(N_{gap}, N_{RB}^{DL} - N_{gap}^{(1)})$ or $N_{VRB}^{DL} = 2N_{gap}^{(2)} \left\lfloor N_{RB}^{DL}/2N_{gap}^{(2)} \right\rfloor$ depending on the value of N_{gap}.

Figure 9.33 illustrates the interleaving table for a 10 MHz channel bandwidth when using $N_{gap}^{(1)}$. The VRB indices are written to the table row-by-row and are sequentially read off the table column-by-column. The nulls in the table are ignored and do not generate outputs. When using $N_{gap}^{(1)}$, the VRBs associated with the first half of the output indices from the interleaving table are mapped to the lower set of PRBs starting from the first PRB. The VRBs corresponding to the second half of the output indices from the interleaving table are then mapped to the upper set of PRBs ending at the last PRB. This mapping is applied to the first slot of the subframe. The mapping is cyclically shifted by the gap value $N_{gap}^{(1)}$ for the second slot of the subframe [13]. As shown in Figure 9.34, when using $N_{gap}^{(2)}$ in a 10 MHz bandwidth, two interleaving tables each with six rows are used to define the mapping from VRBs to PRBs. There are some null entries that are inserted in each table [3].

To better understand the difference between the localized and distributed allocations, consider the example shown in Figure 9.35. The frequency response of a multipath fading channel varies with time and frequency due to path loss, shadowing, and user mobility effects. Let us assume that user A and user B are two users with different channel conditions. The base station receives channel quality reports (typically in the form of CINR or SINR measurements) from the mobile stations in the cell. The base station scheduler may adopt either of the following allocation schemes depending on the channel condition reports and other considerations that will be discussed later in this chapter. One allocation strategy is to allocate user A and user B in the sub-channels where the corresponding SINR is

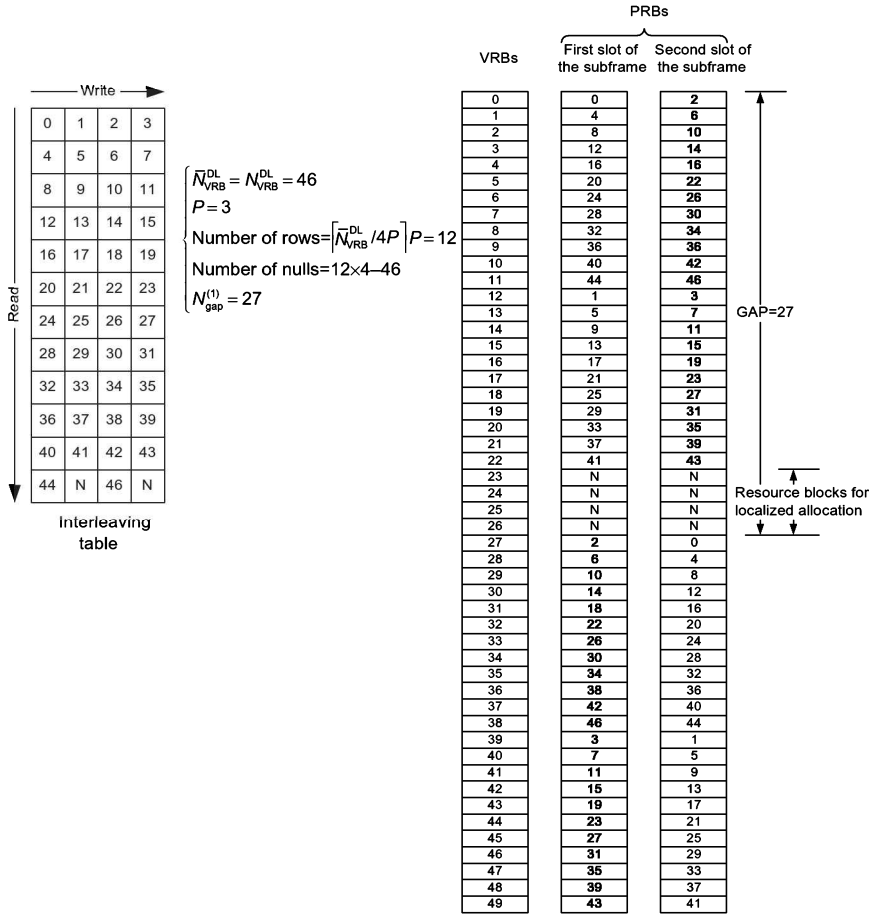


FIGURE 9.33

Illustration of distributed VRB to PRB mapping procedure in 10 MHz using $N_{gap}^{(1)}$ [13].

the best for that user. This is when a frequency-selective scheduling gain can be achieved through allocation of a group of physically adjacent sub-carriers in the sub-channels with relatively strong SINR for that user. Another strategy is to permute the sub-carriers over the entire channel, form a logical group of distributed sub-carriers and assign them to a user. In this case, frequency diversity gain can be achieved through use of distributed resource units for this user.

9.8.2 Resource allocation

One of the main design objectives for signaling of resource allocations, in terms of a set of resource blocks in each subframe, to the active UEs in the cell is to find a

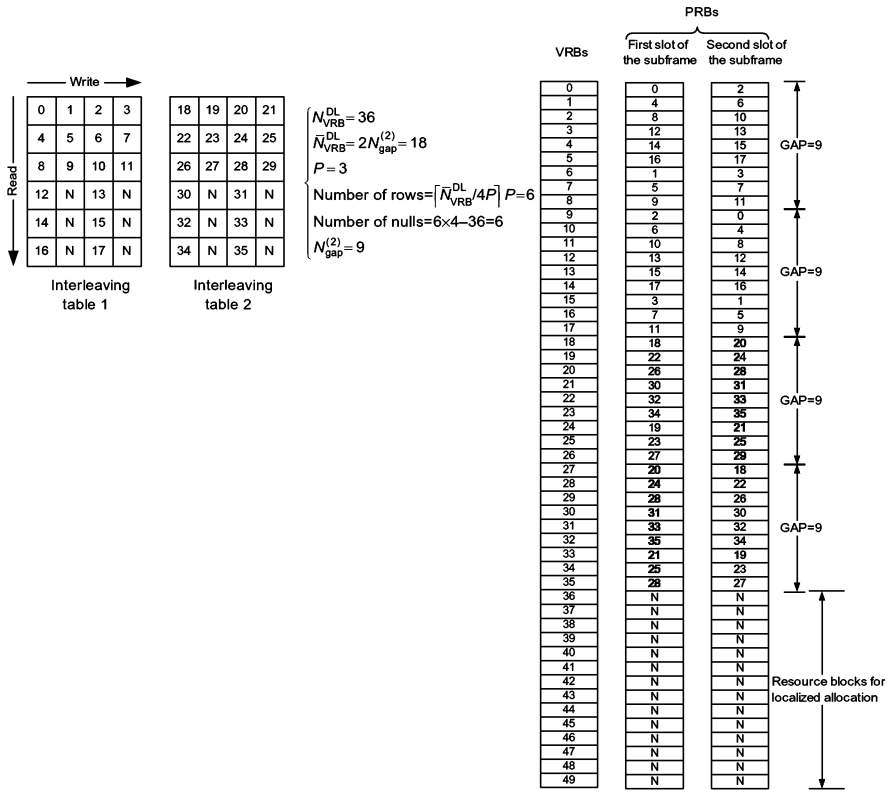


FIGURE 9.34

Illustration of distributed VRB to PRB mapping procedure in 10 MHz using $N_{gap}^{(2)}$ [13].

good trade-off between flexibility and signaling overhead. Indications of localized/distributed resource allocations to different UEs are transmitted via PDCCH. While the exact use of PDCCH depends on the scheduling algorithms implemented in the eNB, it is nevertheless possible to outline some general principles of this operation. The resource allocation field in downlink control information (carried via PDCCH or ePDCCH) is interpreted by the UE depending on the DCI format. A resource allocation field in each DCI includes two parts, a resource allocation header field and information pertaining to the actual resource block assignment. The DCI formats 1, 2, 2A, 2B, 2C and 2D with type 0, and DCI formats 1, 2, 2A, 2B, 2C and 2D with type 1 resource allocation (carried via PDCCH) have the same format and are distinguished from each other via the single bit resource allocation header field which exists depending on the downlink system bandwidth. The PDCCH with DCI format 1A, 1B, 1C, and 1D support type 2 resource allocation, while PDCCH with DCI format 1, 2, 2A, 2B, 2C and 2D support type 0 or type 1

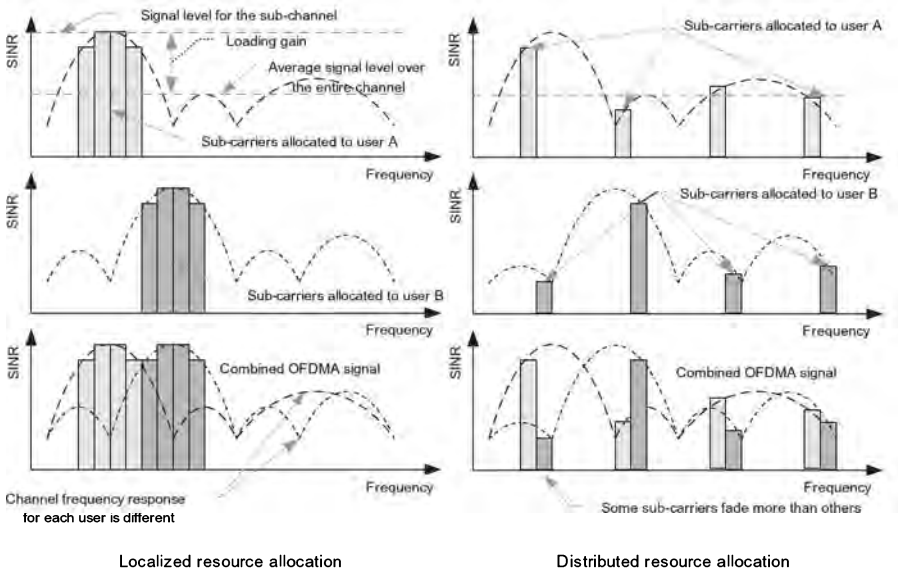


FIGURE 9.35

Comparison of localized and distributed allocation schemes (example) [18].

resource allocation. The DCI formats with a type 2 resource allocation do not have a resource allocation header field [5]. LTE/LTE-Advanced specify three resource allocation schemes in the downlink, where their usage depends on the number resources assigned to the UE, and whether a localized or distributed allocation must be used for the UE considering the channel conditions that the UE is experiencing. In LTE Rel-11, the DCI formats 1, 2, 2A, 2B, 2C and 2D with type 0 and DCI formats 1, 2, 2A, 2B, 2C and 2D with type 1 resource allocation (carried via ePDCCH) have the same format and are distinguished from each other via the single-bit resource allocation header field which exists depending on the downlink system bandwidth. The ePDCCH with DCI format 1A, 1B, and 1D have a type 2 resource allocation.

9.8.2.1 Resource allocation type 0

In resource allocation type 0, the resource block assignment information includes a bitmap indicating the RBGs that are allocated to the scheduled UE where an RBG is a set of consecutive VRBs of localized type. Note that in this case, the granularity of resource assignment is an RBG. The RBG size P is a function of the system bandwidth as given in Table 9.6. The total number of RBGs is given by $N_{\text{RBG}} = \lceil N_{\text{RB}}^{\text{DL}}/P \rceil$. The bitmap is of size N_{RBG} bits and each bit corresponds to a unique RBG in the set. The RBGs are indexed in the order of increasing

frequency and non-increasing RBG sizes starting at the lowest frequency (i.e., RBG_0 to $RBG_{N_{RBG}-1}$ are mapped to the most-significant-bit and least-significant-bit of the bitmap, respectively). The RBG is allocated to a UE, if the corresponding bit value in the bitmap is set to "1." Therefore, each bit in the bitmap will select a small contiguous group of resource blocks whose size depends on the bandwidth. The maximum number of resource blocks assigned using type 0 allocation is the entire bandwidth. As an example, consider a total of 50 resource blocks (i.e., 10 MHz system bandwidth), the number of bits in the bitmap is 17. Each bit in the 17-bit bitmap selects a group of three consecutive resource blocks, with the exception of the last group which will only contain two resource blocks, as shown in Figure 9.36.

9.8.2.2 Resource allocation type 1

In resource allocation type 1, a bitmap indicates PRBs inside a selected RBG subset $0 \leq p < P$. The maximum number of resource blocks assignable through type 1 resource allocation is a subset of the entire bandwidth, i.e., a type 1 allocation, even with all bits in the bitmap set to "1" does not span the entire bandwidth. Each bit in the bitmap will select a single resource block from a set of contiguous group of resources whose size and separation depend on the total bandwidth. In resource allocation type 1, the resource block assignment signaling consists of three parts: (1) select a RBG subset; (2) whether to apply an offset when interpreting the bitmap using a shift flag; and (3) the bitmap that indicates to the UE the specific PRBs inside the selected RBG subset.

The size of the bitmap used to address VRBs in a selected RBG subset is given by $N_{RB}^{(1)} = \lceil N_{RB}^{DL}/P \rceil - \lceil \log_2(P) \rceil - 1$. The addressable VRBs within a selected RBG subset start from an offset $n_{shift}(p)$ to the smallest VRB number within the selected RBG subset. The offset is interpreted in terms of the number of VRBs, and is given by $n_{shift}(p) = N_{RB}^{RBG_{subset}}(p) - N_{RB}^{(1)}$ (if the value of the shift bit is non-zero) where the parameter $N_{RB}^{RBG_{subset}}(p)$ denotes the number of VRBs in RBG subset p and is given as follows [5]:

$$N_{RB}^{RBG_{subset}}(p) = \begin{cases} P\left(\left\lfloor \frac{N_{RB}^{DL}-1}{P^2} \right\rfloor + 1\right) & p < \left\lfloor \frac{N_{RB}^{DL}-1}{P} \right\rfloor \bmod P \\ P\left\lfloor \frac{N_{RB}^{DL}-1}{P^2} \right\rfloor + (N_{RB}^{DL}-1) \bmod P + 1 & p = \left\lfloor \frac{N_{RB}^{DL}-1}{P} \right\rfloor \bmod P \\ P\left\lfloor \frac{N_{RB}^{DL}-1}{P^2} \right\rfloor & p > \left\lfloor \frac{N_{RB}^{DL}-1}{P} \right\rfloor \bmod P \end{cases} \tag{9.23}$$

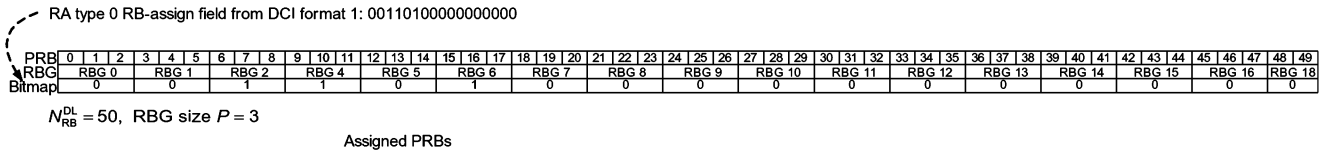


FIGURE 9.36

Example resource allocation type 0 [3,5].

When RBG subset p is addressed, the ith bit ($i = 0, 1, \ldots, N_{\text{RB}}^{(1)} - 1$) in the bitmap field corresponds to VRB number $n_{\text{VRB}}^{\text{RBG}_{\text{subset}}}(p) = \lfloor [i + n_{\text{shift}}(p)]/P \rfloor P^2 + pP + (i + n_{\text{shift}}(p)) \bmod P$. In comparison to resource allocation type 0, the bitmap size for type 1 is always shorter by $\lceil \log_2 P \rceil + 1$ number of bits. As an example, consider a total of 50 resource blocks, the number of bits in the bitmap is 14 (i.e., three bits shorter relative to type 0). Each bit in the 14-bit bitmap will select an individual resource block inside the selected subset, as demonstrated in Figure 9.37.

9.8.2.3 Resource allocation type 2

In resource allocation type 2, the resource block assignment information comprises a set of contiguously allocated localized VRBs or distributed VRBs. When the resource allocation is signaled in PDCCH/ePDCCH with DCI format 1A, 1B, or 1D, a 1-bit flag indicates whether localized VRBs or distributed VRBs are assigned (value "0" indicates localized and value "1" indicates distributed VRB assignment), while distributed VRBs are always assigned in the case of resource allocation signaled with DCI format 1C. A localized VRB assigned to a UE can vary from a single VRB up to a maximum number of VRBs over the system bandwidth. For DCI format 1A, the distributed VRBs allocated to a UE may vary from a single VRB up to $N_{\text{VRB}}^{\text{DL}}$, if the DCI CRC bits are scrambled with P-RNTI, RA-RNTI, or SI-RNTI. With DCI format 1B, 1D, and the CRC scrambled with C-RNTI, or with DCI format 1A and the CRC scrambled with C-RNTI, SPS C-RNTI or temporary C-RNTI, distributed VRB allocations allocated to a UE vary from a single VRB up to $N_{\text{VRB}}^{\text{DL}}$ VRBs, if $6 \le N_{\text{RB}}^{\text{DL}} < 50$, and vary from a single VRB up to 16 if $50 \le N_{\text{RB}}^{\text{DL}} < 110$. Using DCI format 1C, the number of distributed VRB allocations for a UE may range from $N_{\text{RB}}^{\text{step}}$ up to $\lfloor N_{\text{VRB}}^{\text{DL}}/N_{\text{RB}}^{\text{step}} \rfloor N_{\text{RB}}^{\text{step}}$ with an increment of $N_{\text{RB}}^{\text{step}}$, where $N_{\text{RB}}^{\text{step}} = 2$ if $6 \le N_{\text{RB}}^{\text{DL}} < 50$, and $N_{\text{RB}}^{\text{step}} = 4$ if $50 \le N_{\text{RB}}^{\text{DL}} < 110$ [5].

For PDCCH/ePDCCH with DCI format 1A, 1B, or 1D, the type 2 resource allocation field consists of a parameter referred to as resource indication value (RIV) which corresponds to the starting resource block index RB_{start} and the length of the allocation in terms of the number of virtually contiguous resource blocks L_{RB}. The RIV is defined by $\text{RIV} = N_{\text{RB}}^{\text{DL}}(L_{\text{RB}} - 1) + \text{RB}_{\text{start}}$ if $(L_{\text{RB}} - 1) \le \lfloor N_{\text{RB}}^{\text{DL}}/2 \rfloor$; otherwise $\text{RIV} = N_{\text{RB}}^{\text{DL}}(N_{\text{RB}}^{\text{DL}} - L_{\text{RB}} + 1) + (N_{\text{RB}}^{\text{DL}} - 1 - \text{RB}_{\text{start}})$, where $1 \le L_{\text{RB}} < N_{\text{VRB}}^{\text{DL}} - \text{RB}_{\text{start}}$. For DCI format 1C; however, the RIV in type 2 resource allocation field corresponds to starting resource block index $\text{RB}_{\text{start}} = \{0, N_{\text{RB}}^{\text{step}}, 2N_{\text{RB}}^{\text{step}}, \ldots, (\lfloor N_{\text{VRB}}^{\text{DL}}/N_{\text{RB}}^{\text{step}} \rfloor - 1)N_{\text{RB}}^{\text{step}}\}$ and the length in terms of virtually contiguous resource blocks $L_{\text{RB}} = \{N_{\text{RB}}^{\text{step}}, 2N_{\text{RB}}^{\text{step}}, \ldots, \lfloor N_{\text{VRB}}^{\text{DL}}/N_{\text{RB}}^{\text{step}} \rfloor N_{\text{RB}}^{\text{step}}\}$. In other words, in type 2 resource allocation, PRBs are not directly allocated. The type 2 allocation supports both localized and distributed VRB allocation which are distinguished by a flag. The information regarding the starting point of VRB and the length in terms of contiguously allocated VRB can be derived from RIV signaled within the DCI.

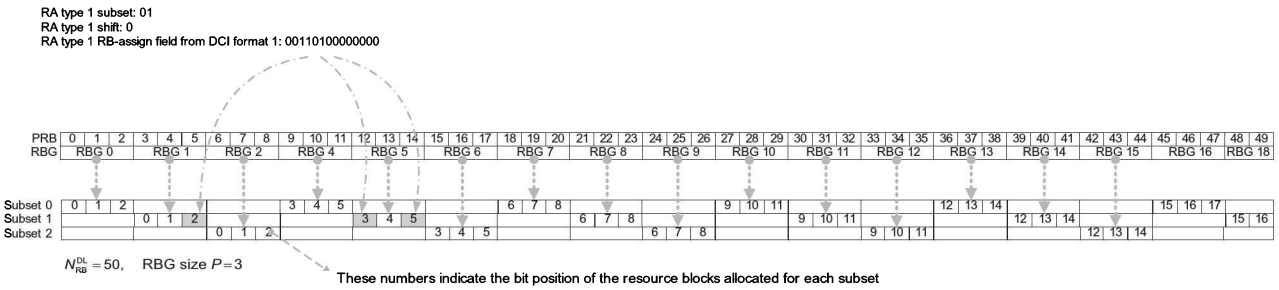


FIGURE 9.37

Example resource allocation type 1 [158].

As an example, consider a total of 50 resource blocks. Let us assume that the eNB is going to assign six localized VRBs starting from resource block 9 in the frequency-domain. Using the above equations for DCI format 1A, the RIV value is calculated as RIV = 259. This RIV is signaled in DCI format 1A, and the UE can unambiguously derive the starting resource block and the number of allocated resource blocks from RIV, as shown in Figure 9.38.

9.8.2.4 PRB bundling

When a UE is configured for operation in transmission mode 9, it may assume that the precoding granularity is multiple resource blocks in the frequency-domain when configured for PMI/RI reporting. In a given serving cell, if a UE is configured in transmission mode 10 and if PMI/RI is reported for all configured CSI processes, then the UE may assume that precoding granularity is multiple resource blocks; otherwise, the precoding granularity is going to be one resource block in the frequency-domain. Fixed bandwidth-dependent precoding resource block groups (PRGs) P' partition the available resource blocks and each PRG consists of a number of consecutive PRBs. If $N_{RB}^{DL} \bmod P' > 0$, then the size of one of the PRGs is adjusted to $N_{RB}^{DL} - P'\lfloor N_{RB}^{DL}/P' \rfloor$. The PRG size is in descending order starting at the lowest frequency index. The UE may assume that the same precoder is applied to all scheduled PRGs within the bundle. The size of the PRG depends on the system bandwidth, and is given in Table 9.7.

9.8.2.5 Resource element group

Resource element groups (REGs) are the basic unit of resource allocation for PCFICH, PHICH, and PDCCH that are represented by an index pair (k', l'), where k' is the sub-carrier index of the resource element within the REG with the lowest sub-carrier index, and l' is the OFDM symbol index of the REG. In other words, resource-element groups are used for defining the mapping of control channels to resource elements. The mapping of a symbol-quadruplet $\{a(i), a(i+1), a(i+2), a(i+3)\}$ to a resource-element group is defined such that elements $a(i)$ are mapped to resource elements (k, l) of the resource-element group not used for cell-specific reference signals (CRSs) in increasing order of indices k and l, as shown in Figure 9.39.

A REG always contains four consecutive resource elements which are not occupied by a CRS on any antenna port in use. All resource elements within a resource block in the first three OFDM symbols are allocated to REGs. Therefore, the number of resource elements within each REG and the number of REGs within an OFDM symbol are affected by the number of CRSs present on all antenna ports. The number and location of cell-specific reference signals are dependent on the number of antenna ports and the type of cyclic prefix used. Each antenna port has a unique CRS associated with it. Considering the dependency of REG arrangement on CRS configuration, the REG structure for one, two, or four antenna port

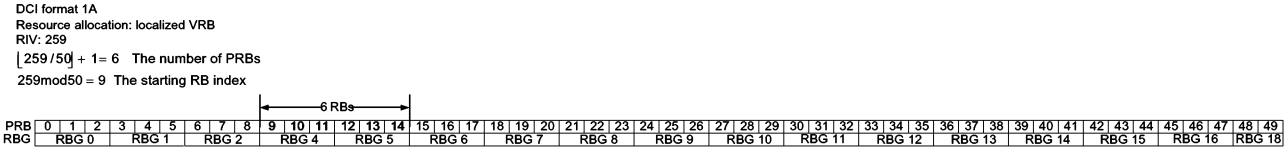


FIGURE 9.38

Example resource allocation type 2 [158].

Table 9.7 Precoding Resource Block Group Size [5]

Number of Resource Blocks N_{RB}^{DL}	Precoding Resource Block Group Size P'
≤10	1
11–26	2
27–63	3
64–110	2

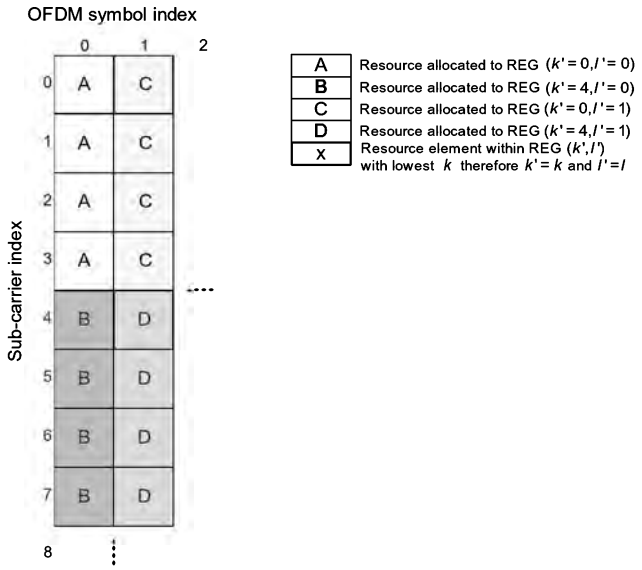


FIGURE 9.39

Resource element group structure [3,160].

configuration is different. However, the REG arrangement for each resource block within a subframe and for each antenna port is identical.

When antenna port 0 or antenna ports 0 and 1 are used, it is assumed that the CRSs are present on both antenna ports 0 and 1. This leads to an REG arrangement for each resource block on the first slot of each subframe, as shown in Figure 9.40. The CRSs are present within the first OFDM symbol. Since four resource elements not including CRSs are required to form a REG, the 12 resource elements over the first symbol in a resource block are divided into two REGs, each containing six resource elements (including two CRSs). On the

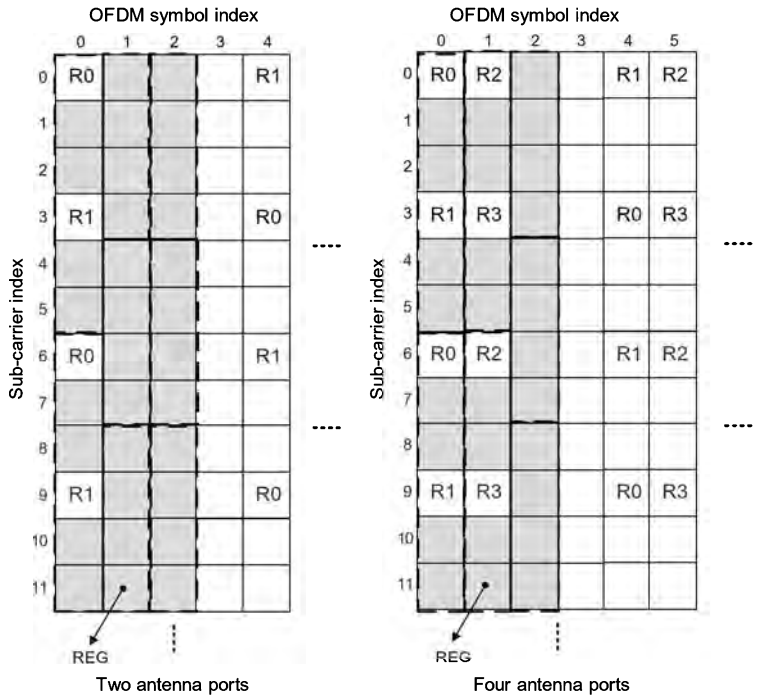


FIGURE 9.40

Resource element group structure with two and four antenna ports [3,160].

second and third OFDM symbols, no CRS is present; therefore; the 12 resource elements over each OFDM symbol are divided into three REGs, each containing four resource elements.

The REG organization in each resource block for four antenna port configuration is illustrated in Figure 9.40. The REG allocation within the first and second OFDM symbols is the same as one or two antenna port configuration. Four CRSs are present on the second OFDM symbol; therefore, eight resource elements are available for the mapping of control information. The 12 reference elements are divided into two REGs, each containing six resource elements. The third OFDM symbol contains no reference signals, so three REGs are available. An extended cyclic prefix subframe contains 12 OFDM symbols as opposed to 14 for the normal cyclic prefix. Since the number of CRSs in a normal or extended cyclic prefix subframe is the same, the limited number of OFDM symbols in an extended cyclic prefix subframe requires the OFDM symbol spacing of the CRSs to be reduced compared to a normal cyclic prefix. This reduction in spacing causes CRSs to be present within the fourth OFDM symbol of an extended cyclic prefix subframe, whereas in a normal cyclic prefix subframe no

CRSs are present on the fourth OFDM symbol of the subframe. The number of cell-specific reference symbols within the first three OFDM symbols is identical for normal or extended cyclic prefix, thus the REG configurations are identical.

9.9 Time and frequency synchronization

Synchronization signals are transmitted in every radio frame. The synchronization signals are used for time and frequency synchronization during initialization. As part of the system acquisition process, the UE synchronizes to the OFDM symbol, slot, subframe, half-frame, and radio frame (in that order) using a synchronization signal. Two synchronization signals are defined for system acquisition in E-UTRA: (1) a primary synchronization signal is used to obtain slot, subframe, and half-frame boundary. It also provides cell identity within the cell identity group. (2) A secondary synchronization signal is used to obtain the radio frame boundary. It also enables the UE to determine the cell identity group, which may range from 0 to 167. The synchronization signals are transmitted in the 0th and 5th subframes of every radio frame. In the frequency-domain, the synchronization signals occupy the center 1.08 MHz (equivalent to six resource blocks) of the radio channel (Figure 9.41).

Figure 9.41 shows the location of the primary and secondary synchronization signals in a radio frame. In the time-domain, primary and secondary synchronization signals are transmitted in the 0th and 10th slot (1st slot within 0th and 5th subframe) of every radio frame. Within these slots, the primary and secondary synchronization signals are transmitted in the last two symbols. The primary synchronization signal is transmitted in the last symbol of the 0th and 10th slot. The secondary synchronization signal is transmitted in one before the last symbol of the 0th and 10th slot. The PBCH is mapped to the first four symbols in slot 1 of the frame (just after the primary and secondary synchronization signals). In the frequency-domain, the center 1.08 MHz of the radio channel is reserved for primary and secondary synchronization signals. This is equivalent to six resource blocks or 72 sub-carriers. However, the primary and secondary synchronization signals only require 62 sub-carriers around the DC sub-carrier. The remaining sub-carriers (10 sub-carriers) in these resource blocks are not used for transmission. They provide a guard band between primary and secondary synchronization signal transmissions and other data transmissions.

There are 504 unique physical layer cell identities. The physical layer cell identities are grouped into 168 unique physical layer cell-identity groups, each group contains three unique identities. The grouping is done such that each physical layer cell identity is part of one and only one physical layer cell-identity group. A physical layer cell identity $N_{\mathrm{ID}}^{\mathrm{cell}} = 3N_{\mathrm{ID}}^{(1)} + N_{\mathrm{ID}}^{(2)}$ is thus uniquely defined by a number $N_{\mathrm{ID}}^{(1)}$ in the range of 0−167, representing the physical layer cell-identity group, and a number $N_{\mathrm{ID}}^{(2)}$ in the range of 0−2, denoting the physical layer identity within the physical layer cell-identity group.

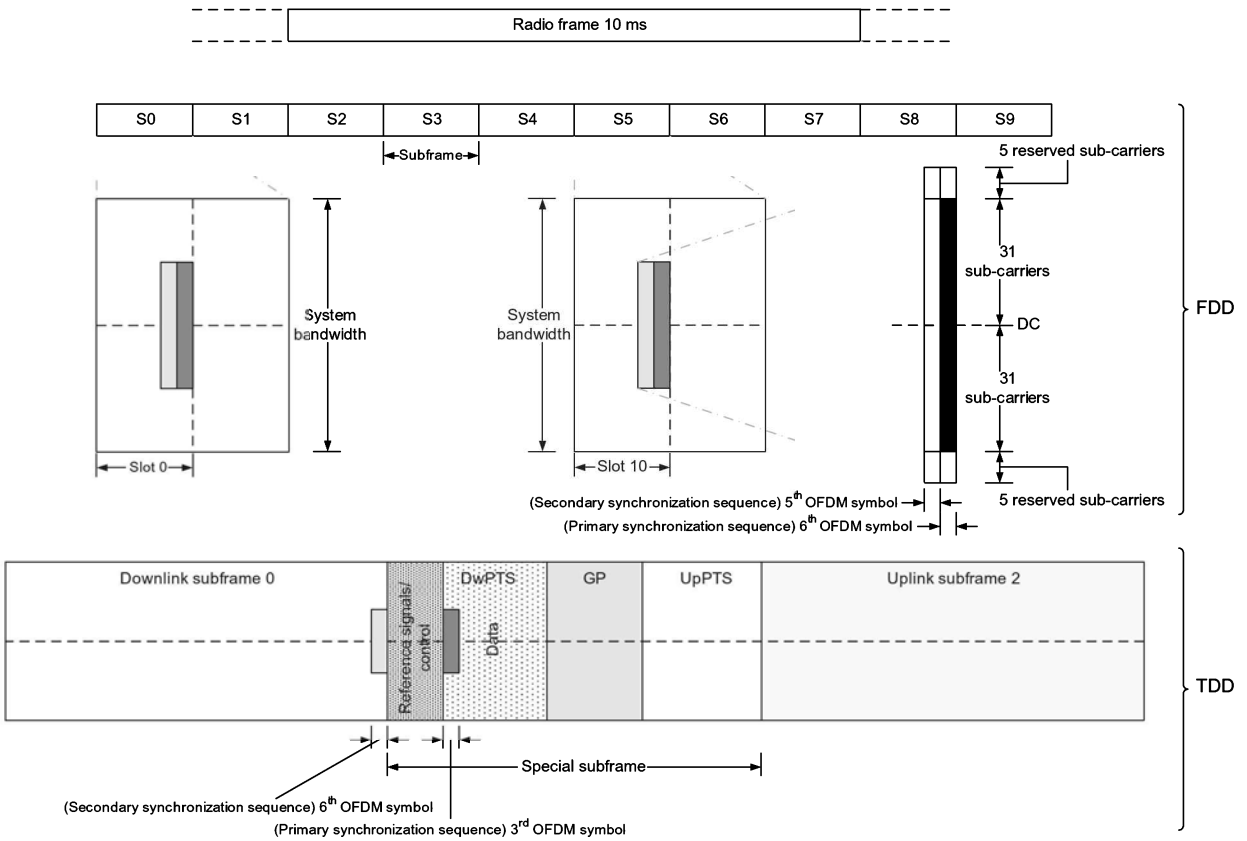


FIGURE 9.41

Location and structure of primary and secondary synchronization signals [3,174].

9.9.1 Primary synchronization signals

The primary synchronization signal is generated from a frequency-domain Zadoff–Chu sequence[9] according to the following equation [3]:

$$d_{\mathrm{u}}(n) = \begin{cases} \mathrm{e}^{-j\frac{\pi un(n+1)}{63}} & n = 0, 1, \ldots, 30 \\ \mathrm{e}^{-j\frac{\pi u(n+1)(n+2)}{63}} & n = 31, 32, \ldots, 61 \end{cases} \tag{9.24}$$

where the Zadoff–Chu root sequence index $u = 25, 29, 34$ if $N_{\mathrm{ID}}^{(2)} = 0, 1, 2$, respectively. The mapping to resource elements depends on the frame structure. The sequence $d(n)$ is mapped to the resource elements according to $a_{\mathrm{kl}} = d(n), n = 0, \ldots, 61$ and $k = n - 31 + 6N_{\mathrm{RB}}^{\mathrm{DL}}$. As shown in Figure 9.41, for frame structure type 1, the primary synchronization signal is mapped to the last OFDM symbol in slots 0 and 10. For frame structure type 2, the primary synchronization signal is mapped to the third OFDM symbol in subframes 1 and 6. The

[9] A Zadoff–Chu sequence is a complex-valued sequence with constant amplitude property whose cyclically shifted versions exhibit low cross-correlation. Thus, under certain conditions, the cyclically shifted versions of each sequence remain orthogonal to one another. A Zadoff–Chu sequence that has not been shifted is referred to as a root sequence. The uth root Zadoff–Chu sequence of prime length N is defined as follows [172]:

$$x_u(n) \triangleq \begin{cases} \mathrm{e}^{-\mathrm{j}[\pi un(n+1)]/N} & 0 \le n < N-1 \quad (N \text{ is an odd integer}) \\ \mathrm{e}^{-\mathrm{j}\pi un^2/N} & 0 \le n < N-1 \quad (N \text{ is an even integer}) \end{cases}$$

where N is an integer, which denotes the length of the Zadoff–Chu sequence. It is easily verified that $x_u(n)$ is periodic with period N, i.e., $x_u(n) = x_u(n+N), \forall n$. In other words, the sequence index u is prime relative to N. For a fixed value of u, the Zadoff–Chu sequence has an ideal periodic autocorrelation property (i.e., the periodic autocorrelation is zero for all time shifts other than zero). For different values of index u, the Zadoff–Chu sequences are not orthogonal, but rather exhibit low cross-correlation. If the sequence length N is selected as a prime number, there are different sequences with periodic cross-correlation of $1/\sqrt{N}$ between any two sequences regardless of time shift. The Zadoff–Chu sequences are a subset of constant amplitude zero autocorrelation (CAZAC) sequences. The properties of Zadoff–Chu sequences can be summarized as follows [172]:

- They are periodic with period N, if N is a prime number, i.e., $x_u(n+N) = x_u(n)$.
- Given N is a prime number, the DFT of a Zadoff–Chu sequence is another Zadoff–Chu sequence conjugated and time-scaled multiplied by a constant factor, i.e., $X_u[k] = x_u^*(vk)X_u[0]$ where v is the multiplicative inverse of u modulo N. It can be shown that $x_u^*(vk) = x_v^*(k)\mathrm{e}^{\mathrm{j}\pi(1-v)k/N}$.
- The autocorrelation of a prime-length Zadoff–Chu sequence with a cyclically shifted version of itself also yields zero autocorrelation sequence, i.e., it is non-zero only at one instant which corresponds to the cyclic shift zero.
- The cross-correlation between two prime-length Zadoff–Chu sequences, i.e., different u, is constant and equal to $1/\sqrt{N}$.
- The Zadoff–Chu sequences have low PAPR.

resource elements a_{kl} over OFDM symbols used for transmission of the primary synchronization signal where $k = n - 31 + 6N_{RB}^{DL}, n = \{-5, -4, \ldots, -1, 62, 63, \ldots, 66\}$ are reserved as guard sub-carriers.

9.9.2 Secondary synchronization signals

The secondary synchronization signal is generated from interlacing of two length-31 M-sequences[10] $s_0^{(m_0)}(n)$ and $s_1^{(m_1)}(n)$. There are 168 sequences that are generated from cyclic shifts of sequences $s_0^{(m_0)}(n)$ and $s_1^{(m_1)}(n)$. The cyclic shifts are generated based on cell-identity group number, which ranges from 0 to 167. The two sequences are scrambled, with the scrambling sequences $c_0(n)$ and $c_1(n)$ that are obtained based on the primary synchronization signal identity $N_{ID}^{(2)} \in \{0, 1, 2\}$. The second sequence is further scrambled with additional sequence $z_1^{(m_0)}(n)$. The two sequences are interleaved and mapped to 62 sub-carriers around the DC sub-carrier using BPSK modulation. As shown in Figure 9.42, five sub-carriers on either side are not used, providing a guard band between the primary and secondary synchronization sequences and other data-bearing sub-carriers. The secondary synchronization signals are also used to detect the radio frame boundary. To achieve this, a different variation of the sequence is generated by switching $s_0^{(m_0)}(n)$ and $s_1^{(m_1)}(n)$ in slot 10. The switch produces a different sequence so the UE can distinguish between slot 0 and slot 10.

More specifically, the secondary synchronization signal $d(0), d(1), \ldots, d(61)$ is an interleaved concatenation of two length-31 binary sequences as shown in Figure 9.42. The concatenated sequence is scrambled with a scrambling sequence given by the primary synchronization signal. The combination of two length-31 sequences defining the secondary synchronization signal differs between subframe 0 and subframe 5 according to the following rule:

$$d(2n) = \begin{cases} s_0^{(m_0)}(n)c_0(n) & \text{in subframe 0} \\ s_1^{(m_1)}(n)c_0(n) & \text{in subframe 5} \end{cases} \tag{9.25}$$

[10]Maximum length sequences are pseudo-random binary sequences that are generated using maximal linear feedback shift registers. The *M*-sequences are periodic and reproduce every binary sequence that can be reproduced by the shift registers (i.e., for length-*m* registers, they produce a sequence of length $2^m - 1$). An *M*-sequence is spectrally flat with the exception of a near-zero DC term. Since *M*-sequences are periodic and shift registers cycle through every possible binary value with the exception of the zero vectors, the registers can be initialized to any state with the exception of the zero vectors. A binary polynomial over *GF*(2) can be associated with the linear feedback shift register. The degree of polynomial is equal to the length of the shift register and the coefficients that are either 0 or 1 correspond to the taps of the register [173].

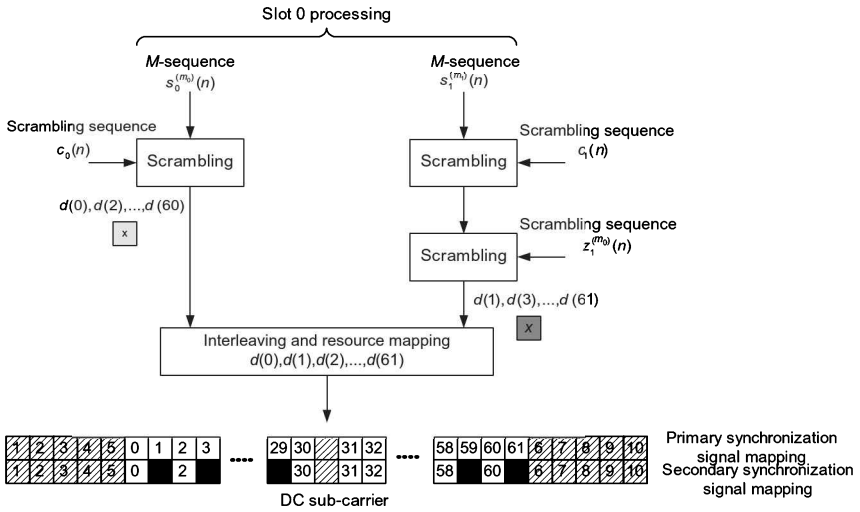


FIGURE 9.42

Secondary synchronization signal generation and mapping.

and:

$$d(2n+1) = \begin{cases} s_1^{(m_1)}(n)c_1(n)z_1^{(m_0)}(n) & \text{in subframe 0} \\ s_0^{(m_0)}(n)c_1(n)z_1^{(m_1)}(n) & \text{in subframe 5} \end{cases} \tag{9.26}$$

where $0 \leq n \leq 30$ and indices m_0 and m_1 are derived from the physical layer cell-identity group $N_{\text{ID}}^{(1)}$. The two sequences $s_0^{(m_0)}(n)$ and $s_1^{(m_1)}(n)$ are derived from two different cyclic shifts of an M-sequence $\tilde{s}(n)$. The two scrambling sequences $c_0(n)$ and $c_1(n)$ depend on the primary synchronization signal and are characterized by two different cyclic shifts of the M-sequence $\tilde{c}(n)$. The scrambling sequences $z_1^{(m_0)}(n)$ and $z_1^{(m_1)}(n)$ are defined by a cyclic shift of the M-sequence $\tilde{z}(n)$. The mapping of the sequence to resource elements depends on the frame structure. In a subframe for frame structure type 1 and in a half-frame for frame structure type 2, the same antenna port used for transmission of the primary synchronization signal is used for transmission of the secondary synchronization signal. The sequence $d(n)$ is mapped to resource elements according to $a_{\text{kl}} = d(n)$, $\forall n = 0, \ldots, 61$ and $k = n - 31 + 6N_{\text{RB}}^{\text{DL}}$, where $l = N_{\text{s}}^{\text{DL}} - 2$ in slots 0 and 10 for frame structure type 1 and $l = N_{\text{s}}^{\text{DL}} - 1$ in slots 1 and 11 for frame structure type 2 [3].

The primary synchronization signals are detected using non-coherent detection, whereas the secondary synchronization signals can be detected using either coherent or non-coherent detection schemes. Figure 9.43 shows an example algorithm for coherent detection of the secondary synchronization signals. The channel compensation block is used for coherent detection. The received secondary synchronization signals in the frequency-domain can be expressed as

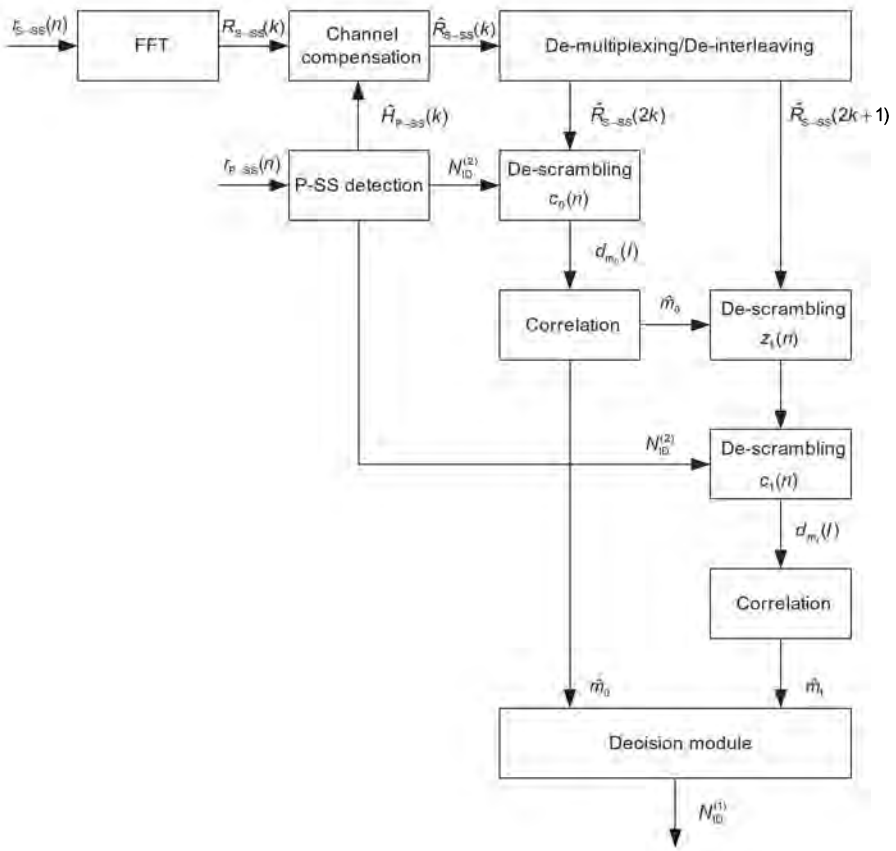


FIGURE 9.43

Example algorithm for coherent detection of secondary synchronization signal [174].

$r_{S-SS}(k) = H(k)d_{S-SS}(k) + w(k)$, where $d_{S-SS}(k)$ is the secondary synchronization signal in the frequency-domain, $H(k)$ is channel frequency response during secondary synchronization signal transmission, and $w(k)$ is the additive white Gaussian noise. For coherent detection, the UE estimates the channel frequency response $\hat{H}(k)$ by using the received primary synchronization sequence. From the UE perspective, the secondary synchronization signal detection is performed after the primary synchronization signal detection; thus the channel can be assumed to be known based on the primary synchronization signal sequence. In the frequency-domain, channel compensated secondary synchronization signal is given by $\hat{R}_{S-SS}(k) = \hat{H}(k)R_{S-SS}(k)$. After channel compensation, the de-interleaved and de-scrambled signal can be expressed as $d_{m_0}(l) = \hat{R}_{S-SS}(2k)c_0(k), d_{m_1}(k) = \hat{R}_{S-SS}(2k+1)c_1(k)z_1^{(\hat{m}_0)}(k)$. The output of coherent detection is the cross-correlation

between the de-scrambled signal and the cyclic shift of a secondary synchronization sequence that can be represented as follows:

$$\hat{m}_0 = \arg \max_i \left| \sum_{l=0}^{30} d_{m_0}(l) s_i(l) \right|^2$$
$$\hat{m}_1 = \arg \max_i \left| \sum_{l=0}^{30} d_{m_1}(l) s_i(l) \right|^2 \quad (9.27)$$

where $\hat{m}_0$ and $\hat{m}_1$ are the ith secondary synchronization signal sequence. Once $\hat{m}_0$ and $\hat{m}_1$ are identified, the frame timing and cell identity group $N_{\mathrm{ID}}^{(1)}$ can be detected (Figure 9.44).

The E-UTRA supports normal and extended cyclic prefix types for various deployment scenarios. The timing of the secondary synchronization signals will change depending on the cyclic prefix type, as shown in Figure 9.44. Prior to successful detection of the secondary synchronization signal, the cyclic prefix type is unknown to the UE and it must be blindly detected by examining the secondary synchronization signals at two possible locations in time. A simple algorithm for determining the cyclic prefix type is introduced in this section as an example. Note that E-UTRA specifications do not specify the receiver architecture and detection methods. This algorithm is based on the principle developed by Van de Beek, where the correlation between the cyclic prefix and its repetition part of the data symbol is calculated for two choices of cyclic prefix type [175]. For each cyclic prefix type, the correlation function is normalized by the associated received secondary synchronization signal power, because the range of values over which the summation is calculated varies with the cyclic prefix type. Let $r(n)$ denote the secondary synchronization signal at the receiver in the time-domain, we define [174]:

$$\rho_i = \sum_{n=0}^{L_i-1} r(n+\tau_i) r^*(n+\tau_i+N)$$
$$P_i = \sum_{n=0}^{L_i-1} |r(n+\tau_i+N)|^2 \quad (9.28)$$

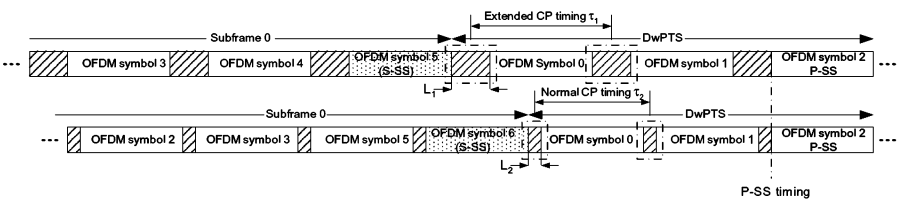


FIGURE 9.44

Illustration of the cyclic prefix detection algorithm [174].

where τ_1 and τ_2 represent the normal and extended cyclic prefix timing, respectively, and L_1 and L_2 denote the corresponding cyclic-prefix lengths. The normalized correlation function $U_i = \mathrm{Re}\{\rho_i\}/P_i$ with the larger value indicates the type of cyclic prefix. Therefore, a normal cyclic prefix is considered, if $U_1 > U_2$ and an extended cyclic prefix is assumed, if $U_2 > U_1$. The main advantage of this algorithm is that a correlation value can be accumulated by using an OFDM symbol within one radio frame in order to improve the detection performance at low SNR. In addition, this algorithm can reduce the computational complexity compared to blind detection using secondary synchronization signal detection at the two possible instances in time.

9.10 Downlink reference signals and channel estimation

To facilitate estimation of the multipath communication channel and reliable, coherent detection of data/control signaling, an OFDM system uses reference signals (or pilot symbols). The pilot symbols provide an estimate of the channel frequency response at the pilot locations over the time–frequency grid. It is possible to estimate the channel at other time–frequency locations using interpolation techniques. In E-UTRA, the CRSs are assigned to specific positions within a subframe, depending on the physical cell identity and the specific antenna port from which they are transmitted. The unique positioning of the pilots ensures that they do not interfere with each other and can be used to provide a reliable estimate of the complex-valued channel gains scaling each resource element within the transmitted grid. Using known pilot symbols to estimate the channel matrix, it is possible to equalize the effects of the channel and to reduce noise and interference effects of the received resource blocks. Downlink spatial multiplexing in LTE Rel-10 was enhanced to support up to eight transmission layers, together with an enhanced reference-signal structure. Relying on cell-specific reference signals for higher order spatial multiplexing is less attractive, as the reference signal overhead is not proportional to the instantaneous transmission rank, but rather to the maximum supported transmission rank. LTE Rel-10, therefore, introduced extensive support of UE-specific reference signals for demodulation of up to eight layers. In addition, feedback of channel-state information (CSI) is based on a separate set of reference signals, known as CSI reference signals (CSI-RSs). The CSI-RSs are relatively sparse in frequency but regularly transmitted from all antennas at the base station, while the UE-specific reference signals are denser in frequency but only transmitted when data is transmitted on the corresponding layer. Separating the reference-signal structures supporting demodulation and channel estimation helps to reduce reference-signal overhead, especially for high degrees of spatial multiplexing, and allows for implementation of various beam-forming schemes. In LTE Rel-8, the reference signal is added to the signal after precoding, one CRS per antenna port. From the received CRS, the UE can estimate the channel matrix and determine how the radio channel affected the signal. Using the estimated channel matrix

together with the knowledge about the code-book used for precoding, the UE can demodulate the received signal and regenerate the original information. In LTE Rel-10, the demodulation reference signals (DM-RSs) are added to the different data streams before precoding. The knowledge about the reference signal will then provide information about the combined effect of radio channel and precoding, thus prior knowledge about the precoder is not required by the receiver, this case is referred to as non-codebook-based precoding.

The LTE/LTE-Advanced systems make use of reference signals for downlink and uplink channel estimation. The downlink reference signals are used to measure CSI and channel estimation for demodulation. The design of downlink reference signals in LTE Rel-8 primarily relied on a set of cell-specific reference signals that could be used for both purposes. The CRS are wideband reference signals transmitted across the entire downlink system bandwidth in every subframe, and are defined for up to four antenna ports. The presence of cell-specific reference signals in every subframe, while necessary for decoding essential physical channels, results in unreasonably high overhead. To support eight-layer spatial multiplexing in LTE-Advanced, new low duty-cycle wideband reference signals, known as CSI-RSs, were designed for CSI measurements. Furthermore, a new set of UE-specific or DM-RSs were designed to enable channel estimation for demodulation. The LTE Rel-10 DM-RS design efficiently supports high-rank transmissions because it is present only in those PRBs in which higher rank is used [121]. The enhanced PDCCH (ePDCCH) DM-RSs with an identical structure to Rel-10 DM-RSs were added in Rel-11 to support various use cases involving the new FDM-based control channel structure.

LTE/LTE-Advanced defines an antenna port as a logical entity which is distinct from a physical antenna and is associated with a specific set of reference signals such that the channel over which a symbol on the antenna port is conveyed can be distinguished from the channel over which another symbol on the same antenna port is conveyed. There is one resource grid per antenna port. The set of downlink antenna ports supported depends on the reference signal configuration in the cell as given in Table 9.8 [3].

The set of downlink antenna ports supported depends on the reference signal configuration in the cell as follows [3]:

- Cell-specific reference signals support one, two, or four antenna port configurations and are transmitted on antenna ports $p = 0$, $p \in \{0, 1\}$, and $p \in \{0, 1, 2, 3\}$, respectively.
- MBSFN reference signals are exclusively transmitted on antenna port $p = 4$.
- UE-specific or DM-RSs are transmitted on antenna port $p = 5, 7,$ or 8, or one or several of $p \in \{7, 8, 9, 10, 11, 12, 13, 14\}$.
- Positioning reference signals are exclusively transmitted on antenna port $p = 6$.
- CSI reference signals support configuration of one, two, four, or eight antenna ports, and are transmitted on antenna ports $p = 15$, $p = \{15, 16\}$, $p = \{15, 16, 17, 18\}$, and $p = \{15, 16, 17, 18, 19, 20, 21, 22\}$, respectively.

Table 9.8 Downlink Antenna Ports and Their Use Cases [3,13]

3GPP Release	Downlink Antenna Port	Associated Reference Signals	Use Cases
8	0–3	Cell-specific reference signals	Single-layer transmission, transmit diversity, SU-MIMO/MU-MIMO
8	4	MBSFN reference signal	Multimedia broadcast and multicast services
8	5	UE-specific reference signal/ demodulation reference signal	Single-layer beamforming
9	6	Positioning reference signal	Location-based services
9	7,8	UE-specific reference signals/demodulation reference signals	Dual-layer beamforming, MU-MIMO
10	9–14	UE-specific reference signals/demodulation reference signals	Multi-layer beamforming, MU-MIMO
10	15–22	CSI reference signals	CSI measurement and reporting
11	107–110	ePDCCH demodulation reference signals	ePDCCH-based transmission and demodulation

- ePDCCH DM-RSs are transmitted on antenna ports $p \in \{107, 108, 109, 110\}$, and are used to support ePDCCH transmission and demodulation.

Based on the above discussion, six types of downlink reference signals are defined as follows, which will be discussed in detail in the next sections:

- Cell-specific reference signals (CRS)
- MBSFN reference signals
- UE-specific or DM-RSs associated with PDSCH
- Positioning reference signals
- CSI-RSs
- DM-RSs associated with ePDCCH

Note that there is one reference signal transmitted per downlink antenna port.

9.10.1 Reference signal design criteria

In this section, we discuss the theoretical criteria for efficient pilot pattern design for OFDM systems. Pilot-based channel estimation is defined as the use of the channel samples estimated at the pilot tones to reconstruct channel samples at the

remaining data tones. As a result, the pilot pattern design is essentially a conventional sampling rate selection problem in the two-dimensional signal processing space. In order to avoid aliasing during reconstruction of the channel time–frequency function, the pilot tone selection should follow the two-dimensional sampling theorem. When multiple antennas are used, the receiver must estimate the channel impulse response (or the transfer function) from each of the transmit antennas in order to correctly detect the signal. This is achieved through distributing reference signals (or pilot tones) among the transmit antennas.

Let $\Delta f = 1/T_{\mathrm{u}}$ and $T_{\mathrm{symbol}} = T_{\mathrm{g}} + T_{\mathrm{u}}$ denote the sub-carrier spacing and the OFDM symbol duration (inclusive of the guard interval), respectively. Let us further assume that the pilot sub-carriers are transmitted at integer multiples of sub-carrier spacing and OFDM symbol duration in frequency and time directions, respectively (i.e., $f_{\mathrm{p}} = m/T_{\mathrm{u}}$ and $T_{\mathrm{p}} = nT_{\mathrm{symbol}}$, where m and n are integers). The (m, n) pair represents the pilots' separation in terms of sub-carrier spacing and OFDM symbol duration. From the sampling theorem point of view, the channel's two-dimensional delay-Doppler response $h(\tau, \nu)$ can be fully reconstructed, if the two-dimensional transform function $H(t,f)$ is sampled greater than or equal to the Nyquist rate across time and frequency dimensions. Hence, the time-domain sampling rate must be greater than or equal to the channel's maximum Doppler spread, i.e., $T_{\mathrm{p}} \leq 1/\nu_{\max}$ (the sampling rate in time must be less than the coherence time) and the frequency-domain sampling rate must be greater than or equal to the channel's maximum delay spread, i.e., $f_{\mathrm{p}} \leq 1/\tau_{\max}$ (the sampling rate in frequency must be less than coherence bandwidth). Assuming a wide-sense stationary uncorrelated scattering channel model, and assuming the channel to be constant over one OFDM symbol, the frequency response $H(t,f)$ of the L-path channel is given as [23,140]:

$$H(t,f) = \frac{1}{\sqrt{L}} \sum_{l=1}^{L} \mathrm{e}^{\mathrm{j}(\psi_l + 2\pi\nu_l t - 2\pi f \tau_l)} \tag{9.29}$$

where ψ_l, ν_l, and τ_l denote the phase, Doppler frequency, and delay of the lth path, respectively. All these parameters are independent random variables. In general, the pilot signals are over-sampled in order to ensure a good trade-off between performance and overhead. Therefore, the choice of (m, n) depends on the channel's maximum delay spread and maximum Doppler spread, and must satisfy the following equation according to the two-dimensional sampling theorem [23]:

$$\begin{aligned} n &\leq \frac{1}{2T_{\mathrm{symbol}}\nu_{\max}} \\ m &\leq \frac{T_u}{2\tau_{\max}} \end{aligned} \tag{9.30}$$

It should be noted that the pilot density for a regular pattern can be calculated using Eq. (9.30). From Eq. (9.30), it can be seen that for large values of $\nu_{\max}$

(i.e., large Doppler spread, or alternatively small channel coherence time, means that the channel is time-varying), n should be small in order to appropriately track channel time variations. On the other hand, for large values of τ_{max} (i.e., large delay spread, or alternatively small coherence bandwidth, means that the channel is frequency selective), m should be small in order to closely follow channel frequency variation. In a regularly spaced pilot pattern, the pilot symbols are evenly spaced in frequency and in time.

Using the default OFDM symbol duration and the sub-carrier spacing given in Table 9.2 (OFDM parameters $T_{symbol} = 71.36\ \mu s$ and $\Delta f = 1/T_u = 15$ kHz), and assuming that the channel maximum delay spread $\tau_{max} = 5\ \mu s$ for the radio channels under consideration in LTE/LTE-Advanced evaluation methodology, the most appropriate range for the values of (m, n) can be determined. Since an E-UTRA system is required to support vehicle speeds up to 350 km/h, the maximum Doppler spread is estimated as $\nu_{max} \approx 810$ Hz at 2.5 GHz carrier frequency. From Eq. (9.30) and the above design parameters, the following pilot spacing constraints are obtained $n \leq 8, m \leq 6$. Figure 9.45 illustrates a regularly spaced

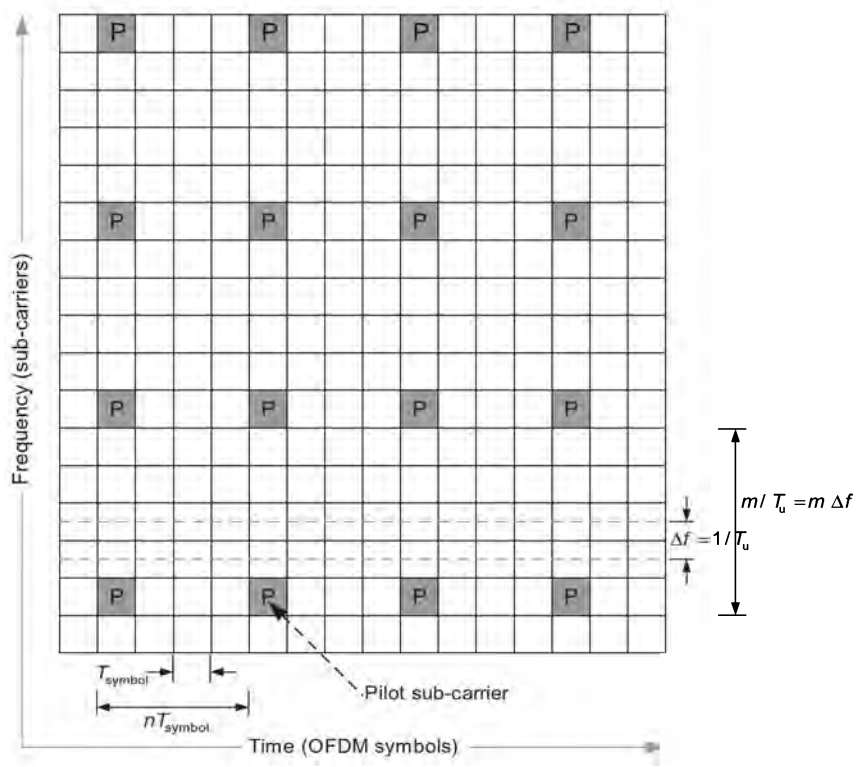


FIGURE 9.45

An example of regular pilot pattern for one spatial layer with $n = 4$ and $m = 5$.

two-dimensional pilot pattern for one spatial stream in the time–frequency domain where $n = 4$ and $m = 5$. The above criteria can be generalized to the case of multi-antenna transmission where each transmit antenna is associated with a group of pilot sub-carriers.

In the case of irregular pilot patterns, the pilot symbols can be irregularly placed in time, in frequency, or in both. The irregular patterns can be designed according to a certain criteria, yielding a specific pattern. Other important considerations in the design of pilot structures for support of multi-antenna OFDM systems include the following:

- Identical pilot pattern for each PRB: Pilot spacing in time and frequency must be in proportion with the resource block size. If the location of the pilot tones within a resource block is not maintained the same across all resource blocks within the system bandwidth, or alternatively the pilot sub-carriers have different positions within each resource block, the filtering and interpolation operations during channel estimation become excessively complex.
- Pilot density: Proper pilot overhead must be considered as a trade-off between accurate channel estimation under various mobility conditions and higher throughput.
- Types of pilots: Pilots are typically classified as common (cell-specific) and dedicated (UE-specific). Each UE in the cell can estimate the channel over the entire bandwidth using the common pilots, while dedicated pilots in the UE assigned resource block can be used for channel estimation. In conventional MIMO schemes, the dedicated pilots are typically adaptively precoded. The dedicated pilots can be used in conjunction with common pilots, or in place of common pilots.
- Pilot power boosting: In general, the power of pilot sub-carriers is boosted relative to data sub-carriers in order to enhance the channel estimation with pilot hopping or shifting. To avoid symbol to symbol power fluctuation when using pilot boosting, it is desirable to place the pilot sub-carriers regularly on every symbol, or to employ power adjustment for data sub-carriers over a symbol.
- Per antenna power balance: When using multiple transmit antennas, it is important to maintain balanced power distribution across antennas. For this purpose, if there are two transmit antennas, each OFDM symbol should contain the same number of pilot tones associated with different antenna ports.

As an example and by taking the above additional criteria into consideration, the structure and link-level performance (i.e., bit error rate versus SNR curves) of several hypothetical common pilot patterns for two transmit antennas and two different multi-antenna transmission schemes are shown in Figure 9.46. In this example, the pilot overhead of type A and type B patterns is 11.11% (i.e., 5.5% per antenna port), and the transmit power per antenna is balanced. The type C

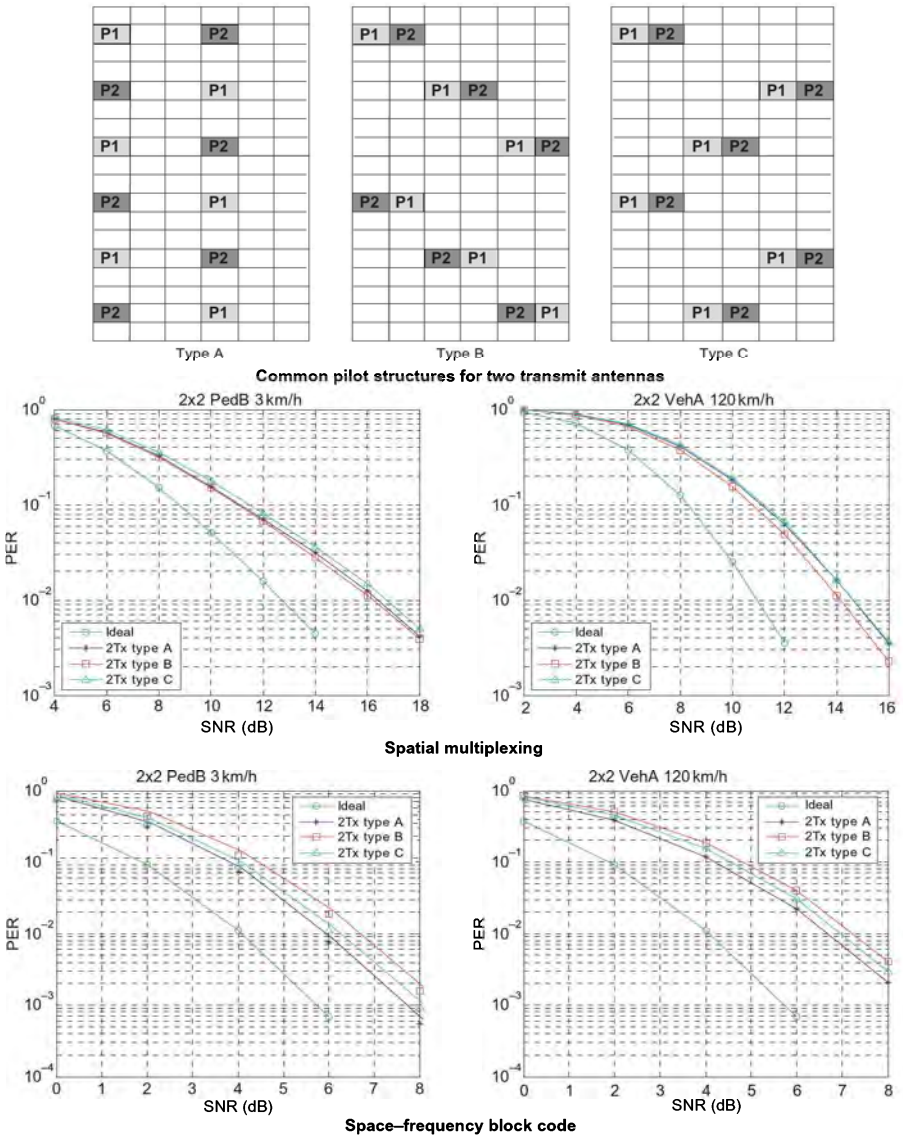


FIGURE 9.46

Channel estimation accuracy of various pilot patterns for two transmit antennas and QPSK-1/2 in different mobility and channel conditions [140].

pilot pattern has unbalanced power across antennas and symbols. The pilot tone powers have been boosted by 3 dB relative to data sub-carriers [140].

The accuracy of channel estimation using common pilots affects the performance of downlink/uplink data and control channel decoding. Figure 9.46 shows the link-level performance of spatial multiplexing with the example pilot patterns, where type B provides the best performance under low and high mobility conditions. The ideal channel estimation is used as a reference. Also shown in Figure 9.46 is the link-level performance of space-frequency block codes (SFBCs) with the example pilot patterns. In this case, type A outperforms the other structures. It can be concluded that type A is the best among three example patterns. Note that the example pilot patterns satisfy the sampling theorem constraint of Eq. (9.30). This example underlines the criteria for the design of robust and high performance pilot structures for OFDM-based cellular system operation in a wide range of mobility and channel conditions. The time/frequency characteristics of a number of typical radio channel models, mobility classes, and the associated time–frequency sampling rate requirements are summarized in Table 9.9.

Table 9.9 Characteristics of Typical Radio Channel Models ($T_u = 66.67$ μs and $T_{symbol} = 71.36$ μs)

Channel Model	RMS Delay Spread (ns) ν_{RMS}	Maximum Delay Spread (μs) τ_{max}	Maximum Sampling Rate in Frequency (Number of Sub-carriers) $\frac{T_u}{2\tau_{max}}$
ITU-T Pedestrian A (PedA)	45	0.41	81
ITU-T Pedestrian B (PedB)	750	3.7	9
ITU-T Vehicular A (VehA)	370.4	2.51	13
ITU-T Vehicular B (VehB)	4000	20	1
Typical Urban 6-Ray Model (TU)	1000	5	6

Vehicular Speed Range (km/h)	Maximum Doppler Spread (Hz) ν_{max}	Coherence Time (ms) T_c	Maximum Sampling Rate in Time (Number of OFDM Symbols) $\frac{1}{2T_{symbol}\nu_{max}}$
0–15	~27.78	15.23	252
15–120	~222.22	1.90	31
120–350	~648.15	0.65	10

9.10.2 Pilot-based channel estimation

While perfect knowledge of the radio channel can be used to find an upper bound for system performance, such knowledge is not available in practice and the channel needs to be frequently estimated. Channel estimation can be performed in various ways, including the use of frequency and/or time correlation properties of the wireless channel, blind or pilot (reference signal) based channel estimation, and adaptive or non-adaptive channel estimation. Non-parametric methods attempt to estimate the frequency response without relying on a specific channel model. In contrast, the parametric estimation methods assume a certain channel model and determine the parameters of this model. Spaced-time and spaced-frequency correlation functions, discussed earlier, are specific properties of a channel that can be incorporated in the estimation method, improving the quality of estimations. Pilot-based estimation methods are the most commonly used methods in OFDM systems which are applicable in systems where the sender transmits some known signals to the receiver. Blind estimation, on the other hand, relies on some properties of the signal and is rarely used in practical OFDM systems. Adaptive channel estimation methods are typically used for a rapid time-varying channel [69–71].

In this section, we describe a generic pilot-based non-adaptive channel estimation method. In a linear time-invariant AWGN channel, the relationship between the transmitted signal $X_k|_{k=0,1,\ldots,N_{\text{FFT}}-1}$ and received signal Y_k can be expressed as $Y_k = H_k X_k + Z_k$, where Z_k is the frequency-domain noise sampled at the kth sub-carrier frequency, and H_k is the channel transfer-function sampled at the kth sub-carrier frequency. To estimate the channel, pilot symbols are used. Let us assume that every pth sub-carrier contains known pilot symbols $X_{pk}|_{kp=0,p,2p,\ldots,N_p-1}$, where N_p is the total number of pilot sub-carriers across frequency. Using the known pilot symbols X_{pk} and the received symbols Y_{pk} sampled at pilot sub-carriers, the channel transfer function $\hat{H}_k$ sampled at pilot sub-carriers can be written as follows:

$$\hat{H}_{pk} = \frac{Y_{pk}}{X_{pk}} + \frac{Z_{pk}}{X_{pk}} = H_{pk} + W_{pk} \tag{9.31}$$

where W_{pk} is a scaled-noise component at the pkth sub-carrier. Different methods such as one- or two-dimensional linear interpolation can be used to estimate the channel samples at other sub-carrier frequencies.

Channel estimation in OFDM systems is a two-dimensional problem, i.e., the channel transfer function or channel impulse response is a function of time and frequency variables. Due to the computational complexity of two-dimensional estimators, one-dimensional channel estimation methods are more practical. The idea behind one-dimensional estimators is to estimate the channel in one

dimension (e.g., frequency), and then estimate the channel in another dimension (e.g., time), thus obtaining a two-dimensional channel estimate. Linear interpolation is a simple method to estimate the channel from channel samples estimated at pilot sub-carriers. This is done by linearly interpolating the channel samples at the two nearest pilot sub-carriers. Although linear interpolation provides some limited noise reduction of the channel estimates at data locations (due to averaging function), it is the simplicity of the solution that is more attractive. It must be noted that averaging window length (i.e., the number of pilots contained in the averaging window) is inversely proportional to the coherence bandwidth of the channel. The one-dimensional linear interpolation method estimates the channel by interpolating the channel transfer function between $\hat{H}_{m,k}$ and $\hat{H}_{m,l}$ (interpolation across frequency), or between $\hat{H}_{m,k}$ and $\hat{H}_{n,k}$ (interpolation across time). The algorithm shown in Figure 9.47 illustrates the procedure for channel estimation across the time–frequency grid based on the known values of the channel at the reference signal locations. Depending on the relative location of the sub-carrier in the time–frequency grid, the channel is estimated via one-dimensional linear interpolation of channel values at neighboring pilots or merely copying or extrapolation of values at the neighboring pilot, if the previous or the next pilot does not exist [160].

The channel transfer function $H_{m,k}$ at time–frequency index (m, k) can be modeled as a linear weighed sum of two-dimensional basis functions evaluated at the kth sub-carrier (frequency index) and at the mth OFDM symbol (time index):

$$H_{m,k} = \sum_{n=0}^{N-1} \alpha_n \phi_n(m, k) \tag{9.32}$$

where $\phi_n(m, k)$ denotes the nth basis function sampled at the mth OFDM symbol and at the kth sub-carrier, α_n is the coefficient of the nth basis function, and N is number of basis functions used in the linear model. By taking a one-dimensional approach, the above two-dimensional problem is reduced to one-dimensional, if one of the time or frequency indices is kept unchanged:

$$H_k = \sum_{n=0}^{N-1} \alpha_n \phi_n(k) \tag{9.33}$$

The channel samples at pilot sub-carriers shown in Eq. (9.33) can be written as:

$$\hat{H}_{pk} = \sum_{n=0}^{N-1} \alpha_n \phi_n(pk) + W_{pk} \tag{9.34}$$

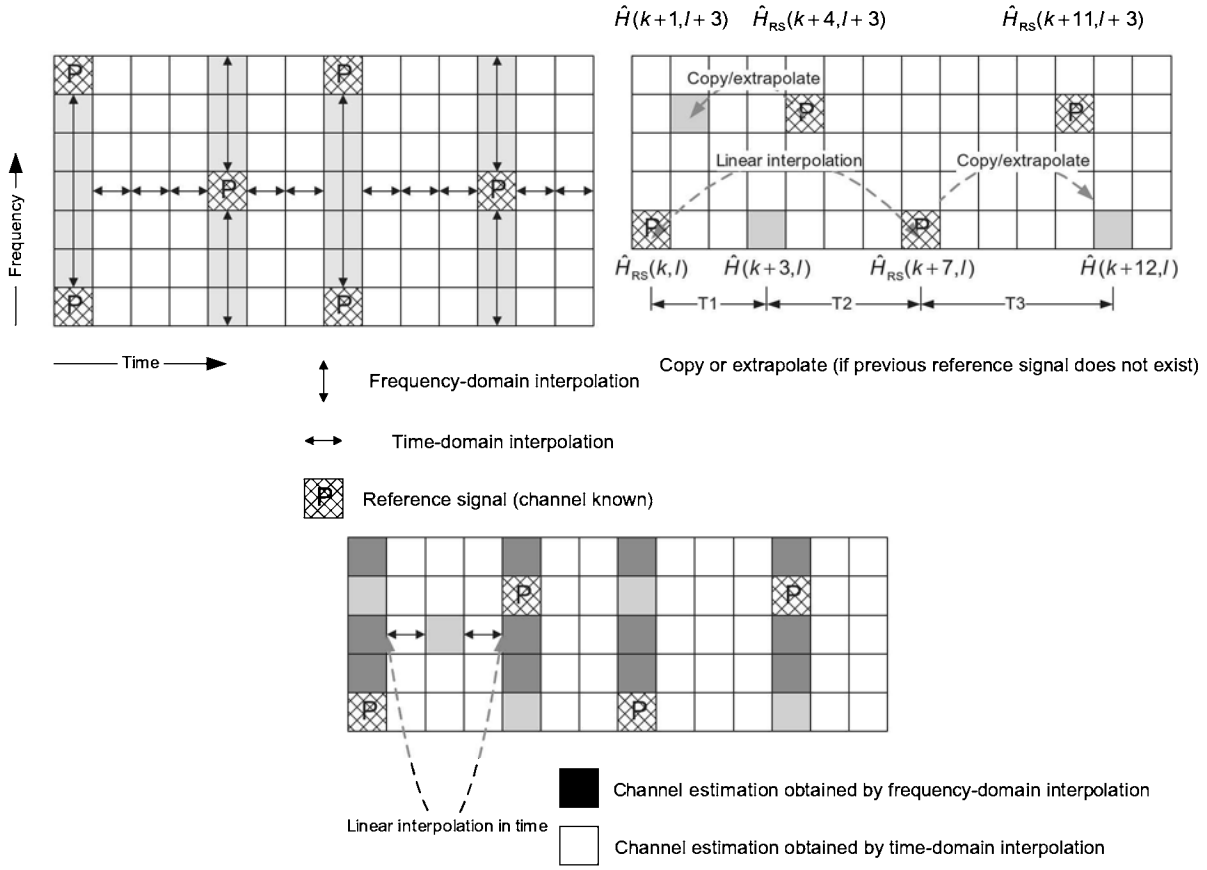


FIGURE 9.47

Two-dimensional channel estimation across frequency and time [176].

The above equation can be represented in matrix form as follows:

$$\begin{bmatrix} \hat{H}_0 \\ \hat{H}_p \\ \hat{H}_{2p} \\ \vdots \\ \hat{H}_{N_p-1} \end{bmatrix} = \begin{bmatrix} \phi_0(0) & \phi_1(0) & \phi_2(0) & \dots & \phi_{N-1}(0) \\ \phi_0(p) & \phi_1(p) & \phi_2(p) & \dots & \phi_{N-1}(p) \\ \phi_0(2p) & \phi_1(2p) & \phi_2(2p) & \dots & \phi_{N-1}(2p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_0(N_p-1) & \phi_1(N_p-1) & \phi_2(N_{p-1}) & \dots & \phi_{N-1}(N_{p-1}) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{N-1} \end{bmatrix} + \begin{bmatrix} Z_0 \\ Z_p \\ Z_{2p} \\ \vdots \\ Z_{N_p-1} \end{bmatrix} \tag{9.35}$$

Alternatively, by using matrix notations, the above equation can be rewritten as follows:

$$\hat{\mathbf{H}} = \mathbf{\Phi}\boldsymbol{\alpha} + \mathbf{Z} \tag{9.36}$$

The least square (LS) estimate of the coefficients is calculated by minimizing the squared distance between the actual channel vector $\mathbf{H}$ and the estimated channel vector $\hat{\mathbf{H}}$ as follows:

$$\hat{\boldsymbol{\alpha}}_{\text{LS}} = (\mathbf{\Phi}^{\mathbf{H}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathbf{H}}\hat{\mathbf{H}} = \mathbf{\Phi}^{-1}\hat{\mathbf{H}} \tag{9.37}$$

where the superscript H denotes conjugate transpose operation, hence:

$$\hat{H}_k = \sum_{n=0}^{N-1} \hat{\alpha}_n \phi_n(k) \tag{9.38}$$

A number of different options exist while selecting the basis functions such as orthogonal polynomials, Fourier series, discrete cosine/sine transform function, etc. There are a number of reasons for using orthogonal polynomials including ease of calculation of $\mathbf{\Phi}^{\mathbf{H}}\mathbf{\Phi}$ since it becomes a diagonal matrix, if matrix $\mathbf{\Phi}$ is orthogonal, resulting in ease of calculation of $\hat{\alpha}_n's$ and the inverse matrix $(\mathbf{\Phi}^{\mathbf{H}}\mathbf{\Phi})^{-1}$ with minimal computational complexity. The degree of orthogonal polynomials can be increased without changes in recursive calculation of $\hat{\alpha}_n's$. When applying the polynomial-based channel estimation, a sliding window is typically used to estimate the channel. This is done to avoid use of high-order polynomials to estimate highly frequency-selective channels. The sliding window approach allows for local approximations to the channel transfer function using a window length larger than the polynomial order, providing a better estimate.

In a special case where the OFDM channel estimation symbols are transmitted periodically and all sub-carriers over an OFDM symbol are used as pilots, the channel estimation algorithm is more simplified. In this case, the receiver uses the estimated channel transfer function to decode the received data within the block until the next pilot symbol arrives. If $Y_k = H_k X_k + Z_k|_{k=0,1,2,\dots,N_{\text{FFT}}-1}$ denotes

the received signal at kth sub-carrier (pilot sub-carrier), the LS estimate of the channel transfer function in this case is given as:

$$\hat{\mathbf{H}}_{\mathbf{LS}} = (Y_1/X_1, Y_2/X_2, \ldots, Y_{N_{\text{FFT}}-1}/X_{N_{\text{FFT}}-1})^H = (Y_1, Y_2, \ldots, Y_{N_{\text{FFT}}-1})^H; X_k = 1, \forall k \tag{9.39}$$

The channel transfer function at non-pilot sub-carriers can be estimated using interpolation in time. For convenience, we assume that the pilot sub-carriers have unity magnitude.

An alternative approach to pilot-based channel estimation is the minimum mean squared error (MMSE) method. The MMSE channel estimator exploits the second-order statistics of the channel transfer function to minimize the mean-squared error. Let $\mathbf{R_{HH}} = E(\mathbf{HH}^{\mathrm{H}})$ and σ_Z^2 denote the autocorrelation matrix of the channel vector $\mathbf{H}$ and the additive white Gaussian noise variance, respectively. The MMSE estimate of the channel transfer function is given as:

$$\hat{\mathbf{H}}_{\text{MMSE}} = \mathbf{R_{HH}}[\mathbf{R_{HH}} + \sigma_Z^2(\mathbf{XX^H})^{-1}]^{-1}\hat{\mathbf{H}}_{\mathbf{LS}} = \mathbf{R_{HH}}[\mathbf{R_{HH}} + \sigma_Z^2\mathbf{I}]^{-1}\hat{\mathbf{H}}_{\mathbf{LS}}; \quad \mathbf{XX^H} = \mathbf{I} \tag{9.40}$$

The MMSE estimator performs much better than LS estimators, especially under the low SNR conditions. A major drawback of the MMSE estimator is its high computational complexity, particularly if matrix inversion is necessary every time the input data changes. Figure 9.48 shows an example where the

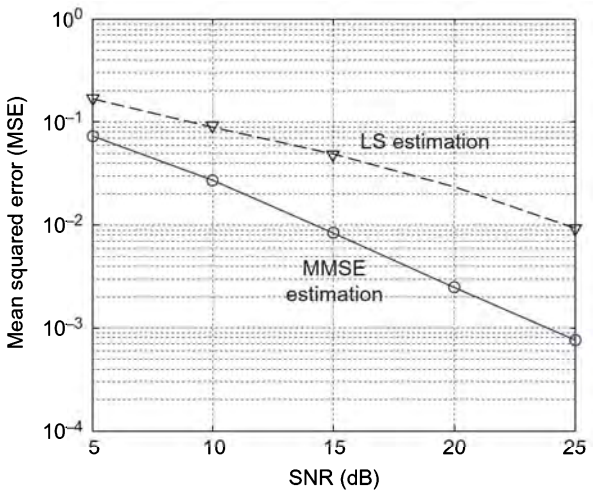


FIGURE 9.48

Example comparison of the LS and MMSE channel estimation algorithms [141].

performance of the LS and MMSE channel estimation algorithms has been compared. In this example $N_{FFT} = 64$ and the channel is modeled as a two-tap delay line filter. The superiority of MMSE over LS estimation in low SNR conditions can be seen in Figure 9.48.

9.10.3 Cell-specific reference signals

The cell-specific reference signals are used during initial cell selection for channel estimation and coherent detection of downlink physical channels. The measurements pertaining to scheduling and handover functions are based on the received signal strength of the CRS transmitted from antenna port zero (and antenna port one). The CRS sequence is a complex pseudo-random sequence, which is a length-31 Gold sequence.[11] The initialization seed for the second M-sequence used for Gold sequence generation is derived from the OFDM symbol number, slot number, and the physical cell identity. The seed is reinitialized at the beginning of each OFDM symbol within a slot. Pseudo-random sequences which are cell-specific are used to differentiate the reference signals transmitted from different cells. There is also a frequency shift that is applied to mapping of the CRS to resource elements to prevent the CRS from adjacent cells overlapping over the time–frequency grid. The generated sequence is transmitted in every slot and is spread throughout the channel bandwidth. The frequency of CRS transmission per slot and actual mapping to resource elements depend on the antenna configuration (Figure 9.49). The CRS symbols between antennas from the same cell are time-division multiplexed; the resource elements used for CRS symbols on one antenna are reserved and are not used on other antenna ports.

The CRSs are transmitted in all downlink subframes in a cell supporting PDSCH transmission. The CRSs are transmitted on one or several of antenna ports 0–3, and are only defined for sub-carrier spacing of 15 kHz. The reference-signal sequence $r_l^{n_s}(m)$ is defined as $r_l^{n_s}(m) = (1/\sqrt{2})(1 - 2c(2m)) + j(1/\sqrt{2})(1 - 2c(2m+1))$, $\forall m = 0, 1, \ldots, 2N_{RB}^{DL} - 1$, where index n_s denotes the slot number within a radio frame, $l = 0, 1, \ldots, 6$ (or 5 in case of an extended cyclic prefix) represents the OFDM symbol number within slot n_s, and $c(i)$ is a pseudo-random sequence, which is initialized with $c_{init} = 2^{10}[7(n_s + 1) + l + 1](2N_{ID}^{cell} + 1) + 2N_{ID}^{cell} + N_{CP}$ at the start

[11] A Gold sequence is a type of binary sequence that is often used in telecommunication and satellite navigation. Gold sequences have bounded small cross-correlations within a set. A set of Gold sequences consists of $2^n - 1$ sequences each with a period of $2^n - 1$. A set of Gold sequences can be generated by taking two maximum-length sequences of the same length $2^n - 1$ such that their absolute cross-correlation is less than or equal to $2^{(n+2)/2}$, where n is the size of the linear feedback shift register used to generate the maximum length sequence [130]. The set of $2^n - 1$ (logical) exclusive OR of the two sequences in their various phases is a set of Gold codes. The highest absolute cross-correlation in this set of codes is $2^{(n+2)/2} + 1$ for even n, and $2^{(n+1)/2} + 1$ for odd n. The exclusive OR of two Gold codes from the same set is another Gold code with arbitrary phase [130].

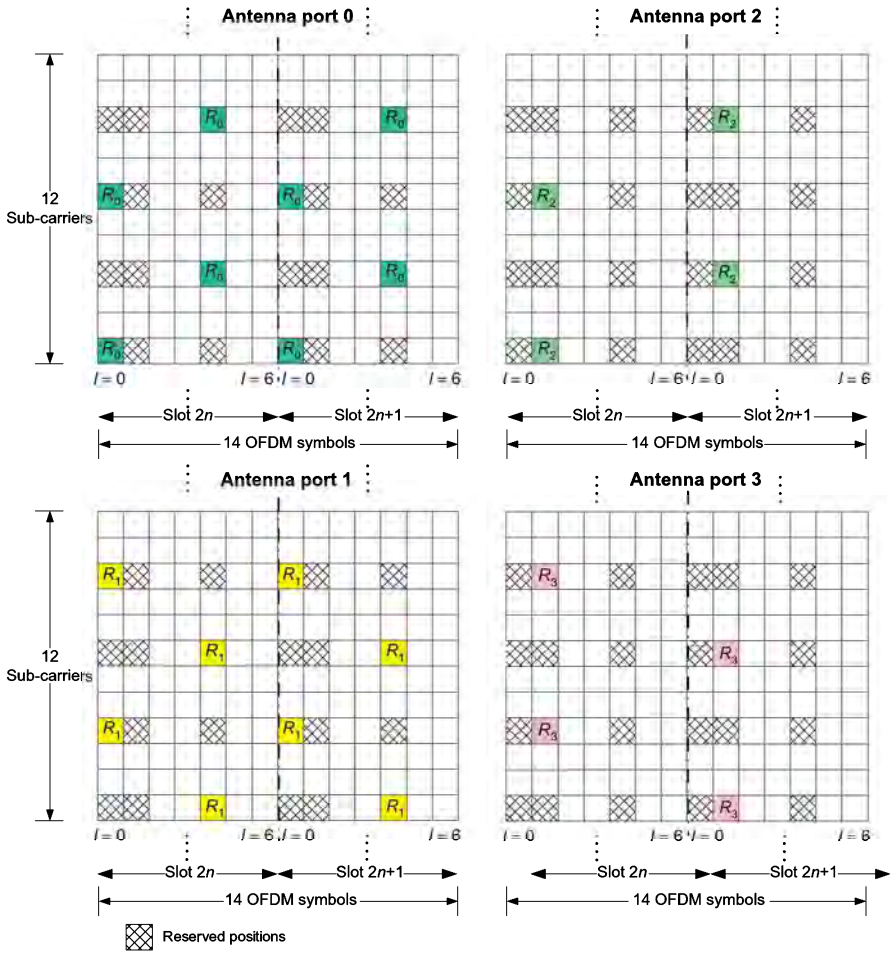


FIGURE 9.49

Mapping of cell-specific reference signals to antenna ports 0, 1, 2, and 3 for four transmit antennas over a resource block using a normal cyclic prefix [3].

of each OFDM symbol.[12] Note that $N_{CP} = 1$ for a normal cyclic prefix and $N_{CP} = 0$ for an extended cyclic prefix. The reference signal sequence $r_l^{n_s}(m)$ is mapped to complex-valued modulation symbols a_{kl}^p that are designated as reference symbols for

[12]Pseudo-random sequences are defined by a length-31 Gold sequence. The output sequence $c(n), \forall n = 0, 1, \ldots, M-1$ of length M is defined by $c(n) = (x_1(n + N_C) + x_2(n + N_C))\text{mod } 2$ in which $x_1(n + 31) = (x_1(n + 3) + x_1(n))\text{mod}2, x_2(n + 31) = (x_2(n + 3) + x_2(n + 2) + x_2(n + 1) + x_2(n))\text{mod } 2$, and $N_C = 1600$. The first m-sequence is initialized with $x_1(0) = 1, x_1(n) = 0, n = 1, 2, \ldots, 30$. The second M-sequence is initialized by $c_{\text{init}} = \sum_{i=0}^{30} x_2(i)2^i$ with the value depending on the application of the sequence [3].

antenna port p in slot n_s as $a_{kl}^{p} = r_{l}^{n_s}(m')$, where the parameters are given as follows [3]:

$$k = 6m + (v + v_{\text{shift}}) \bmod 6; \quad l = \begin{cases} 0, N_{s}^{\text{DL}} - 3 & \text{if } p \in \{0, 1\} \\ 1 & \text{if } p \in \{2, 3\} \end{cases}$$

$$m = 0, 1, \ldots, 2N_{\text{RB}}^{\text{DL}} - 1; \qquad m' = m + \max(N_{\text{RB}}^{\text{DL}}) - N_{\text{RB}}^{\text{DL}} \tag{9.41}$$

The variables v and v_{shift} define the position in the frequency-domain for different reference signals, where v is given by the following expression [3]:

$$v = \begin{cases} 0 & \text{if } p = 0 \text{ and } l = 0 \\ 3 & \text{if } p = 0 \text{ and } l \neq 0 \\ 3 & \text{if } p = 1 \text{ and } l = 0 \\ 0 & \text{if } p = 1 \text{ and } l \neq 0 \\ 3(n_s \bmod 2) & \text{if } p = 2 \\ 3 + 3(n_s \bmod 2) & \text{if } p = 3 \end{cases} \tag{9.42}$$

A cell-specific frequency shift is given by $v_{\text{shift}} = N_{\text{ID}}^{\text{cell}} \bmod 6$ (Figure 9.50). As shown in Figure 9.50, the resource element allocation for CRSs cycles once every six physical cell identity values, i.e., cell identity values that are integer multiples of 6 has the same resource element allocation as cell identity 0. This corresponds to one cycle for every two physical layer cell-identity groups. The resource elements

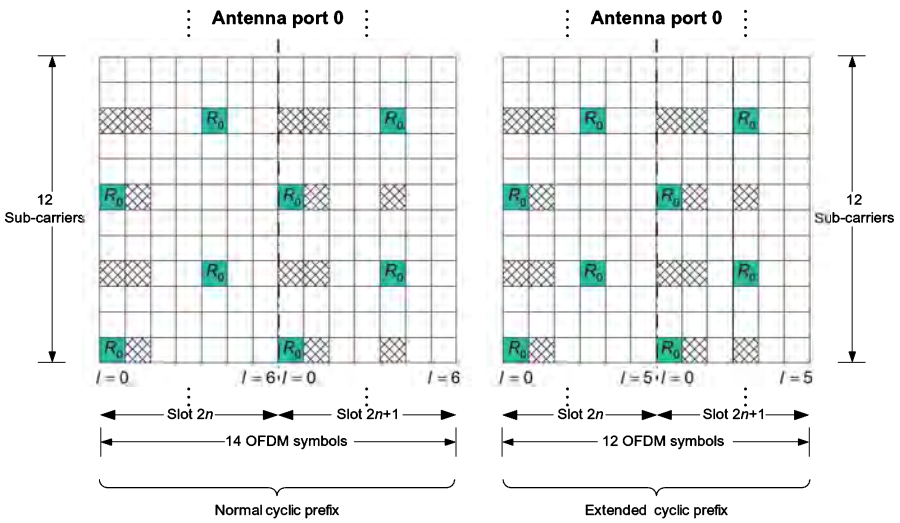


FIGURE 9.50

Cell-specific reference signal mapping for normal and extended cyclic prefix subframes [3].

(k, l) used for transmission of CRSs on an any of the antenna ports in a slot are not used for any transmission on another antenna port in the same slot and are set to zero. In an MBSFN subframe, CRSs are only transmitted in the non-MBSFN region of the MBSFN subframe. Figure 9.49 shows the CRS mapping to the resource grid. The procedure for generation of the CRSs is illustrated in Figure 9.51.

Given a different number of available OFDM symbols in a subframe with an extended cyclic prefix, the mapping of the CRSs is different than that for subframes with a normal cyclic prefix. Figure 9.50 illustrates and compares the CRS mapping in subframes with a normal and extended cyclic prefix associated with antenna port 0 and 4 transmit antennas. While the number of resource elements mapped with reference signals is the same, the CRS overhead in subframes with an extended cyclic prefix is higher due to a lower number of total sub-carriers in a resource block.

As shown in Figure 9.51, the CRSs are based on length-31 Gold pseudo-random sequences. The length-31 Gold pseudo-random sequence is generated with the seed c_{init}, based on the slot number, symbol number, cell identity, and cyclic prefix type. While the sequence itself is 230 bits long, the number of bits from the sequence selected for transmission is based on the largest channel bandwidth, which is currently 20 MHz. Therefore, $4 \max(N_{\text{RB}}^{\text{DL}})$ bits are selected from

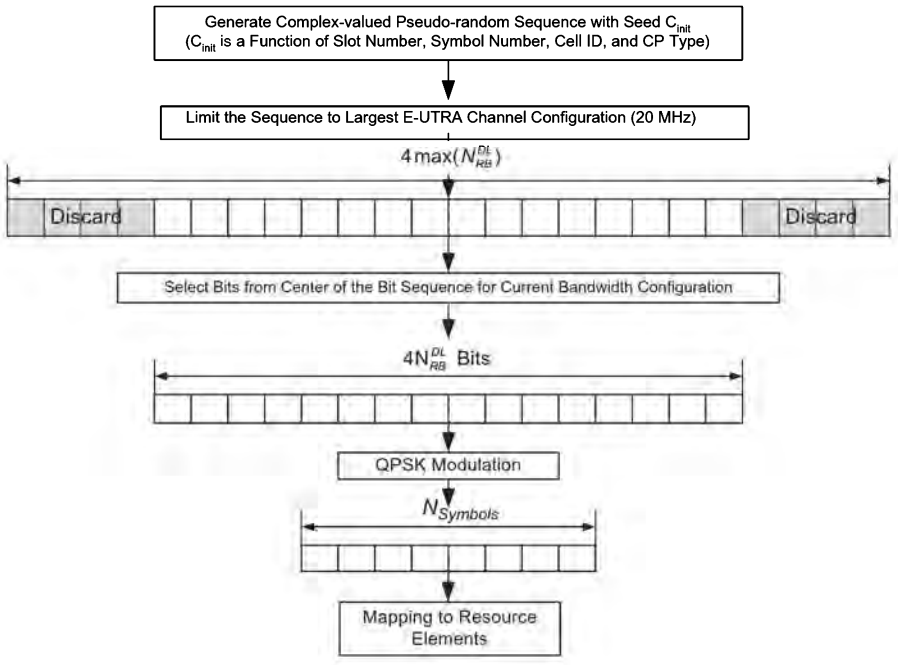


FIGURE 9.51

Generation of cell-specific reference signals.

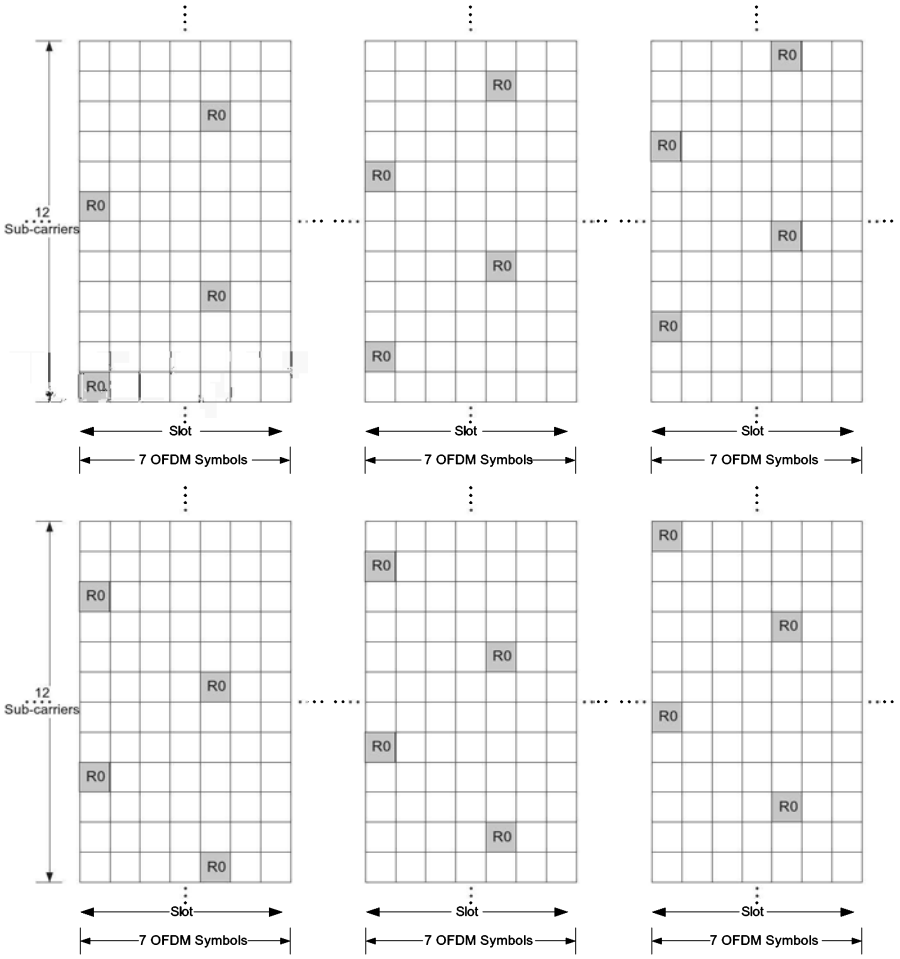


FIGURE 9.52

Dependency of cell-specific reference signals location on physical cell identifier [13].

the sequence where $\max(N_{\mathrm{RB}}^{\mathrm{DL}})$ is the maximum number of resource blocks in 20 MHz. From 4 $\max(N_{\mathrm{RB}}^{\mathrm{DL}})$, $4N_{\mathrm{RB}}^{\mathrm{DL}}$ bits are selected, where $N_{\mathrm{RB}}^{\mathrm{DL}}$ is the number of resource blocks corresponding to the system bandwidth. The bits are selected from the center of the truncated sequence for the maximum channel configuration. This ensures that the same reference symbols are present around PBCH regardless of channel bandwidth. The PBCH uses the six resource blocks (72 subcarriers) at the center of the channel. The UE can use these symbols for channel estimation for PBCH detection without knowing the overall channel bandwidth. The channel bandwidth in terms of resource blocks is determined from the

PBCH. After PBCH is decoded, the UE determines the length of the reference signal used for the current channel configuration.

9.10.3.1 Received signal strength indicator measurement

Received signal strength indicator (RSSI) is the total power that UE observes across the entire system bandwidth. This includes the main signal and co-channel non-serving cell signal, adjacent channel interference, and the thermal noise within the specified frequency band. This is the power of the received signal prior to demodulation, thus the UE can measure this power without any synchronization and demodulation [6].

9.10.3.2 Reference signal received power measurement

Reference signal received power (RSRP) is defined as the linear average over the power contributions of the resource elements that carry cell-specific reference signals within the considered measurement bandwidth. For the purpose of RSRP calculation, the R_0 cell-specific reference signals are used; however, if the UE can reliably detect R_1 cell-specific reference signals, then it may use R_1 in addition to R_0 in order to calculate RSRP. The reference point for the RSRP measurement is the antenna connector of the UE. If the UE uses a receive diversity scheme, the reported RSRP value must not be lower than individual RSRP measured at any of the diversity branches. There are requirements for both absolute and relative RSRP. The absolute requirements range from ± 6 to ± 11 dB, depending on the noise level and environmental conditions. Measuring the difference in RSRP between two cells on the same frequency (intra-frequency measurement) is a more accurate operation for which the requirements vary from ± 2 to ± 3 dB. The requirements widen again to ± 6 dB when the cells are on different frequencies (inter-frequency measurement) [6].

9.10.3.3 Reference signal received quality measurement

Reference signal received quality (RSRQ) is defined as the ratio $\text{RSRQ} = N_{\text{RB}}\,\text{RSRP}/\text{RSSI}_{E\text{-UTRA}}$ (dB), where N_{RB} is the number of resource blocks over which the measurement is conducted. Note that the value of RSRQ in decibels is always negative, since the denominator is larger than the numerator. The measurements in the numerator and denominator are conducted over the same set of resource blocks. The E-UTRA carrier RSSI comprises the linear average of the total received power observed only in OFDM symbols containing the cell-specific reference symbols for antenna port zero over N_{RB} resource blocks in the measurement bandwidth by the UE from all sources, including co-channel serving and non-serving cells, adjacent channel interference, and thermal noise. If higher-layer signaling indicates certain subframes for performing RSRQ measurements, then RSSI is measured over all OFDM symbols in the designated subframes. The reference point for the RSRQ is the antenna connector of the UE. If the UE uses a receive diversity scheme, the reported RSRP value must not be lower than individual RSRQ measured at any of the diversity branches.

Table 9.10 RSRQ Mapping [6]

Reported RSRQ Value	Measured Quantity (dB)
RSRQ_00	RSRQ < −19.5
RSRQ_01	−19.5 ≤ RSRQ < −19
RSRQ_02	−19 ≤ RSRQ < −18.5
...	...
RSRQ_32	−4 ≤ RSRQ < −3.5
RSRQ_33	−3.5 ≤ RSRQ < −3
RSRQ_34	−3 ≤ RSRQ

Measuring RSRQ becomes particularly important near the cell-edge when decisions need to be made regardless of absolute RSRP to perform a handover to the target cell. Reference signal receive quality is used only during connected states. Intra- and inter-frequency absolute RSRQ accuracy varies from ±2.5 to ±4 dB, which is similar to the inter-frequency relative RSRQ accuracy of ±3 to ±4 dB.

It must be noted that RSRQ measurement provides additional information when RSRP is not sufficient to make a reliable handover or cell reselection decision. The RSRQ is the ratio of the RSRP and the RSSI, which depends on the measurement bandwidth, and thereby the number of resource blocks. The RSSI value includes the total received wideband power from all interference and thermal noise sources. Since RSRQ combines signal strength as well as interference level, its value further assists mobility management in E-UTRA. The UE typically measures RSRP or RSRQ per instructions from the network (via RRC messages) and reports the values back to the eNB. When it reports the RSRQ value, it does not use the real RSRQ value; rather it sends a non-negative value ranging from 0 to 34, and each of these values are mapped to a specific range of real RSRQ values as given in Table 9.10. The reporting range of RSRQ is defined from −19.5 dB to −3 dB with 0.5 dB resolution. As an example, assume that only reference signals are transmitted in a resource block and that data, noise, and interference are not considered. In this case, the RSRQ is equal to −3 dB. If cell-specific reference signals and the sub-carriers carrying data are equally powered, the RSRQ is equal to 1/12 or −10.79 dB.

9.10.4 UE-Specific reference signals associated with PDSCH

It was mentioned that cell-specific reference signals are transmitted on each antenna port. An antenna port is generally used as a generic term for signal transmission under identical channel conditions. The symbols that are transmitted via identical antenna ports are subject to the same channel conditions. In order to determine the characteristics of the channel from an antenna port, a UE must carry out a separate channel estimation for each antenna port. Therefore, separate

reference signals that are suitable for estimating the respective channel are defined for each antenna port. There is one resource grid per antenna port. The cell-specific reference signals are used for both coherent detection and mobility-related measurements. Their structure in time and frequency ensures sufficient channel estimation accuracy. An LTE-Advanced eNB is required to support LTE Rel-8 user terminals. Since the structure of PCFICH, PHICH, PDCCH, PBCH, and PDSCH (transmit diversity only) is unchanged in Rel-10, the support of CRS is also required for LTE-Advanced UEs in order to detect these essential physical channels. The UE-specific reference signals are supported for transmission of PDSCH, and are transmitted on antenna ports $p = 5$, $p = 7$, $p = 8$, or $p = 7, 8, \ldots, N_l + 6$, where N_l is the number of spatial layers used for transmission of PDSCH. The UE-specific reference signals or DM-RS can be precoded and support non-codebook-based precoding. The DM-RS pattern for a higher number of layers has been extended from dual-layer format of transmission mode 8 in Rel-9 to eight layers in Rel-10. The CSI-RS is transmitted on each physical or virtual antenna port, and is only used for channel state measurement/reporting purposes. Its channel estimation accuracy can be relatively poorer compared to that of DM-RS. In Rel-8, the CRS is used for both demodulation and measurement purposes. When using precoding in conjunction with CRSs, the precoding information must be sent to the UE, and the CRS power is boosted to support various types of geometry. The Rel-8 E-UTRA downlink MIMO transmission modes are based on CRS. Figure 9.53 shows the CRS-based procedure for precoding where the cell-specific reference signals are inserted after the data is precoded. Therefore, the cell-specific reference signals do not contain information about the

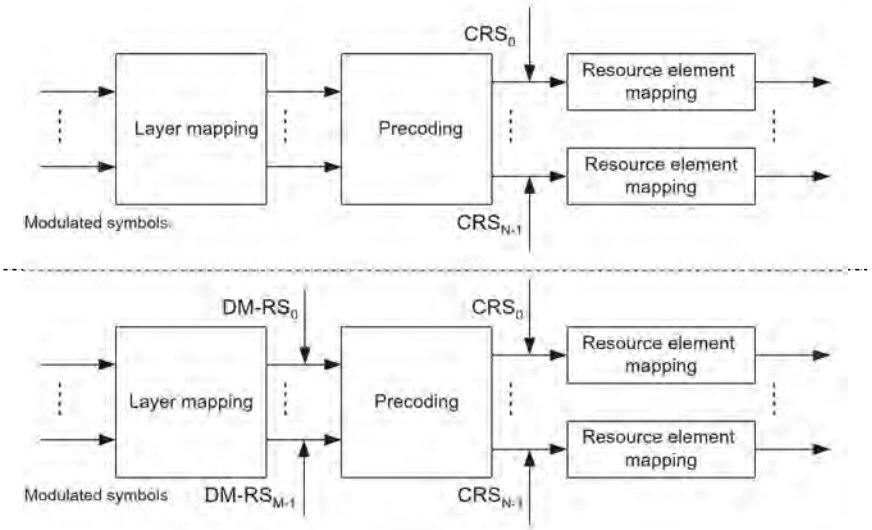


FIGURE 9.53

Comparison of the CRS and DM-RS usage in Rel-8 and Rel-10 of E-UTRA [110].

precoding scheme, and codebook information must be separately sent to the UE. In contrast, in LTE-Advanced, downlink MIMO transmission is based on DM-RS and CSI-RS. Figure 9.53 shows the DM-RS-based procedure for precoding where UE-specific reference signals are precoded in the same way as the data. Therefore, no separate precoding information is sent to the UE. Note that in LTE Rel-10, DM-RS is used for demodulation and CSI-RS is used for channel state measurement purposes. Unlike CRS-based precoding, the precoding information is not needed for demodulation.

It was stated earlier that UE-specific reference signals are used for transmission of the downlink shared channel and are transmitted on antenna port(s) $p = 5$, $p = 7$, $p = 8$ or $p = 7, 8, \ldots, N_l + 6$, where N_l is the number of layers (up to eight). It must be noted that the UE-specific reference signals on any antenna port are a valid reference for data demodulation, if the data transmission is associated with the corresponding antenna port. The UE-specific reference signals are exclusively transmitted on the resource blocks to which the corresponding user data is mapped. The UE-specific reference signal is not mapped to resource elements (k, l) that are reserved for transmission of control signaling, physical channels, or physical signals.

For antenna port 5, the UE-specific reference-signal sequence $r_{n_s}(m)$ is defined by $r_{n_s}(m) = (1/\sqrt{2})[1 - 2c(2m)] + \mathrm{j}(1/\sqrt{2})[1 - 2c(2m+1)]$, $\forall m = 0, 1, \ldots, 12N_{\mathrm{RB}} - 1$, where N_{RB} denotes the number of resource blocks allocated to the corresponding data transmission. The pseudo-random sequence $c(i)$ was defined earlier. The pseudo-random sequence generator is initialized with $c_{\mathrm{init}} = \left(\lfloor n_s/2 \rfloor + 1\right)\left(2N_{\mathrm{ID}}^{\mathrm{cell}} + 1\right)2^{16} + n_{\mathrm{RNTI}}$ at the start of each subframe where n_{RNTI} is the 16-bit radio network temporary identifier (RNTI) of the UE. For antenna ports $p \in \{7, 8, \ldots, N_l + 6\}$, the UE-specific reference-signal sequence $r_{n_s}(m)$ is defined by $r_{n_s}(m) = (1/\sqrt{2})[1 - 2c(2m)] + \mathrm{j}(1/\sqrt{2})[1 - 2c(2m+1)]$, where $m = 0, 1, \ldots, 12N_{\mathrm{RB}}^{\mathrm{DL}} - 1$ (for normal cyclic prefix) or $m = 0, 1, \ldots, 16N_{\mathrm{RB}}^{\mathrm{DL}} - 1$ (for extended cyclic prefix). The pseudo-random sequence $c(i)$ is defined similar to the previous section. Prior to LTE Rel-11, the pseudo-random sequence generator was initialized with $c_{\mathrm{init}} = \left(\lfloor n_s/2 \rfloor + 1\right)\left(2N_{\mathrm{ID}}^{\mathrm{cell}} + 1\right)2^{16} + n_{\mathrm{SCID}}$ at the start of each subframe. The value of the scrambling identity n_{SCID} is zero unless specified otherwise. For a data transmission on ports 7 or 8, n_{SCID} is indicated by the DCI format 2B or 2C. In the case of DCI format 2B, $n_{\mathrm{SCID}} = 0$ if the scrambling identity field is zero. It must be noted that CQI estimation at the terminal is always based on cell-specific reference signals. In LTE Rel-11, and as a result of the introduction of the new PDCCH structure based on the UE-specific reference signals, the initialization of the DM-RS sequences associated with antenna ports $p \in \{7, 8, \ldots, N_l + 6\}$ was modified as follows: $c_{\mathrm{init}} = \left(\lfloor n_s/2 \rfloor + 1\right)\left(2N_{\mathrm{ID}}^{(n_{\mathrm{SCID}})} + 1\right)2^{16} + n_{\mathrm{SCID}}$, where $N_{\mathrm{ID}}^{(i)}|_{i=0,1}$ is defined as $N_{\mathrm{ID}}^{(i)} = N_{\mathrm{ID}}^{\mathrm{cell}}$, if no value for $N_{\mathrm{ID}}^{\mathrm{DM\text{-}RS_i}}$ is conveyed by higher layers, or if DCI format 1A is used for the DCI associated with PDSCH transmission; otherwise $N_{\mathrm{ID}}^{(i)} = N_{\mathrm{ID}}^{\mathrm{DM\text{-}RS_i}}$ [3].

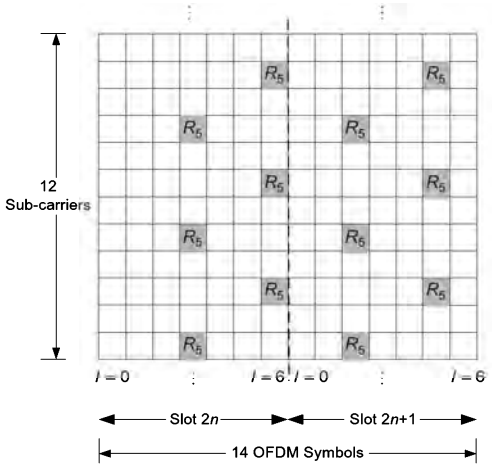


FIGURE 9.54

Mapping of UE-specific reference signals to antenna port 5 [3].

In a PRB with index n_{PRB} allocated for data transmission on antenna port 5, and assuming the normal cyclic prefix, the DM-RS $r_{n_s}(m)$ is mapped to the complex-valued modulation symbols a_{kl}^{p} over slot n_s as $a_{kl}^{5} = r_{n_s}(3l'N_{\text{RB}} + m')$, where $m' = 0, 1, \ldots, 3N_{\text{RB}} - 1$ denotes the number of UE-specific reference signal resource elements within a respective OFDM symbol allocated to user data transmission, $l' = \{0, 1\}$ for even-numbered slots, and $l' = \{2, 3\}$ for odd-numbered slots.

The UE-specific reference signal pattern for up to two layers, or alternatively rank-2 transmission, is identical to that of LTE Rel-9 to maintain backward compatibility. For up to four layers, the DM-RS pattern is obtained by extending LTE Rel-9 UE-specific reference signal pattern using a combination code-division and frequency-division multiplexing, as shown in Figure 9.55. The four precoded layers (antenna ports 7–10) are grouped into two groups of two resource elements that are spread with the same length-2 Walsh–Hadamard[13] orthogonal cover codes (OCCs) as in Rel-9. The UE-specific reference signals in different groups are frequency-multiplexed on adjacent sub-carriers. This pattern has been shown to provide reasonable performance for a variety of channel conditions and UE speeds [121].

[13]In coding theory, the Walsh–Hadamard code is a linear code over a binary alphabet that maps messages of length 2^n. The Walsh–Hadamard code is unique in that each non-zero codeword has a Hamming weight of exactly 2^{n-1}, which implies that the distance of the code is also 2^{n-1} resulting in a $[2^n, n, 2^{n-1}]_2$ code. The Walsh–Hadamard code is locally decodable, which enables recovery of parts of the original message with high probability, while only looking at a small fraction of the received codeword [27,40].

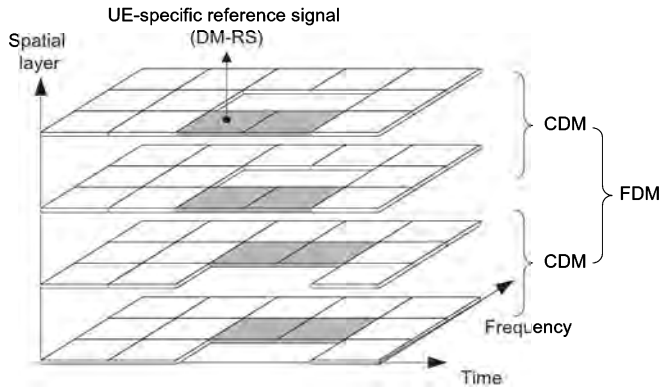


FIGURE 9.55

Code and frequency division multiplexing of the UE-specific reference signals [177].

For rank-8 transmission on antenna ports 7−14, the UE-specific reference signal pattern of Figure 9.55 is further extended using a hybrid CDM/FDM approach with two CDM groups each using a Walsh−Hadamard OCC of length 4, as shown in Figure 9.56. The UE-specific reference signals for each layer are spread across four resource elements, all having the same frequency location but different time locations within the subframe. This approach maintains the balanced power property of the rank-4 UE-specific pilot design and is optimized for pedestrian mobility class, which is the most likely use case for eight-layer transmission. The OCCs of length two are embedded in the OCC of length four to ensure backward compatibility with LTE Rel-9. The mapping of the OCCs to resource elements is shown in Figure 9.56. The mapping for odd and even resource blocks (in the frequency-domain) is different in order to provide orthogonality in both time and frequency dimensions. Such orthogonality improves the performance of channel estimation in high Doppler scenarios, and reduces the inter-carrier interference across the two CDM groups in the presence of frequency offsets. The OCC mapping also enables power balancing across OFDM symbols by peak power randomization [14]. The reference sequence for up to eight layers is the same as in Rel-9, except for antenna ports $AP_p|_{p=9,10,\ldots,14}$ where only one sequence initialization is possible.

In mathematical terms, for antenna ports $AP_p|_{p=7,8,\ldots,N_l+6}$, in a PRB with frequency-domain index n_{PRB} assigned for the corresponding user data transmission, part of the reference sequence $r(m)$ is mapped to complex-valued modulation symbols a_{kl}^p in a subframe (assuming normal cyclic prefix) as follows: $a_{kl}^p = w_p(l')\, r[3l'\max(N_{\text{RB}}^{\text{DL}}) + 3n_{\text{PRB}} + m']$, where $w_p(i) = \overline{w}_p(i)$ if $(m' + n_{\text{PRB}}) \bmod 2 = 0$ or $w_p(i) = \overline{w}_p(3-i)$ if $(m' + n_{\text{PRB}}) \bmod 2 = 1$, where $\overline{w}_p(i)$ is given in Table 9.11. The UE-specific reference signal patterns for different transmission ranks (number of spatial layers) are shown in Figure 9.57.

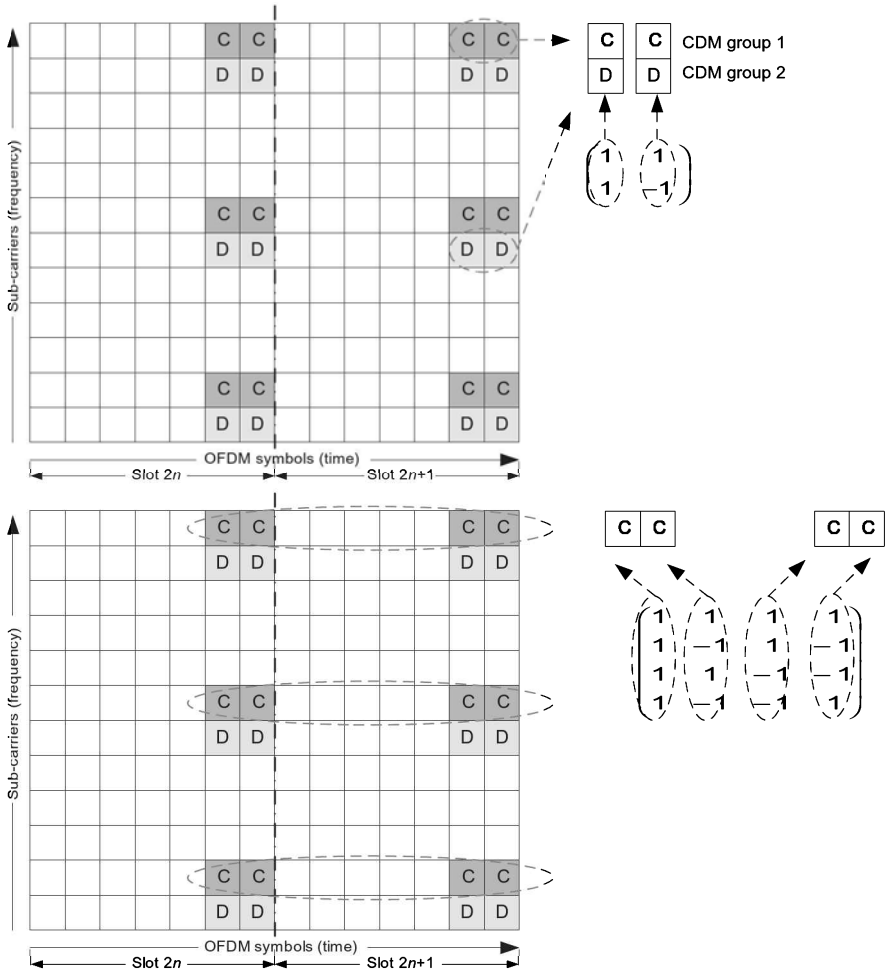


FIGURE 9.56

Length-2 and length-4 OCC mapping and DM-RS structure [121].

9.10.5 CSI reference signals

The new multi-user MIMO techniques that were proposed as part of Rel-10 and beyond require some level of CSI knowledge at the base station so that the system can dynamically adapt to the radio channel conditions experienced by the user terminals in order to optimize the performance. In TDD systems, this information is intuitively obtained from the uplink direction based on the reciprocity of the wireless channel, provided that the channel fading is sufficiently slow and transceivers at the base and mobile stations are calibrated, due to the fact that the

Table 9.11 Orthogonal Cover Code of Length 4 [3]

Antenna Port p	$[\bar{w}_p(0)$	$\bar{w}_p(1)$	$\bar{w}_p(2)$	$\bar{w}_p(3)]$
7	[+1	+1	+1	+1]
8	[+1	−1	+1	−1]
9	[+1	+1	+1	+1]
10	[+1	−1	+1	−1]
11	[+1	+1	−1	−1]
12	[−1	−1	+1	+1]
13	[+1	−1	−1	+1]
14	[−1	+1	+1	−1]

same carrier frequency is used for transmission and reception. On the other hand, due to the asymmetry of FDD systems, feedback information over the uplink is required. Full CSI could cause an excessive overhead in some scenarios, thus quantized or statistical CSI are used in practice. In addition, terminal mobility can cause further degradation of the system performance, if the CSI arriving at the eNB is already outdated. Multi-antenna techniques in a multi-user scenario have the role of delivering streams of data in a spatially multiplexed manner to different users in such a way that all the degrees of freedom of a MIMO system are utilized. The idea is to perform intelligent space-division multiple access or multi-user beamforming so that the transmission of the base station is adapted to each user's instantaneous channel conditions in order to achieve the maximum gain in the direction of that user. The base station collects information on the CSI of each UE and subsequently decides on the optimal resource allocation strategy for each user.

Channel state information reference signals were introduced as part of Rel-10 to enable channel state measurement and reporting by UEs for up to eight downlink transmit-antennas to facilitate eNB precoding functions in transmission mode 9. LTE Rel-10 supports transmission of CSI-RS for 1, 2, 4, and 8 transmit antennas. The CSI-RSs also enable the UE to estimate the CSI for multiple neighbor cells in addition to its serving cell, to support multi-cell cooperative transmission schemes. The following principles have been incorporated in the design of the CSI-RS [14]:

- It is desirable to place CSI-RS at locations with uniform spacing in the frequency-domain.
- It is desirable to minimize the number of subframes containing CSI-RS in the time-domain so that a UE can estimate the CSI for different antenna ports and different cells with a minimal wake-up (active state) duty cycle when the UE is in DRX mode to reduce UE power consumption.
- The overall CSI-RS overhead involves a trade-off between accurate CSI estimation for efficient operation and minimizing the impact on legacy UEs

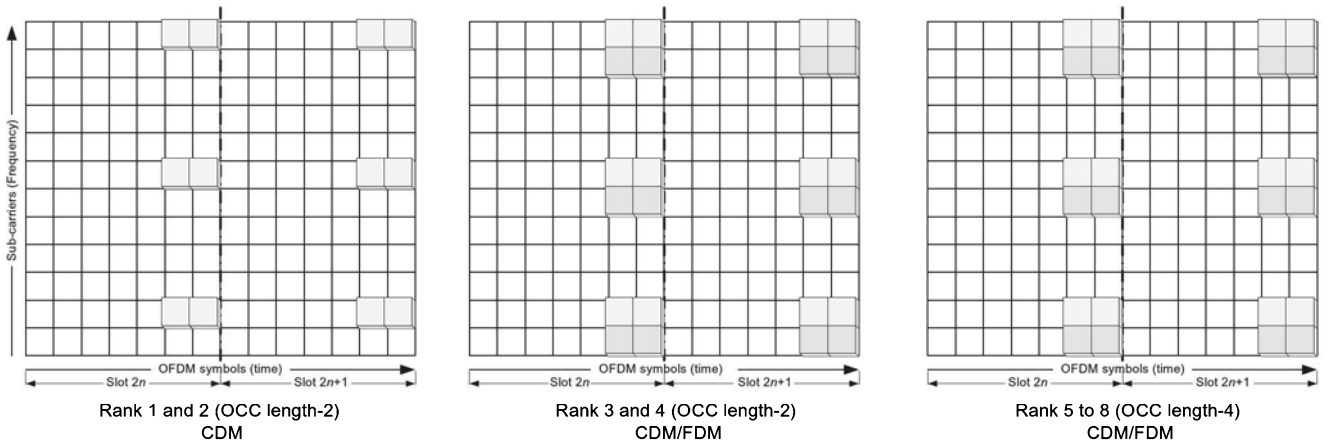


FIGURE 9.57

Mapping of UE-specific reference signals to resource blocks for different numbers of layers [177].

which are unaware of the presence of CSI-RS. Furthermore, the data sub-carriers of the legacy UE are punctured by the CSI-RS transmissions. Figure 9.58 shows that a CSI-RS density of one resource element per resource block per antenna port is a reasonable choice, as the throughput degradation compared to ideal CSI estimation is negligible.

- The CSI-RSs of different antenna ports within a cell, and ideally from other cells, should be orthogonally multiplexed to enable accurate CSI estimation.
- To ensure backward compatibility, CSI-RSs should not be mapped to the resource elements used for CRSs and control channels, as well as avoiding resource elements used for the Rel-10 UE-specific reference signals.

A set of CSI-RSs were designed in Rel-10, taking the above design considerations into account. An example CSI-RS pattern for 1, 2, 4, and 8 antenna ports is shown in Figure 9.58. The OCC codes of length 2 are used, so that CSI-RSs on two antenna ports share two resource elements on a given sub-carrier. The pattern shown in Figure 9.58 is applicable to frame structure type 1 and frame structure type 2 [121].

The CSI-RS patterns follow a nested structure, meaning that the resource elements used in the case of two CSI-RS antenna ports are a subset of those used for four and eight antenna ports; this helps to simplify the implementation. The total number of supported CSI-RS patterns is 40, which can be used to give a frequency-reuse factor of 5 between cells with eight antenna ports per cell, or a factor of 20 in the case of two antenna ports. It can be seen that collisions may occur with resource elements used for the DM-RS antenna ports defined in Rel-8 for PDSCH transmission mode 7. Therefore, it may be desirable to avoid scheduling UEs in transmission mode 7 in subframes containing CSI-RS. This pattern is designed to avoid collisions with the Rel-8 UE-specific reference signal antenna port, as this port is more suited to TDD operation where channel reciprocity can be more effectively exploited to support beamforming. However, this pattern only offers frequency-reuse factors between 3 and 12 depending on the number of antenna ports per cell [14].

The CSI-RS configuration is UE-specific. When the CSI-RS is configured, it is only present in certain subframes conforming to a given duty cycle and subframe offset. The duty cycle and offset of the subframes containing CSI-RSs and the CSI-RS pattern used in those subframes are provided to a Rel-10 UE through RRC signaling. The duty cycle and subframe offset are jointly coded, while the CSI-RS pattern is configured independently of these parameters. It should be noted that in subframes containing CSI-RSs, the rate-matching for PDSCH transmissions to Rel-10 UEs is considerate of the resource elements used for CSI-RS and the encoded PDSCH data is only mapped to the neighboring sub-carriers, whereas PDSCH transmissions to the legacy UEs are punctured by CSI-RS transmissions. Therefore, the density and periodicity of the CSI-RSs should be chosen

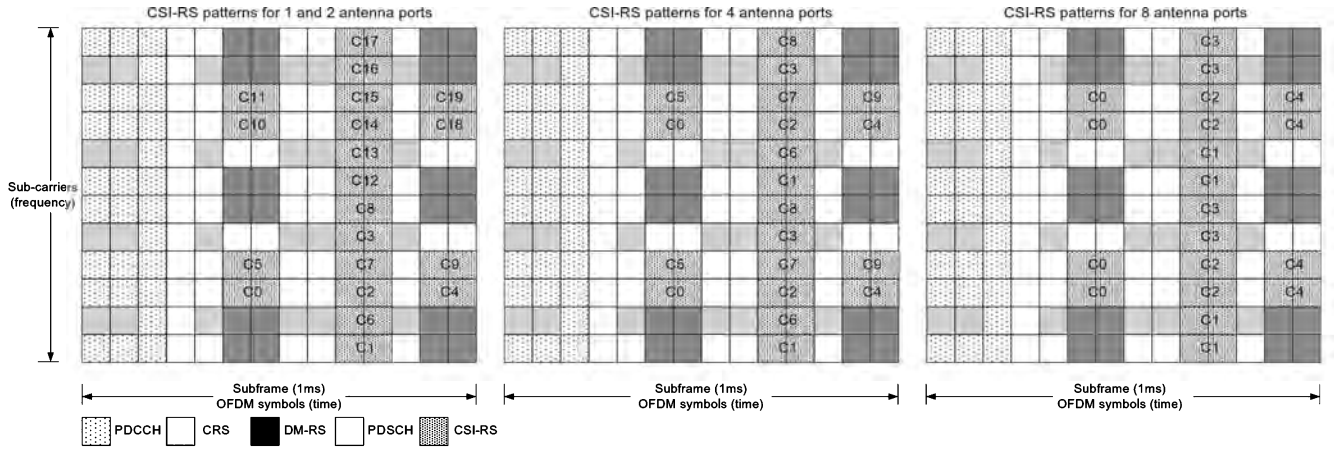


FIGURE 9.58

CSI-RS patterns for 1, 2, 4, and 8 transmit antennas [3,121].

such that the impact of puncturing[14] on the performance of these users is negligible.

In the context of cooperative MIMO, it may be possible to improve the performance of channel estimation, and especially interference estimation, by coordinating CSI-RS transmissions across multiple cells. In Rel-10 it is possible to mute a specific set of sub-carriers in data transmissions from a cell. The locations of these resource elements can be chosen to avoid colliding with CSI-RS transmissions from other cells, thus improving the inter-cell interference measurement quality. The muting pattern is signaled to the UEs using a 16-bit bitmap, where each bit corresponds to a 4-port CSI-RS configuration for frame structure type 1 or type 2. As a result, all the resource elements in the selected 4-port CSI-RS configuration are muted and the UE can assume zero transmission power on those resource elements, unless they are used to transmit CSI-RSs. In subframes in which CSI-RS muting is configured, data transmissions for Rel-10 UEs are rate-matched around the muted resource elements in the same way as for CSI-RSs. Data transmitted to Rel-8/9 UEs is punctured at the muted resource elements [15].

The CSI-RSs are transmitted on one, two, four, or eight antenna ports $AP_p = \{p = 15\}, \{p = 15, 16\}, \{p = 15, \ldots, 18\}, \{p = 15, \ldots, 22\}$ and are only defined for sub-carrier spacing of 15 kHz. The reference signal sequence is defined as $r_l^{n_s}(m) = (1/\sqrt{2})[1 - 2c(2m)] + \mathrm{j}(1/\sqrt{2})[1 - 2c(2m+1)]$, $\forall m = 0, 1, \ldots, \max(N_{\mathrm{RB}}^{\mathrm{DL}}) - 1$, where n_s is the slot number within a radio frame and $l = 0, 1, \ldots, 6$ (or 5 for an extended cyclic prefix) is the OFDM symbol number within the slot. The pseudo-random sequence $c(i)$ was defined in the previous sections. The pseudo-random sequence generator is initialized with $c_{\mathrm{init}} = 2^{10}(7(n_s + 1) + l + 1)(2N_{\mathrm{ID}}^{\mathrm{CSI}} + 1) + 2N_{\mathrm{ID}}^{\mathrm{CSI}} + N_{\mathrm{CP}}$ at the start of each OFDM symbol where $N_{\mathrm{CP}} = 1$ for a normal cyclic prefix and $N_{\mathrm{CP}} = 0$, otherwise. In subframes configured for CSI-RS transmission, the reference signal sequence $r_l^{n_s}(m)$ is mapped to complex-valued modulation symbols $a_{kl}^p = w_\alpha r_l^{n_s}(m')$ which have been designated as reference symbols on AP_p where $m' = m + \lfloor [\max(N_{\mathrm{RB}}^{\mathrm{DL}}) - N_{\mathrm{RB}}^{\mathrm{DL}}]/2 \rfloor; m = 0, 1, \ldots, N_{\mathrm{RB}}^{\mathrm{DL}} - 1$ and $w_\alpha = 1$ for $p \in \{15, 17, 19, 21\}$ and $w_\alpha = (-1)^\alpha$ for $p \in \{16, 18, 20, 22\}$. The virtual CSI identifier in the above equation is equal to the cell identity unless configured by higher layers.

[14] In coding theory, puncturing is the process of removing some of the parity bits after encoding with an error-correction code. This has the same effect as encoding with an error-correction code with a higher rate, or less redundancy. However, with puncturing the same decoder can be used regardless of how many bits have been punctured, thus puncturing considerably increases the flexibility of the system without significantly increasing its complexity. In some cases, a predefined pattern of puncturing is used in an encoder. Then, the inverse operation, known as de-puncturing, is implemented by the decoder. Puncturing is used in UMTS and LTE during the rate-matching process. Puncturing is often used with the soft decoding and Viterbi algorithm [27,40].

The following parameters for CSI-RS are configured via higher layer signaling [5]:

- Number of CSI-RS ports
- CSI-RS configuration
- CSI-RS subframe configuration $I_{\text{CSI-RS}}$
- Subframe configuration period $T_{\text{CSI-RS}}$
- Subframe offset $T_{\text{Offset-CSI-RS}}$
- UE assumption on reference PDSCH transmitted power for CSI feedback P_c. The parameter P_c is the assumed ratio of PDSCH energy per resource element (EPRE)[15] to CSI-RS EPRE when UE derives CSI feedback and takes values in the range of [−8 dB to 15 dB] in 1 dB step sizes.

Multiple CSI-RS configurations can be used in a given cell, which include zero or one configuration for which the UE can assume non-zero transmission power, and zero or more configurations for which the UE can assume zero transmission power. The UE assumes that CSI-RSs are not transmitted in the special subframe(s) in the case of frame structure type 2, in subframes where transmission of a CSI-RS would collide with transmission of synchronization signals, PBCH, or system information block type 1 messages, in subframes configured for transmission of paging messages for any UE with the cell-specific paging configuration. The resource elements (k, l) that are used for transmission of CSI-RS on any of the antenna ports $AP_p = \{15\}, \{15, 16\}, \{17, 18\}, \{19, 20\}, \{21, 22\}$ are not used for transmission of PDSCH on any antenna port in the same slot, and are not used for CSI-RS on any antenna port other than those in the above set of antennas in the same slot. The mapping of CSI-RS for 1, 2, 4, and 8 antenna ports are shown in Figure 9.58 [3].

Depending on the number of CSI-RS antenna ports, there are multiple reuse patterns on different locations which share a common base pattern. In the case of 1, 2, 4, and 8 antenna ports, there are 40, 20, 10, and 5 reuse patterns, respectively, allowing different cells to utilize different reuse patterns to avoid inter-cell CSI-RS collision. In addition, the base patterns for different numbers of CSI-RS antenna ports have a nested structure, allowing for simpler CSI-RS transmitter and receiver implementation. Each antenna port's CSI-RS is spread over two adjacent resource elements across time such that two different CSI-RSs are code-

[15]Downlink power control determines the energy per resource element (EPRE). The term resource element energy denotes the energy prior to cyclic prefix insertion. The EPRE is calculated by averaging the energy over all constellation points based on the modulation scheme applied. A UE may assume downlink cell-specific reference signal EPRE is constant across the downlink system bandwidth and constant across all subframes until different cell-specific RS power information is received. The downlink cell-specific reference-signal EPRE can be derived from the downlink reference-signal transmit-power information provided by higher layers. The downlink reference-signal transmit power is defined as the linear average over the power contributions in Watts of all resource elements that carry cell-specific reference signals within the operating system bandwidth [5].

division multiplexed over two adjacent resource elements. One of the advantages of code-division-multiplexing in this case, relative to other multiplexing schemes such as FDM, is that CDM can balance transmission power across antenna ports in the frequency-domain. The CSI-RS density is related to the channel estimation accuracy. In general, larger CSI-RS density may provide better CSI measurement accuracy while reducing downlink resource utilization. Therefore, adequate CSI-RS density, where the CSI-RS occupies minimum downlink resources while providing reasonable CSI measurement accuracy, is desirable. Since CSI-RS is only used for downlink CSI measurement, its granularity is coarse due to feedback overhead; the system performance is relatively insensitive to the channel measurement accuracy as compared to that for DM-RS. Therefore, a reduced-overhead CSI-RS transmitted only once every 5, 10, 20, 40, or 80 ms is utilized in the LTE-Advanced. A typical transmission of CSI-RS, which may consist of four antenna ports with periodicity of 10 ms, would only require an overhead equivalent to 0.2% of the entire time and frequency resources. The low overhead is achieved by allocating a single resource-element-pair per CSI-RS antenna port, except when the CSI-RS has only one antenna port, where two resource elements are allocated per resource block pair (Table 9.12). The configuration period $T_{\text{CSI-RS}}$ and the subframe offset $T_{\text{Offset-CSI-RS}}$ for CSI-RSs are given in Table 9.13. The parameter $I_{\text{CSI-RS}}$ can be configured separately for CSI-RSs for which the UE assumes non-zero and zero transmission power. Those subframes containing CSI-RSs must satisfy the following equation $\left(10n_{\text{subframe}} + \lfloor n_s/2 \rfloor - T_{\text{Offset- CSI-RS}}\right) \text{mod} T_{\text{CSI-RS}} = 0$ [5,14].

Although different CSI-RS patterns are likely to be used in neighboring cells to minimize CSI-RS collision, neighbor cells' downlink data signals may still significantly interfere with CSI-RS received at a cell-edge UE in a heavily loaded network in interference-limited scenarios. In addition, accurate CSI measurement for neighbor cells is also required for support of coordinated multipoint transmissions. To mitigate the interference effects and to facilitate CSI measurement of neighbor cells, LTE-Advanced supports data reference element muting, whereby data reference elements colliding with CSI-RS in a neighbor cell are transmitted with zero transmission power (i.e., muted resource elements).

Table 9.12 Relative Overhead of CSI-RS

Number of Antenna Ports	1	2	4	8
Cell-specific reference signal	4.76	9.52	14.28	–
UE-specific reference signal	7.14	7.14	14.28	14.28
CSI reference signal	$1.19/T_{\text{CSI-RS}}$	$1.19/T_{\text{CSI-RS}}$	$2.38/T_{\text{CSI-RS}}$	$4.76/T_{\text{CSI-RS}}$

Note that CSI-RS overhead is a function of CSI-RS periodicity $T_{\text{CSI-RS}}$ in number of subframes.

Table 9.13 CSI Reference Signal Subframe Configuration [5]

CSI-RS Subframe Configuration Parameter $I_{CSI\text{-}RS}$	CSI-RS Periodicity $T_{CSI\text{-}RS}$ (Subframes)	CSI-RS Subframe Offset $T_{Offset\text{-}CSI\text{-}RS}$
0–4	5	$I_{CSI\text{-}RS}$
5–14	10	$I_{CSI\text{-}RS} - 5$
15–34	20	$I_{CSI\text{-}RS} - 15$
35–74	40	$I_{CSI\text{-}RS} - 35$
75–154	80	$I_{CSI\text{-}RS} - 75$

For an LTE-Advanced UE, the eNB applies code rate-matching around the muted resource elements so that the LTE-Advanced UE is able to decode the allocated data sub-carriers.

As mentioned earlier, a cell can be configured with one, two, four, or eight antenna-port CSI-RS. The exact CSI-RS structure, including the exact set of resource elements used for CSI-RS in a resource block, depends on the number of antenna ports for which the CSI-RS is configured within the cell and may vary for different cells. There are 40 possible positions within a resource-block pair that the CSI-RS can be mapped, and in a given cell, a subset of the corresponding resource elements is used for CSI-RS transmission. In the case of two antenna-port CSI-RS within a cell, the CSI-RS consists of two adjacent reference symbols per resource-block pair. The CSI-RS pair is separated by applying OCCs to the reference-symbol pair. Thus, there can be 20 different CSI-RS configurations based on the 40 possible reference-symbol positions.

In the case of four or eight antenna-port CSI-RS, the CSI-RS are pairwise frequency multiplexed. Thus, there is a possibility for ten or five different antenna ports for which the CSI-RS is configurations, each supporting four or eight CSI-RS, respectively. In the case of a single antenna-port CSI-RS, the same structure as for two antenna-port CSI-RS is used, although with only one of the OCCs. The overhead of a single antenna-port CSI-RS, in terms of occupied resource elements, is the same as two antenna-port CSI-RSs. In the time-domain, the CSI-RS can be transmitted with different periods, ranging from 5 to 80 ms. With a 5 ms period, the overhead per CSI-RS is roughly 0.12% per CSI-RS (one resource element per CSI-RS per resource-block pair, in every fifth subframe), thus longer periods result in lower overhead. The exact subframe in which CSI-RS is transmitted in a cell can also be configured, allowing for separation of CSI-RS transmissions among cells in the time-domain, in addition to using different sets of resource elements within the same subframe. In subframes in which CSI-RS is transmitted, the CSI-RS is transmitted in every resource block across the frequency. In other words, the CSI-RS transmission covers the entire cell bandwidth and by this means it is considered a wideband reference signal.

As mentioned above, when DM-RSs are transmitted within a resource block, the corresponding resource elements on which the reference signals are transmitted are explicitly avoided when mapping PDSCH symbols to the resource block. This restricted resource mapping, which is obviously not understood by legacy UEs, is possible as DM-RSs can be assumed to be transmitted only in resource blocks in which terminals supporting such reference signals are scheduled. In general, a legacy terminal is assumed to be never scheduled in a resource block where DM-RSs are transmitted. The situation is somewhat different for CSI-RS. As CSI-RS is transmitted within all resource blocks in the frequency-domain, it would not be practically possible to assume that legacy terminals are never scheduled in a resource block in which CSI-RS is transmitted. If the PDSCH mapping were modified to explicitly avoid the resource elements in which CSI-RS is transmitted, the mapping would not be recognized by legacy terminals. In order to alleviate this problem, the PDSCH is mapped consistent with LTE Rel-8, i.e., the mapping is not modified to avoid the resource elements on which CSI-RS is transmitted, and CSI-RS is transmitted on top of the corresponding PDSCH symbols. This will obviously impact the PDSCH demodulation performance, as some PDSCH symbols will be interfered by CSI-RS transmission. However, the remaining PDSCH symbols will not be impacted and the PDSCH will still be decodable with some acceptable performance degradation. On the other hand, if a Rel-10 terminal is scheduled in a resource block in which CSI-RS is transmitted, the PDSCH mapping is modified to explicitly avoid the resource elements on which the CSI-RS is transmitted. The remaining resource elements are used for PDSCH transmission within the cell. However, it is also possible to additionally configure one or several subsets of the resource elements as muted CSI-RS.

A muted CSI-RS has the same structure as an unmuted CSI-RS, except that nothing is actually transmitted on the corresponding resource elements. A muted CSI-RS can thus be seen as a normal CSI-RS with zero power. The motivation for inclusion of the muted CSI-RS is to enable creation of transmission holes corresponding to actual CSI-RS transmissions in neighboring cells in order to: (1) make it possible for a terminal to receive CSI-RS of neighboring cells, without being severely interfered by transmissions in its serving cell. The CSI estimation on neighboring cells would be necessary for multi-cell-transmission techniques and (2) to reduce the interference to CSI-RS transmissions in other cells. This is especially applicable to heterogeneous networks where overlapping layers with cells of substantially different transmit power are expected to coexist in the same frequency band. There can be one or multiple sets of muted CSI-RS in a cell. In the case of interference avoidance to neighboring cells, it is typically sufficient to use a single set of muted CSI-RS, as one should typically avoid interference to a single lower-power cell. When muted CSI-RS are used to alleviate intra-cell interference, in order to improve reception of CSI-RS from neighboring cells, multiple sets of muted CSI-RS are needed since neighboring cells may use different CSI-RS configurations.

9.10.6 Enhanced PDCCH demodulation reference signals

The LTE Rel-8 PDCCH was designed with the underlying assumption that each eNB with collocated antennas serves each cell in the network and different cells are distinguished through different physical cell identifiers. Furthermore, the UE relies on CRSs to detect and decode PDCCH content. However, the increasing mobile data traffic necessitates the use of dense and heterogeneous networks wherein multiple low-power remote radio heads (RRHs) with geographically displaced antennas share the same cell identity with the overlaying macro eNB. The cell capacity may be dramatically increased due to spatially split gain of RRHs. In such cases, if one common PDCCH region is shared among the macro eNB and all RRHs, the time-division-multiplexed PDCCH suffers from limited capacity when confined to three OFDM symbols. Moreover, since the legacy interleaved PDCCH spans across the entire system bandwidth, frequency-domain inter-cell interference coordination schemes for control channels transmitted by different RRHs cannot be utilized by PDCCH.

A Rel-11 work item initiated the design of a new downlink control channel structure referred to as ePDCCH which relies on DM-RSs for coherent detection and is frequency-division-multiplexed with data region. The use of DM-RSs would allow precoding and beamforming of the downlink physical control channels in the future. Figure 9.59 illustrates the structure of ePDCCH in a subframe and how it can coexist with the legacy PDCCH without affecting the interoperability with the legacy LTE systems [178]. As shown in the figure, the ePDCCH regions are time-division-multiplexed with the legacy PDCCH region, are frequency-division-multiplexed with PDSCH, and avoid other essential physical channels and signals. Since ePDCCH only relies on UE-specific reference signals for demodulation, the use of ePDCCH in MBSFN subframes would eliminate the reliance on CRS for control channel demodulation in the future releases of LTE.

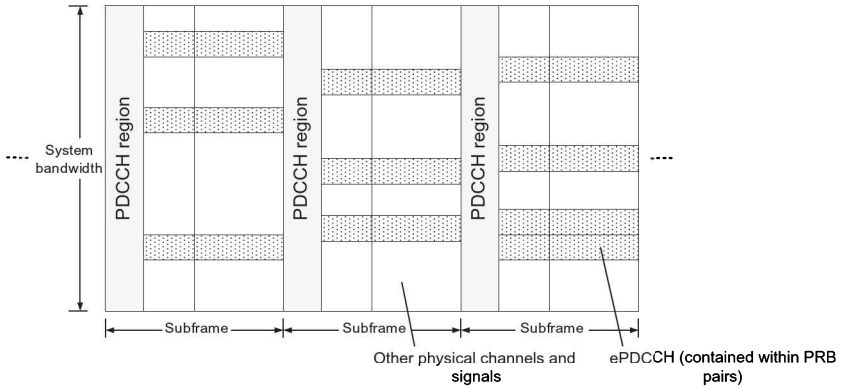


FIGURE 9.59

Enhanced PDCCH structure [149–155].

Each ePDCCH may contain multiple enhanced control channel elements (eCCEs), and multiple eCCEs can be aggregated to achieve different coding rates which depend on the channel conditions of the UE to guarantee a block error rate (BLER) of less than 1%. The existing structure of the PDCCH relies on CRS for channel estimation and coherent detection. It is shown that CRS has more overhead and is less effective when using closed-loop precoding techniques, beamforming, and MU-MIMO. The beamforming of the control channels can significantly improve the performance of the cell-edge users. The use of DM-RS or UE-specific reference signals instead of CRS would facilitate the use of beamforming for the control channels. Figure 9.60 illustrates a hypothetical beam-formed ePDCCH and depicts the localized and distributed ePDCCH mapping to the resource elements and antenna ports [149–155].

The ePDCCH DM-RSs are transmitted on antenna ports $p \in \{107, 108, 109, 110\}$ and are used as a reference for ePDCCH coherent demodulation. An ePDCCH DM-RS is not transmitted over resource element (k, l), where k denotes the frequency index and l denotes the OFDM symbol index within a slot, when one of the physical channels or physical signals are transmitted using the same resource element regardless of their antenna port. For the antenna port $p \in \{107, 108, 109, 110\}$, the reference-signal sequence $r(m)$ is defined by the equation $r(m) = (1/\sqrt{2})(1 - 2c(2m)) + j(1/\sqrt{2})(1 - 2c(2m + 1))$, where $m = 0, 1, \ldots,$

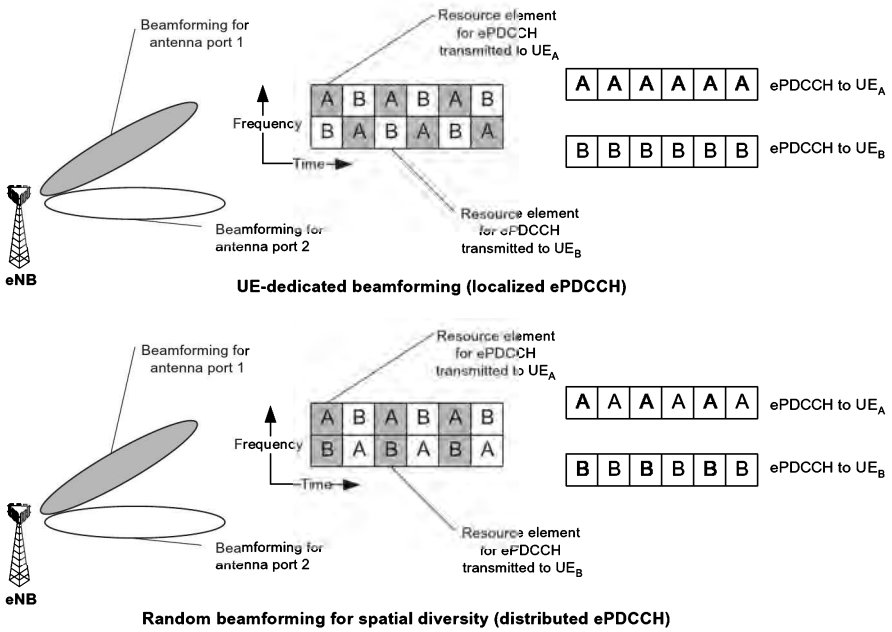


FIGURE 9.60

Possible use of beamforming for transmission of ePDCCH [149–155].

$12\max(N_{RB}^{DL}) - 1$ for a normal cyclic prefix, and $m = 0, 1, \ldots, 16\max(N_{RB}^{DL}) - 1$ for an extended cyclic prefix. The pseudo-random sequence $c(i)$ was defined in the previous sections and is initialized with the initial value $c_{init} = \left(\lfloor n_s/2 \rfloor + 1\right)\left(2N_{ID}^{ePDCCH(i)} + 1\right)2^{16} + N_{SCID}^{ePDCCH}$ at the start of each subframe where $N_{SCID}^{ePDCCH} = 2$ and $N_{ID}^{ePDCCH(i)}|_{i=0,1}$ is configured by higher layers, and index i denotes the set of ePDCCH associated with the DM-RS. For the antenna port $p \in \{107, 108, 109, 110\}$ in a PRB n_{PRB} assigned to the associated ePDCCH, the reference signal sequence $r(m)$ is mapped to complex-valued modulation symbols a_{kl}^{p} in a subframe according to the following mapping rule $a_{kl}^{p} = w_p(l')r(3l'\max(N_{RB}^{DL}) + 3n_{PRB} + m')$ for a normal cyclic prefix, where $w_p(i) = \overline{w}_p(i)$, if $(m' + n_{PRB}) \bmod 2 = 0$, and $w_p(i) = \overline{w}_p(3 - i)$ if $(m' + n_{PRB}) \bmod 2 = 1$ and the frequency index is given by $k = 5m' + 12n_{PRB} + k'$. The length-4 orthogonal cover sequence $\overline{w}_p(i)$ is defined in Table 9.14.

For an extended cyclic prefix, ePDCCH DM-RSs are not supported on antenna ports $p = 109$ and $p = 110$. The resource elements (k, l) used for transmission of ePDCCH DM-RSs to a single UE on any antenna ports in the set $S = \{107, 108\}$ or $S = \{109, 110\}$ is not used for transmission of ePDCCH on any other antenna port in the same slot, and further it is not used for DM-RSs to the same UE on any antenna port other than those in set S in the same slot. The mapping of the ePDCCH DM-RSs to resource elements within a UE-allocated resource block is identical to that corresponding to UE-specific reference signals on antenna port numbers 7–10 when replacing the antenna port numbers with 107–110 in Figure 9.57 for a normal cyclic prefix [3].

9.10.7 Positioning reference signals

Positioning is defined as the process of determining the location and/or velocity of a device using radio signals. The positioning reference signals (PRS) were introduced in LTE Rel-9 on antenna port 6, after the cell-specific reference signals were not found suitable for positioning due to the fact that the required probability of detection could not be guaranteed. A neighbor cell with its synchronization signals (primary/secondary synchronization signals) and reference

Table 9.14 The Sequence $\overline{w}_p(i)$ for Normal Cyclic Prefix [3]

Antenna Port p	$[\overline{w}_p(0)$	$\overline{w}_p(1)$	$\overline{w}_p(2)$	$\overline{w}_p(3)]$
107	[+1	+1	+1	+1]
108	[+1	−1	+1	−1]
109	[+1	+1	+1	+1]
110	[+1	−1	+1	−1]

signals is seen as detectable, when the SINR is at least −6 dB [121,179]. Simulations during development of Rel-9 have shown that this can be only guaranteed for 70% of all cases for the third best-detected cell, which means the second best neighboring cell by assuming an interference-free environment, which cannot be ensured in a realistic scenario. The PRS design has some similarities with cell-specific reference signals as defined in LTE Rel-8. The PRS is a pseudo-random QPSK sequence that is mapped to resource elements in diagonal patterns with shifts in frequency and time to avoid collision with CRSs, and to avoid overlapping with the control channels.

The PRS is defined by its bandwidth N_{RB}^{PRS} in terms of number of resource blocks used for PRS, time offset ($T_{\text{offset-PRS}}$), duration $N_{\text{subframe}}^{PRS}$ in terms of number of consecutive subframes, and its periodicity T_{PRS}. It must be noted that the PRS bandwidth is always smaller than the actual system bandwidth ($N_{RB}^{PRS} < \max(N_{RB}^{DL})$). The PRS is always mapped around the carrier center frequency and the unused DC sub-carrier in the downlink. In a subframe where PRS is configured, typically no PDSCH is transmitted. The positioning reference symbols of a certain cell can be configured to correspond to empty resource elements of the neighboring cells, thus enabling detection of the PRSs with satisfactory SIR values when receiving neighbor-cell PRSs [121].

Positioning reference signals are only transmitted in resource blocks over downlink subframes that are configured for PRS transmission. If both normal and MBSFN subframes are configured as positioning subframes within a cell, the OFDM symbols in the MBSFN subframe use the same cyclic prefix type as the first subframe in the frame. However, if MBSFN subframes are only configured to contain PRSs in a cell, the OFDM symbols configured for PRSs in the MBSFN region of these subframes use an extended cyclic prefix type. In a subframe configured for PRS transmission, the starting positions of the OFDM symbols configured for PRS transmission is identical to those in a subframe in which all OFDM symbols have the same cyclic prefix length as the OFDM symbols configured for PRS transmission. Positioning reference signals are transmitted on antenna port 6, and are not mapped to resource elements (k, l) allocated to BCH and primary and secondary synchronization channels regardless of their associated antenna ports. PRSs are only defined for sub-carrier spacing of 15 kHz.

The reference-signal sequence $r_l^{n_s}(m)$ is defined by $r_l^{n_s}(m) = (1/\sqrt{2})[1 - 2c(2m)] + j(1/\sqrt{2})[1 - 2c(2m+1)]$, $\forall m = 0, 1, \ldots, 2\max(N_{RB}^{DL}) - 1$, where $n_s = 0, 1, \ldots, 19$ is the slot number within a radio frame, $l = 0, 1, \ldots, 6$ is the OFDM symbol number within the slot. The pseudo-random sequence $c(i)$ is defined similar to the previous sections. The pseudo-random sequence generator in this case is initialized by $c_{\text{init}} = 2^{10}[7(n_s + 1) + l + 1](2N_{ID}^{\text{cell}} + 1) + 2N_{ID}^{\text{cell}} + N_{CP}$ at the start of each OFDM symbol where $N_{CP} = 1$ for a normal cyclic prefix and $N_{CP} = 0$ for an extended cyclic prefix. The reference signal sequence $r_l^{n_s}(m)$ is mapped to complex-valued modulation symbols a_{kl}^p designated to the PRS on antenna port $p = 6$ such that $a_{kl}^p = r_l^{n_s}(m')$ where in the case of a normal cyclic prefix

Table 9.15 Positioning Reference Signal Subframe Configuration Parameters [3]

PRS Configuration Index I_{PRS}	PRS Periodicity T_{PRS}(Subframes)	PRS Subframe Offset $T_{offset-PRS}$(Subframes)
0–159	160	I_{PRS}
160–479	320	$I_{PRS} - 160$
480–1119	640	$I_{PRS} - 480$
1120–2399	1280	$I_{PRS} - 1120$
2400–4095		Reserved

$m' = m + \max(N_{RB}^{DL}) - N_{RB}^{PRS}$. The bandwidth for PRSs and N_{RB}^{PRS} is configured by higher layers and the cell-specific frequency shift is given by $\nu_{shift} = N_{ID}^{cell} \bmod 6$.

The cell-specific subframe configuration period T_{PRS} and the cell-specific subframe offset $T_{offset-PRS}$ for transmission of PRSs are given in Table 9.15 (see Figure 9.61). The PRS configuration index I_{PRS} is configured by higher layers. Positioning reference signals are transmitted only in configured downlink subframes and, in the case of TDD duplex scheme, are not transmitted in special subframes. The PRS is transmitted in N_{PRS} consecutive subframes that are configured by higher layers. The PRS location in time is given by $(10n_{subframe} + \lfloor n_s/2 \rfloor - T_{offset-PRS}) \bmod T_{PRS} = 0$. A UE may assume that downlink PRS EPRE is constant across the PRS bandwidth and across all OFDM symbols that contain positioning reference signals in a given PRS occurrence. The structure of the positioning reference signals in the time- and frequency-domain is illustrated in Figure 9.61.

9.10.8 MBSFN reference signals

Support of MBMS in LTE has certain impacts on the physical layer design. It starts with the use of the cyclic prefix, since it is essential for the OFDM system that multipath versions of the signal arrive at the receiver within the cyclic prefix to avoid ISI. As we have multiple, but synchronized transmission from different broadcast/multicast sources, the expected delay spread is much larger than that for unicast operation. Therefore, in order to support multicast and broadcast services, MBSFN uses an extended cyclic prefix while the majority of today's LTE networks are based on a normal cyclic prefix. In the case of the mixed (or time-division-multiplexed) MBSFN and unicast operation, the subframes that are supposed to carry MBMS content are divided into a non-MBSFN and an MBSFN region. The non-MBSFN region can occupy one or two OFDM symbols at the beginning of the subframe. In this region, PCFICH, PDCCH, and PHICH will be mapped to ensure interoperability with user terminals that do not support MBSFN. For proper operation of the network, it is required to schedule terminals to receive or transmit data, page them, or provide feedback on their recent uplink transmission. In this region, CRSs and a normal cyclic prefix will be used. The MBSFN-region will carry PMCH. The extended cyclic prefix and sub-carrier spacing of

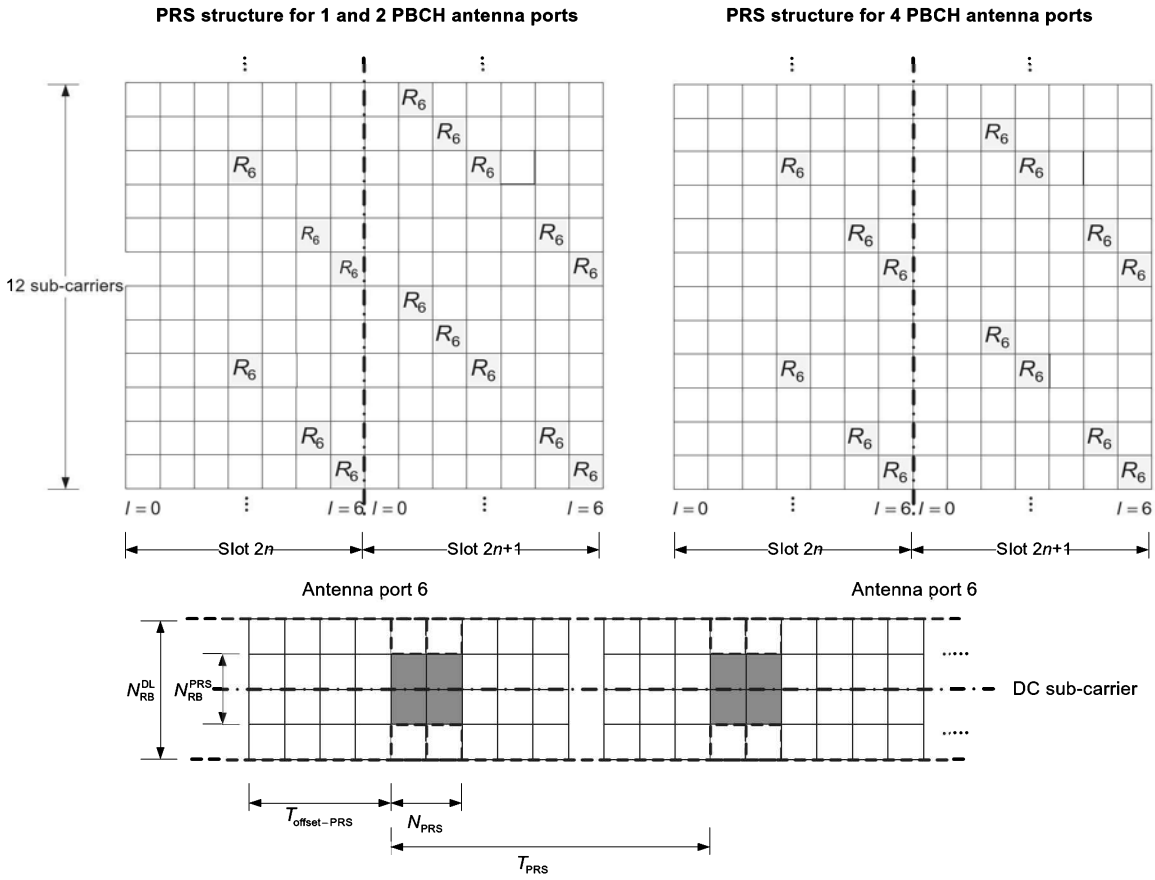


FIGURE 9.61

Structure of PRS in time- and frequency-domain [3,179].

15 kHz (or 7.5 kHz) are used in this region in order to contain the time difference of the cells that belong to an MBSFN area, while transmitting the same content to the terminals. In order to enable coherent demodulation at the terminal, as well as proper channel estimation, the use of CRSs is not sufficient. Thus new reference signals for MBMS transmission have been designed as part of Rel-9. Each cell belonging to the MBSFN area will transmit the same MBSFN reference signal pattern at the exact same time–frequency position. The MBSFN reference signals have a tighter spacing in the frequency-domain due to the frequency-selective nature of the radio channel. In contrast to the CRSs, the initialization sequence for the MBSFN reference signal does not depend on the physical cell identity, rather on the MBSFN identity provided by system information block type 13 [179].

The MBSFN reference signals are transmitted on antenna port 4 in the MBSFN region of MBSFN subframes, only if PMCH is transmitted. These reference signals are only defined for an extended cyclic prefix. The MBSFN reference-signal sequence $r_l^{n_s}(m)$ is obtained as $r_l^{n_s}(m) = (1/\sqrt{2})[1 - 2c(2m)] + j(1/\sqrt{2})\ [1 - 2c(2m+1)]$, $\forall m = 0, 1, \ldots, 6\max(N_{RB}^{DL}) - 1$, where $n_s = 0, 1, \ldots, 19$ is the slot number within a radio frame, $l = 0, 1, \ldots, 5$ is the OFDM symbol number within the slot, and $c(i)$ is the pseudo-random sequence defined similar to the previous sections, which is initialized by $c_{init} = 2^9[7(n_s + 1) + l + 1](2N_{ID}^{MBSFN} + 1) + N_{ID}^{MBSFN}$ at the start of each OFDM symbol. The reference-signal sequence $r_l^{n_s}(m')$ in OFDM symbol l is mapped to complex-valued modulation symbols a_{kl}^p on antenna port $p = 4$ according to $a_{kl}^p = r_l^{n_s}(m')$, where $m' = m + 3(\max(N_{RB}^{DL}) - N_{RB}^{DL})$. Figure 9.62 illustrates the resource elements used for MBSFN reference signal transmission.

9.10.9 Reference signal overhead calculation

The reference signals regardless of their types add to the overhead and reduce the number of available resource elements in a subframe for user data allocation. The amount of overhead typically increases when multiple transmit antennas and/or an extended cyclic prefix is used. Among various types of reference signals used in LTE/LTE-Advanced, the cell-specific reference signals correspond to the highest amount of overhead, as they are always on and span the entire system bandwidth. The overhead concerning other types of reference signals varies depending on the number of antenna ports, periodicity of the configured reference signals, number of UE-allocated resource blocks, etc. The overhead is calculated as the number of resource elements in each resource block pair that is designated to the reference signals relative to the total number of resource elements in each resource block pair. The total number of resource elements is also a function of the cyclic prefix type. Table 9.16 provides the overhead figures associated with each reference signal type for FDD systems. The overhead of the reference signals in TDD systems is also a function of the downlink/uplink subframe configuration and the number of OFDM symbols in the DwPTS portion of the special subframes [13].

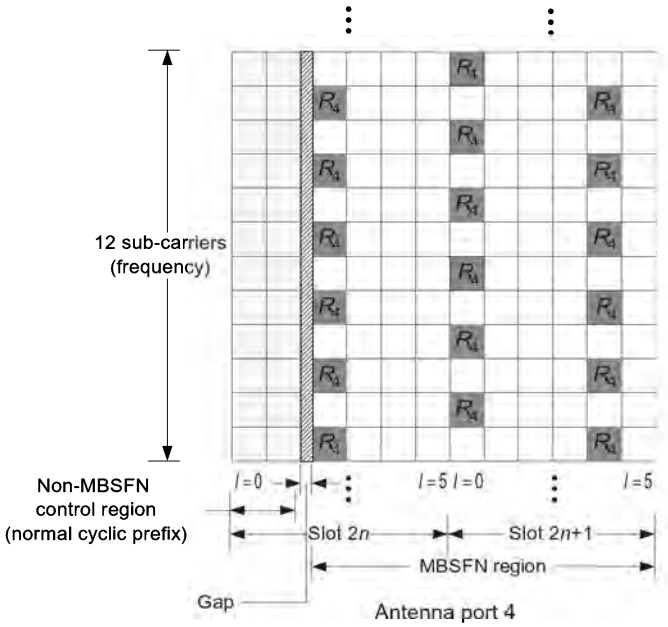


FIGURE 9.62

MBSFN reference signal structure [3,13].

9.11 Channel coding and modulation

Forward error correction (FEC) channel coding schemes are commonly used to improve the efficiency and robustness of wireless digital communication systems. On the transmitter side, an FEC encoder adds redundancy to the input data in the form of parity bits, while at the receiver, the redundancy is utilized by the FEC decoder to correct a number of channel errors. The use of FEC schemes would allow tolerance of more channel errors than otherwise can be tolerated without FEC. The use of FEC schemes enables the wireless communication systems to operate with a lower transmit power, to transmit over longer distances, to tolerate more interference, and to transmit at a higher data rate. A binary FEC encoder receives k bits at a time and produces a codeword of n bits, where $n > k$. While there are 2^n possible sequences of n bits, only 2^k codewords are used. The ratio k/n is the code rate and is denoted by R. Lower code rates corresponding to small values of R can generally correct more channel errors than higher code rates, and therefore are more robust. However, higher code rates are more bandwidth efficient because of lower overhead due to parity or redundancy bits. Thus, the selection of the code rate is a trade-off between energy efficiency and bandwidth efficiency. For every combination of code rate R, code word length n, baseband modulation scheme, channel type, and receiver noise power, there is a theoretical

Table 9.16 Reference Signal Overhead in FDD Mode [13]

Release	Type of Reference Signal	Number of Antenna Ports	Antenna Ports	Number of Reference Signals per Resource Block Pair	Overhead (%) Normal Cyclic Prefix	Overhead (%) Extended Cyclic Prefix
8	Cell-specific reference signals	1	0	8	4.8	5.6
		2	0,1	16	9.5	11.1
		4	0,1,2,3	24	14.3	16.7
	UE-specific (demodulation) reference signals	1	5	12	7.1	8.3
	MBSFN reference signals	1	4	18	–	12.5
9	Positioning reference signals	1	6	16 (14)	9.5	9.7
	UE-specific (demodulation) Reference Signals	2	7,8	12	7.1	8.3
10	UE-specific (demodulation) reference signals	1	7	12	7.1	8.3
		2	7,8	12	7.1	8.3
		4	7–10	24	14.3	16.6
		8	7–14	24	14.3	16.6
	CSI reference signals	1	15	2	$1.19/T_{\text{CSI-RS}}$	$1.39/T_{\text{CSI-RS}}$
		2	15,16	2	$1.19/T_{\text{CSI-RS}}$	$1.39/T_{\text{CSI-RS}}$
		4	15–18	4	$2.38/T_{\text{CSI-RS}}$	$2.78/T_{\text{CSI-RS}}$
		8	15–22	4	$2.38/T_{\text{CSI-RS}}$	$2.78/T_{\text{CSI-RS}}$
11	ePDCCH demodulation reference signals	4	107–110	12(24)	7.1–14.3	8.3–16.6

lower limit on the amount of energy that must be used to convey one bit of information. This limit is called the channel capacity of the communication channel [33]. Since the dawn of information theory, engineers and mathematicians have tried to construct codes that achieve performance close to Shannon capacity [27,33,39]. The introduction of convolutional turbo codes (CTCs) in 1993 followed by the invention of block turbo codes (BTCs) in 1994 closed much of the performance gap with the maximum channel capacity. Turbo codes are essentially parallel concatenated convolutional codes (PCCCs) with an internal interleaving mechanism combined with an iterative soft-decision decoding algorithm. In this section, we first review the principles of convolutional and turbo coding, and then we describe the channel coding and modulation schemes that are used in LTE/LTE-Advanced systems.

Before we proceed to the next sections, we need to clarify that detection, decision-making, hypothesis testing, and decoding are synonymous. The word detection refers to the effort to decide whether some phenomenon is present or not on the basis of some observations. For example, a radar system uses the data to detect whether a target is present. The meaning has been extended in the communication field to detect which one, among a set of mutually exclusive alternatives, is correct. Decision-making is the process of deciding between a number of mutually exclusive alternatives. Hypothesis testing is the same, and here the mutually exclusive alternatives are called hypotheses. Decoding is the process of mapping the received signal into one of the possible sets of codewords or transmitted symbols. We use the word hypotheses for the possible choices in what follows, since the word conjures up the appropriate intuitive image [100–108].

9.11.1 Principles of convolutional coding

Convolutional codes have been extensively used in telecommunication systems such as digital video, radio, mobile, and satellite communication in order to achieve reliable data communication [27]. Prior to the invention of the turbo codes, convolutional code was the most efficient coding scheme that provided the closest performance to the Shannon limit. A generic convolutional encoder (k, n, K) is characterized by a shift register with K stages, k input bits, and n output bits. The rate of the convolutional code is defined as $R = k/n$. A convolutional encoder is a discrete linear time-invariant system and can be viewed as a finite-state machine with 2^{K-1} states. Therefore, convolutional codes are linear codes which do not divide the source stream into blocks, but instead read and transmit bits continuously. The transmitted bits are a linear function of the past source bits. Usually the rule for generating the transmitted bits involves feeding the present source bit into a linear-feedback shift-register of length K, and transmitting one or more linear functions of the state of the shift register in each iteration. The resulting transmitted bit stream is the convolution of the source stream with a linear filter (Figure 9.63). The impulse–response function of this filter may have finite or infinite duration depending on the choice of feedback shift-register,

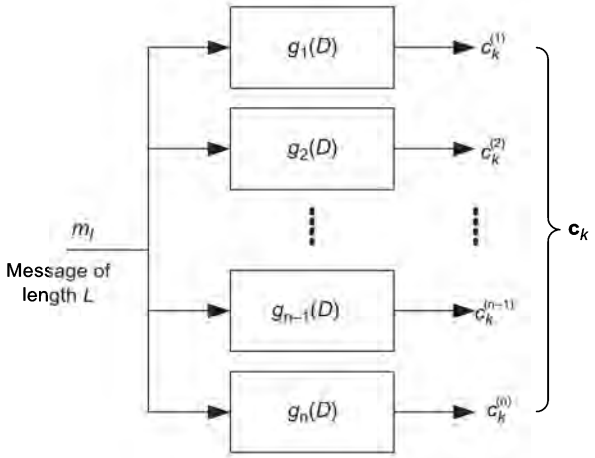


FIGURE 9.63

General form of (1/*n*) convolutional coder [27,40].

i.e., recursive or non-recursive convolutional codes. As an example, consider the LTE convolutional coder with $k = 1, n = 3, K = 6$, and $R = 1/3$. The linear combinations are defined via $n = 3$ generator sequences where $\mathbf{G} = \{\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_{n-1}\}$ and $\mathbf{g}_i = \{g_{i0}, g_{i1}, \ldots, g_{i(K-1)}\}$. The generator sequences used in Rel-8 of LTE are given by $\mathbf{g}_0 = \{1\ 0\ 1\ 1\ 0\ 1\ 1\}$, $\mathbf{g}_1 = \{1\ 1\ 1\ 1\ 0\ 0\ 1\}$, and $\mathbf{g}_2 = \{1\ 1\ 1\ 0\ 1\ 0\ 1\}$ as shown in Figure 9.63. Puncturing is used to make an arbitrary k/n rate code from a basic rate 1/2 or 1/3 code by deleting some bits in the encoder output. For example, if we want to make a code with rate 2/3, we should take the basic 1/2-encoder output and transmit every second bit from the first branch and every bit from the second branch.

Traditional channel coding techniques achieve coding gain at the expense of bandwidth expansion, where coding gain is defined as the reduction in required E_b/N_0 for a given error probability. In the case of band-limited channels, bandwidth expansion is usually not possible. Trellis-coded modulation schemes use redundant non-binary modulation in conjunction with a finite state encoder. For each symbol interval, the encoder selects one of a set of *M*-ary waveforms, thereby generating a sequence of coded waveforms. The noisy signals at the receiver are detected and decoded by a soft-decision maximum-likelihood (ML) detector/decoder. Coding gain can be achieved without sacrificing data rate or expanding bandwidth, but at the expense of increasing decoder complexity.

A convolutional encoder, which is an example of trellis coding, is a finite-state machine, i.e., the output is a function of both the current input and the state of the encoder. A convenient way of characterizing such a state-oriented process is through a trellis diagram. Figure 9.64 illustrates a trellis diagram for a

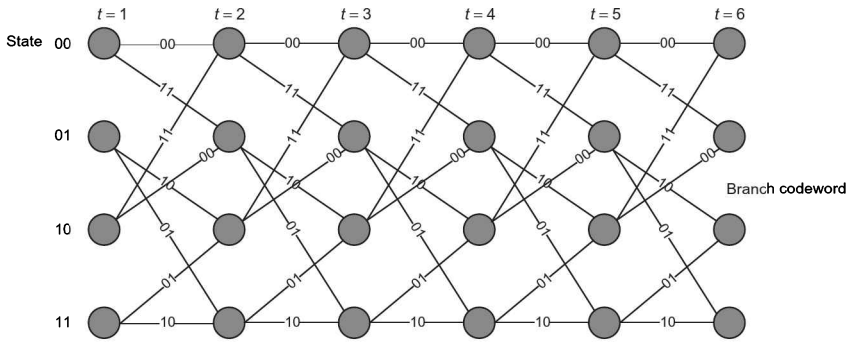


FIGURE 9.64

Four-state Trellis diagram [27,40].

hypothetical four-state convolutional encoder. The number of stages in the encoder shift register is referred to as its constraint length K, and the contents of the rightmost $K-1$ stages define the state of the encoder. Each node of the trellis represents a distinct state at a given time. A branch between any two states represents the transition between those states, and is labeled with the output code sequence associated with that transition. The most important property of the trellis diagram is that every possible sequence of message bits into the encoder corresponds to a unique path through the trellis and a unique sequence of coded bits, and vice versa. The task of the decoder is to estimate the path that the encoded message traverses through the trellis. If all input message sequences are equally likely, a decoder that will achieve the minimum probability of error is one that compares the conditional probabilities $p(\mathbf{z}_j|\mathbf{u}_m)$, where $\mathbf{z}_j$ is the received sequence and $\mathbf{u}_m$ is one of the possible transmitted sequences, and chooses the sequence corresponding to the ML. The decoder chooses $\mathbf{u}_{m'}$ if $p(\mathbf{z}_j|\mathbf{u}_{m'}) = \max_{\forall \mathbf{u}_m} p(\mathbf{z}_j|\mathbf{u}_m)$. Finding the sequence which maximizes $p(\mathbf{z}_j|\mathbf{u}_m)$ is equivalent to finding the sequence $\mathbf{u}_{m'}$ that has the greatest similarity to $\mathbf{z}_j$. Over a hard-decision channel, the similarity is often estimated by the Hamming distance between received and candidate sequences, and over a soft-decision channel it is estimated by the Euclidean distance. Since an ML decoder will choose the trellis path whose corresponding sequence $\mathbf{u}_{m'}$ is at the minimum distance to the received sequence $\mathbf{z}_j$, the ML problem is identical to the problem of finding the shortest distance through the trellis diagram [19]. A comparison of the performance of a rate 1/2 convolutional code with hard- and soft-decision decoding ($K = 7$ and BPSK modulation) over an AWGN channel is shown in Figure 9.66. The advantage of a soft-decision decoding over a hard-decision decoding convolutional coder (approximately $+2$ dB coding gain) can be seen in this example. The performance of the same encoder with hard-decision decoding under Rayleigh multipath fading and AWGN channels for different number paths has been calculated and compared in Figure 9.66.

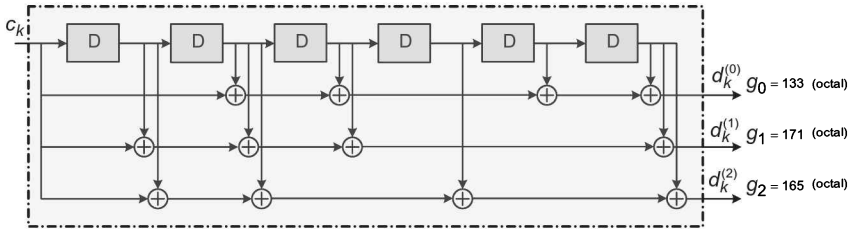


FIGURE 9.65

LTE rate 1/3 tail-biting convolutional encoder [4].

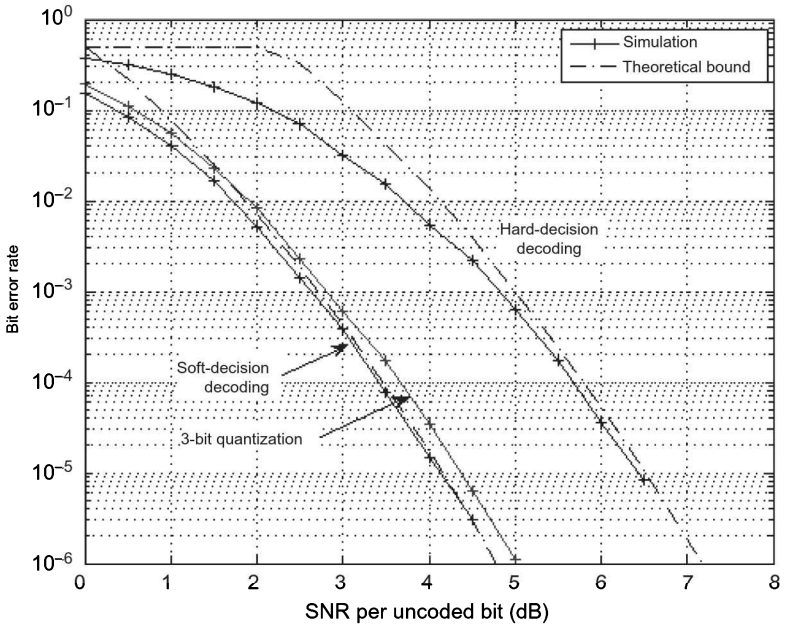


FIGURE 9.66

Performance of rate 1/2 convolutional code ($K = 7$ and BPSK modulation) over AWGN channel [141].

In order to terminate a convolutional code, some bits are appended to the information sequence such that the shift-register's content returns to the initial state. Each of the k input sequences of length L bits is padded with K zeros, and these kL input sequences jointly generate $n(K + L)$ output bits. The effective rate of the terminated convolutional code is now $R_{\text{effective}} = kL/n(K + L) < k/n$, which is less than the nominal rate due to the addition of redundancy. Thus, terminating the trellis of a convolutional code is essential in the code's performance for packet-based communications.

Tail-biting convolutional coding is a technique of trellis termination which avoids the rate loss incurred by zero-tail termination at the expense of a more complex decoder. Tail-biting encoding ensures that the starting state of the encoder is the same as its ending state, and that this state value does not necessarily have to be the all-zero state. For a rate $1/n$ feed-forward encoder, this is achieved by initializing the K memory elements of the encoder with the last K information bits of a block of data of length L and ignoring the output. All of the L bits are then input to the encoder and the resulting nL output bits are used as the codeword. In comparison, the zero-tail termination method appends K zeros to a block of data to ensure the feed-forward encoder starts from and ends in the all-zero state for each block. This incurs a rate loss due to the extra tail bits (i.e., non-informational bits) that are transmitted.

The ML tail-biting decoding involves determining the best path in the trellis under the constraint that it starts and ends in the same state. One way to implement this method is to run M parallel Viterbi algorithms, where M is the number of states in the trellis, and select the decoded bits based on the Viterbi algorithm that gives the best metric. However, this makes the decoding M times more complex relative to the zero-tailed encoding.

The structure of the LTE rate 1/3 tail-biting convolutional encoder with constraint length 7 shown in Figure 9.67. The bit sequence input for a given code block to channel coding is denoted by $(c_0, c_1, \ldots, c_{K-1})$, where K is the number of bits to be encoded. The initial value of the shift register (the initial state) of the encoder is set to the values corresponding to the last 6 information bits in the input stream, so that the initial and final states of the shift register are the same. Therefore, if the content of the shift register is denoted by $(s_0, s_1, s_2, s_3, s_4, s_5)$, then the initial value of the shift register is set to $s_i = c_{(K-1-i)}$. The encoder output streams $d_k^{(0)}$, $d_k^{(1)}$, and $d_k^{(2)}$ correspond to the first, second, and third parity streams [4].

The performance of an example rate 1/2 tail-biting convolution coder in an AWGN channel with different modulation orders, coding rates, and data frame sizes using a Viterbi decoder has been evaluated and shown in Figure 9.68 where the wrap depth is a predetermined number of trellis stages used as a parameter in the figure. One way to estimate the required wrap depth for MAP decoding[16] of a

[16]There are a number of possible criteria to use in making decisions, and initially, we assume that the criterion is to maximize the probability of choosing correctly. That is, when the experiment is performed, the resulting sample point maps into a sample value h for H and a sample value y for Y. The decision-maker observes y (but not h) and maps y into a decision $\hat{H}(y)$. The decision is correct if $\hat{H}(y) = h$. In principal, maximizing the probability of choosing correctly is almost trivially simple. Given y, we calculate $p_{H|Y}(h|y)$ for each $h, 0 \leq h \leq M-1$. This is the probability that h is the correct hypothesis conditional on y. Thus the rule for maximizing the probability of being correct is to choose $\hat{H}(y)$ to be that h for which $p_{H|Y}(h|y)$ is maximized. This is denoted as $\hat{H}(y) = \arg\max_h[p_{H|Y}(h|y)]$. If the maximum is not unique, it makes no difference to the probability of being correct which maximizing h is chosen. To be explicit, we will choose the smallest maximizing h. The conditional probability $p_{H|Y}(h|y)$ is called an *a posteriori* probability, and thus the above decision rule is called the maximum a posteriori probability (MAP) rule [45].

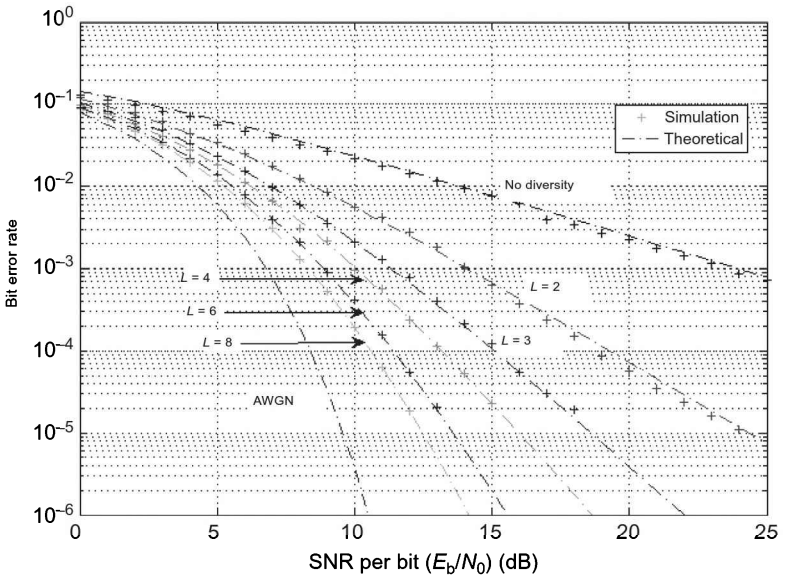


FIGURE 9.67

Performance of rate 1/2 convolutional code (K = 7 and BPSK modulation) in multi-path Rayleigh fading channel [141].

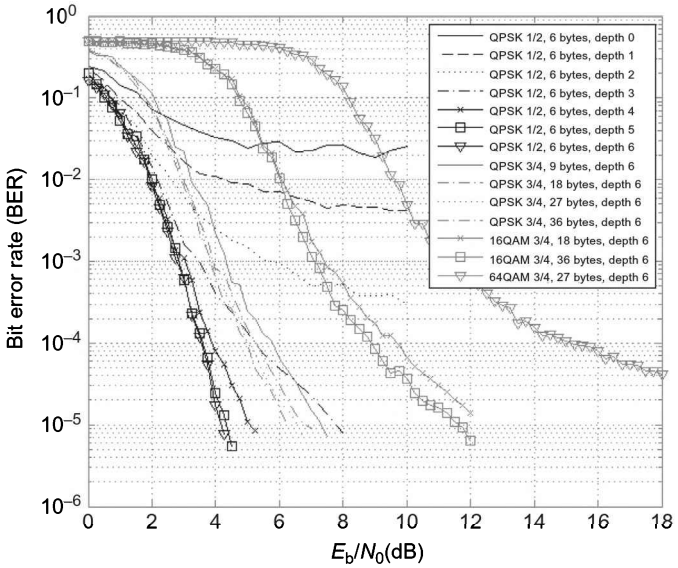


FIGURE 9.68

Performance of an example rate 1/2 tail-biting convolutional coder in an AWGN channel [144].

tail-biting convolutional code is to conduct hardware or software experiments, requiring a circular MAP decoder with a variable wrap depth to be implemented in order to measure the decoded bit error rate versus E_b/N_0 for successively increasing wrap depths. The minimum decoder wrap depth that provides the minimum probability of decoded bit error for a specified E_b/N_0 is found when further increases in wrap depth do not decrease the error probability, as shown in Figure 9.68. It can be seen that the link-level BER performance improves when increasing the depth and the performance does not significantly change after a certain limit. The effect of different data frame sizes (6, 9, 18, 27, and 36 bytes) on the BER performance has been investigated and is shown in the figure.

9.11.2 Principles of turbo coding

In information theory, one can ideally approach the Shannon limit as closely as desired using soft-decision decoding of a long block code, or a convolutional code with a large constraint length. In practice, however, decoding of these codes becomes very computationally intensive. Similar to any conventional error correcting codes, the turbo codes work by imposing a structure on the transmitted bit sequence. If the received bit sequence does not match this known structure, the receiver declares an error has occurred. If the number of errors is sufficiently small and the structure is robust, the receiver can detect the erroneous bits and reconstruct the correct bit sequence. A strong error correcting code has two key characteristics: (1) the encoder imposes a structure that maximizes the difference (or the distance) between any two valid bit sequences; and (2) the decoder utilizes all the information available at the receiver's end including the redundancy and previously unsuccessful transmissions. Thus, the decoder can determine which valid bit sequence is most likely the one that has been transmitted. The first significant difference between a turbo code and conventional code is the use of a recursive systematic encoder. A typical convolutional coder uses a non-recursive structure. By feeding one of the outputs back to the input, a recursive encoding structure is obtained. The recursive structure is systematic, i.e., the input bits appear directly as part of the encoded bit stream. Therefore, a systematic structure enables the combination of two codes in order to construct a stronger composite code. The input bit stream only needs to be transmitted once. The computed parity bits from each of the two constituent encoders are transmitted separately. The recursive structure interacts with the interleaver to give the composite code some unique performance characteristics. The optimum decoder for any code is an ML decoder. In the ML decoding, the receiver produces a probability that each received bit is either 0 or 1. The ML decoder then uses these received bit probabilities, along with its knowledge of the structure imposed by the encoder, to compute the probability of every possible transmitted bit sequence. The most probable transmitted bit sequence is then chosen as the decoded bit sequence. The ML decoding is theoretically optimal

and can achieve the lowest probability of decoder failure or probability of error. However, there is no efficient way to practically compute these probabilities when the block size is large. This leads to the use of suboptimal, but realizable, decoding methods.

The second unique feature of turbo codes is their decoder, which is based on iterative application of ML decoding. The ML decoding of a convolutional code would be easy, if the constraint length (register size) is short; however, short codes are relatively weak. Combining two such codes produces a much stronger code, but now ML decoding becomes intractable. Each shorter code can be ML decoded if we assume the other code has already been decoded. This leads to an iterative decoding strategy, which is to decode the first code and then use the resulting updated probabilities to decode the second code. Once this is done, the updated results can more accurately decode the first code. This decoding strategy shares the limitations of all iterative solution methods, including the possibility of entrapment in a local minimum. The decoder output is not necessarily the true global ML solution. But the study of the corresponding link-level BER curves suggests that it tends to be a very good solution.

In a convolutional code, the input bit sequence usually has tail bits added. This initializes the encoder state to all zeros at the end of the bit sequence. In a turbo coder, this implies that a valid bit sequence cannot contain a single one. It must either be all zeros, or it must contain two or more ones. Because encoding is a linear process, this implies that any valid input sequence must differ from any other valid input sequence in at least two bits. In this case, if we consider only the first decoder's output, it differs from the correct output in two positions. However, the turbo encoder contains an interleaver. It must be noted that the two constituent encoders are operating on the same set of bits, but in a different order. Therefore, although the error bits are close together for the first decoder, they are widely separated in the second decoder's input. As a result, the second decoder's output differs from the correct output in many bit positions. It will see the incorrect bit sequence as highly implausible. This is the basic principle of operation of the turbo codes, and is the key to the high performance of turbo codes at low SNRs. Erroneous information sequences that look reasonable to one decoder are likely to be rejected by the other decoder. This is also the rationale for the error floor at higher SNRs. Since the interleaver is random, there are a few error patterns that will look plausible to both decoders. As the SNR increases, these weak patterns come to dominate the BER curve.

One of the most interesting characteristics of a turbo code is that it is not just a single code. It is, in fact, a combination of two codes that work together to achieve a synergy that would not be possible by merely using one code by itself. As shown in Figure 9.69, a turbo code is formed from the parallel concatenation of two constituent encoders separated by an interleaver. Each constituent encoder can be any form of FEC coder used for conventional data communications. However, in practice, the constituent encoders are identical convolutional

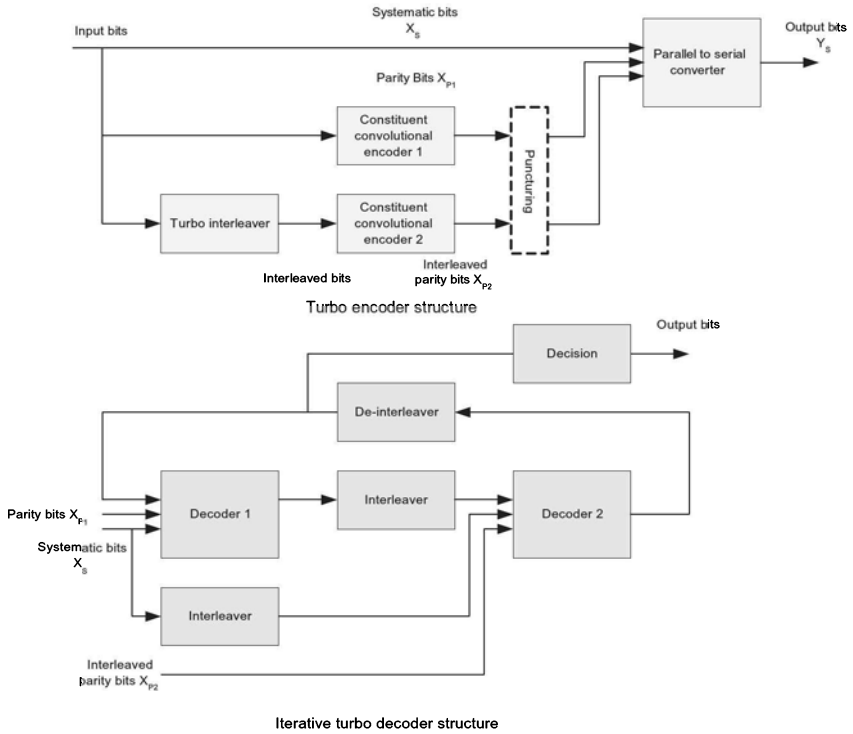


FIGURE 9.69

Generic structure of turbo encoder and iterative decoder [27,40].

encoders. As can be seen in Figure 9.69, the turbo coder consists of two identical constituent encoders. The input data stream and the parity outputs of the two parallel encoders are then serialized into a single turbo codeword. The interleaver is a critical component of the turbo code. It is a simple functional block that rearranges the order of the data bits in a prescribed, but irregular, manner. Although the same set of data bits is present at the output of the interleaver, the order of these bits has been changed. It must be noted that the interleaver used by a turbo coder is quite different than the rectangular interleavers that are commonly used in wireless systems to help mitigate the effect of deep fading. While a rectangular channel interleaver tries to space the data out according to a regular pattern, a turbo interleaver randomizes the ordering of the data in an irregular manner.

The operation of a turbo encoder and decoder can be described as follows. Let the binary input bits $\{0, 1\}$ be represented by bipolar levels $\{+1, -1\}$ and assigned to the variable d, which may take on the values $d = +1$ and $d = -1$. For an AWGN channel, the conditional probability density functions $f(x|d = -1)$

and $f(x|d=+1)$ are referred to as likelihood functions. The common hard-decision criterion, known as ML, selects the symbol $d_k=+1$ or $d_k=-1$ depending on the intercept point of received signal value x_k and the above conditional probability density functions using a fixed threshold of λ (decision point); thus, $d_k=+1$ if $x_k>\lambda$, otherwise, $d_k=-1$. Another decision rule, known as a maximum *a posteriori* probability (MAP) criterion takes into account the *a posteriori* probabilities $f(d=+1|x)$ and $f(d=-1|x)$ to construct the hypotheses H_1 and H_2 as follows:

$$f(d=+1|x) \underset{H_2}{\overset{H_1}{\gtrless}} f(d=-1|x) \tag{9.43}$$

That is, one selects hypothesis $H_1(d=+1)$, if the $f(d=+1|x)>f(d=-1|x)$; otherwise, $H_2(d=-1)$ is selected. Using the Bayes' theorem [30], the above *a posteriori* probabilities can be replaced by their equivalent expressions, yielding:

$$f(x|d=+1)p(d=+1) \underset{H_2}{\overset{H_1}{\gtrless}} f(x|d=-1)p(d=-1) \tag{9.44}$$

The likelihood ratio test is constructed as follows:

$$\frac{f(x|d=+1)p(d=+1)}{f(x|d=-1)p(d=-1)} \underset{H_2}{\overset{H_1}{\gtrless}} 1 \tag{9.45}$$

If the input bits are independent and identically distributed random variables, the above equation is simplified as follows:

$$\frac{f(x|d=+1)}{f(x|d=-1)} \underset{H_2}{\overset{H_1}{\gtrless}} 1 \tag{9.46}$$

By taking the logarithm of the above likelihood ratio, a useful metric known as log-likelihood ratio (LLR) is obtained, which is a real-valued number representing the soft-decision output of the detector:

$$L(d|x)=\log\left(\frac{f(x|d=+1)}{f(x|d=-1)}\right)+\log\left(\frac{p(d=+1)}{p(d=-1)}\right)=L(x|d)+L(d) \tag{9.47}$$

where $L(x|d)$ is the LLR of test statistics x obtained through measuring the channel output x under the condition that either $d=+1$ or $d=-1$ may have been transmitted, and $L(d)$ denotes the *a priori* LLR of the data bit d as shown in Figure 9.70. The introduction of a decoder would improve the reliability of the above decision-making process. For a systematic code, it can be shown that the LLR (soft output) expression at the output of the decoder can be written as $L_{\text{output}}(\hat{d})=L_{\text{input}}(\hat{d})+L_e(\hat{d})$, where $L_{\text{input}}(\hat{d})$ is the LLR of a data bit at the output of the detector (or the input to the decoder), and $L_e(\hat{d})$ represents additional

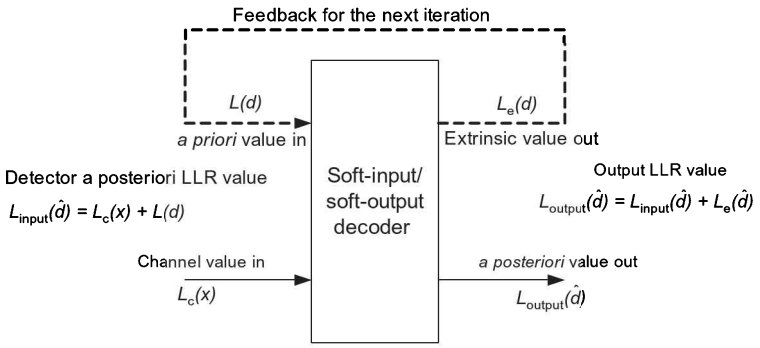


FIGURE 9.70

Illustration of soft-input/soft-output decoder for turbo codes [40,102].

information that is obtained from the decoding process (see Figure 9.70). The output sequence of a systematic decoder consists of values representing data and parity bits. Thus, the decoder output LLR can be decomposed into a data component that is associated with the detector measurement and the extrinsic component that is represented by the decoder contribution due to parity. The soft-decision component $L_{\text{output}}(\hat{d})$ is a real number that provides a hard-decision as well as the reliability of that decision. The sign of the parameter $L_{\text{output}}(\hat{d})$ denotes the hard-decision, i.e., for positive values of $L_{\text{output}}(\hat{d})$, decide $\hat{d} = +1$ and for negative values of $L_{\text{output}}(\hat{d})$, decide $\hat{d} = -1$. The magnitude of $L_{\text{output}}(\hat{d})$ denotes the reliability of the decision [98,102].

Turbo decoding relies on the exchange of probabilistic information between the two soft-input soft-output decoders shown in Figure 9.69. The extrinsic information is the result of the decoder's estimation of bit d, but not taking its own input into account. It is precisely this extrinsic information that is exchanged iteratively between the two soft-decision decoders during the decoding process. Subtracting the decoder's input from its output prevents the decoder from acting as a positive feedback amplifier and introduces stability in the feedback process. Usually, after a given number of iterations, one observes that the two decoders converge to a stable final decision for d. In practice, depending on the nature of the soft-decision decoder, scaling or clipping operations may be applied to the extrinsic information in order to ensure convergence within a small number of iterations. It is shown in the literature that the Max-Log-MAP decoding algorithm can make a good trade-off between performance and complexity, with the added advantage of not requiring any knowledge of the noise level (Figure 9.71). A stop criterion facilitates the convergence of the iterative decoding process and helps reduce the average power consumption of the decoder by reducing the average number of iterations required to decode a block without compromising performance [100–108].

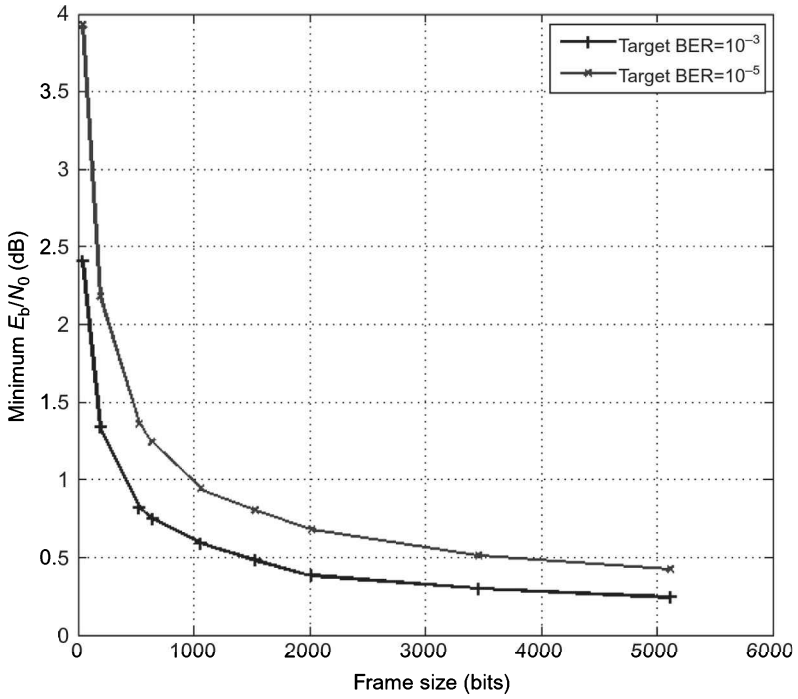


FIGURE 9.71

Minimum $E_b = N_0$ to achieve a target BER using the UMTS/LTE turbo coder ($R = 1/3$) with BPSK modulation over an AWGN channel [144].

For practical applications, it is often necessary to adapt the turbo code parameters to those of the application. The systematic bits are always present at the output of the turbo encoder; however, depending on the desired code rate, the parity bits from the two constituent encoders may be punctured prior to transmission (see Figure 9.69). The puncturing mechanism involves deleting some of the parity bits prior to transmission. As an example, if the minimum code rate of a turbo coder is 1/5, a code rate of 1/3 can be generated by deleting the second parity output of each encoder. Note that the puncturing pattern for each code rate must be known by the decoder for correct interpretation of the received bits [40].

It is shown that for smaller input frame sizes, less iteration is required; however, the performance of the turbo coder is improved by increasing the frame size. This is due to an increase in interleaver gain as the input frame size becomes larger. Figure 9.71 shows the minimum E_b/N_0 required to achieve a BER of 10^{-3} and 10^{-5} for various frame sizes using a UMTS/LTE turbo encoder with a minimum code rate of $R = 1/3$. In general, turbo codes are more suited for applications with large packet sizes and their performance degrades for small packet sizes.

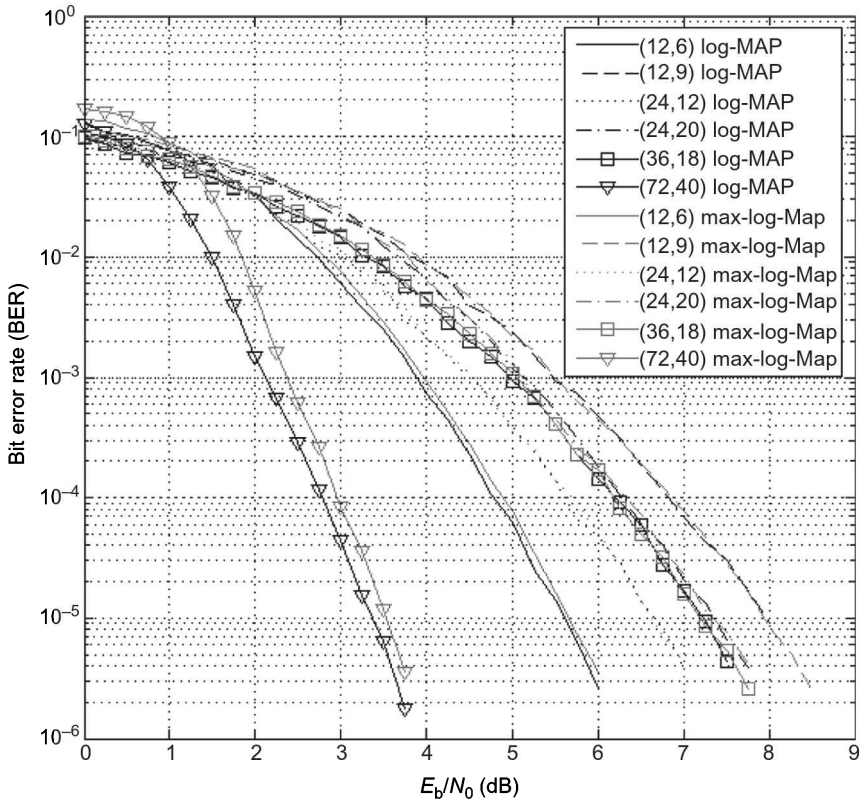


FIGURE 9.72

Comparison of the BER performance of a BTC turbo coder in AWGN channel with different code rates, frame sizes, and detector types [144].

Furthermore, the delay due to turbo interleaver might become prohibitive for delay-sensitive applications. Figure 9.72 shows the BER performance of a BTC turbo coder as a function of E_b/N_0 for various code rates, frame sizes, and detector types. In Figure 9.72, notation (n, k) denotes the number of input k and output n bytes [144].

9.11.3 Baseband digital modulation schemes

Modulation is a process by which a carrier signal is altered according to information in a message signal. For the typical baseband modulation types (PSK, DPSK, FSK, QAM, or PAM), gray-coded constellations are obtained by selecting the gray parameter in the corresponding baseband modulation function. The mapping

or assignment of k information bits to $M = 2^k$ possible signal amplitudes may be done in a number of ways. The preferred assignment is the one in which the adjacent signal amplitudes differ by one binary digit since the most likely errors caused by noise involve the erroneous selection of an adjacent amplitude to the transmitted signal amplitude. In that case, only a single bit error occurs in the k-bit sequence. This mapping is referred to as gray coding. The criteria that must be taken into account when choosing a digital modulation method include the following:

- Power efficiency, i.e., the E_b/N_0 ratio for a specific error probability
- Bandwidth efficiency, i.e., the data rate per unit bandwidth
- Performance in multipath fading channels and under non-linear distortion
- Implementation cost and complexity

However, the above requirements cannot be satisfied at the same time, and some trade-off is necessary. In order to explain the latter statement, let us define a few basic terms in this context. A constellation diagram is a graphical representation of the complex envelope of each possible symbol state. The power efficiency is related to the minimum distance between the points in the constellation. The bandwidth efficiency is related to the number of points in the constellation. The gray coding is used to assign groups of bits to each constellation point. In gray coding, adjacent constellation points differ by a single bit.

The two most common digital modulation schemes used in cellular systems are M-ary PSK and M-ary QAM modulation schemes. In M-ary PSK, a group of n bits are transmitted using $M = 2^n$ different phases, whereas M-ary QAM uses a combination of amplitude and phase modulation to convey the information. Study of the literature suggests that the minimum code distance d_{min} is smaller for M-ary ASK than for M-ary PSK or M-ary QAM for the same average transmitted power, which translates into poorer error performance because the signal envelope is not constant, it results in poor performance under non-linear distortion condition. The minimum distance d_{min} is smaller for M-ary PSK compared to that for M-ary QAM for the same average transmitted power, resulting in poorer error performance. However, a constant envelope results in good performance under nonlinear distortion. The minimum distance d_{min} is larger for M-ary QAM relative to that for M-ary ASK or M-ary PSK for the same average transmitted power, which means better error performance. Nonetheless, since the signal envelope is not constant, poor error performance is expected under non-linear distortion. Pulse shaping is then used to reduce the bandwidth requirements of the modulated signal. The Nyquist technique can be used to reduce the bandwidth requirement and to eliminate ISI. The non-Nyquist techniques reduce the bandwidth requirement, but do not eliminate ISI. Note that in all M-ary modulation schemes, a group of n bits is transmitted in each signaling interval $T = T_b \log_2 M$ and the modulated signal bandwidth is inversely proportional to the signaling interval. It can be shown that the probability of error (i.e., error per bit) equations as a function of E_b/N_0 for M-ary

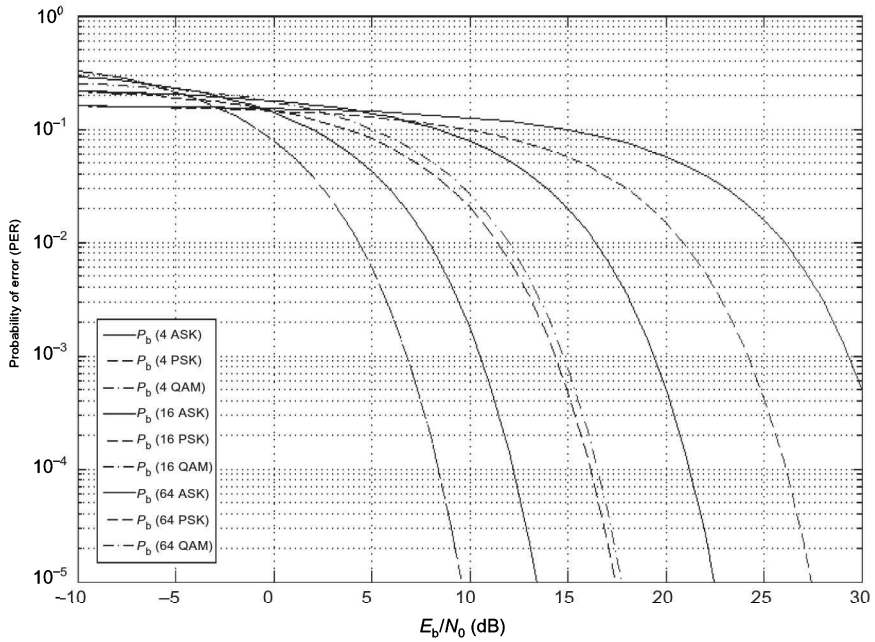


FIGURE 9.73

Probability of error as a function of energy per bit for different digital modulation schemes [159].

ASK, M-ary PSK, and M-ary QAM can be written as follows (see Figure 9.73 for graphical presentation):

$$
\begin{aligned}
p_b &= \frac{1}{\log_2 M}\left(1-\frac{1}{M}\right) Q\left(\sqrt{\frac{3\log_2 M}{M^2-1}\frac{E_b}{N_0}}\right) && M\text{-ary ASK}\\
p_b &= \frac{1}{\log_2 M} Q\left(\sqrt{\log_2 M \frac{E_b}{N_0}\sin^2\left(\frac{\pi}{M}\right)}\right) && M\text{-ary PSK}\\
p_b &= \frac{2}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3\log_2 M}{2(M-1)}\frac{E_b}{N_0}}\right) && M\text{-ary QAM}
\end{aligned}
\tag{9.48}
$$

In the above equations, E_b/N_0 denotes the energy per bit, M is the modulation order, and $Q(\cdot)$ denotes the error function.[17]

[17]In mathematics, the error function is defined as $Q(x) = (2/\sqrt{\pi})\int_0^x e^{-\tau^2}\,d\tau$. An alternative definition is sometimes referred to as the complementary error function in the literature is given as $Q(x) = (2/\sqrt{\pi})\int_x^\infty e^{-\tau^2}\,d\tau$ [30].

The modulation mapping function in the transmission chain takes binary input $b_0 b_1 \ldots b_{N_{\text{modulation order}}-1}$ with $b_i =$ "0" or "1" and produces the complex-valued modulation symbols $x = \gamma(I + jQ)$ in the output with the value of γ chosen to achieve equal average power. For each modulation, symbol, b_0 denotes the most significant bit. In the case of BPSK modulation, $N_{\text{modulation order}} = 1$; thus a single bit is mapped to the modulation symbol x with $\gamma = 1/\sqrt{2}$. In the case of QPSK modulation, $N_{\text{modulation order}} = 2$ and a pair of bits are mapped to complex-valued modulation symbol x with $\gamma = 1/\sqrt{2}$. In the case of 16QAM modulation, $N_{\text{modulation order}} = 4$ and quadruplets of bits $b_0 b_1 b_2 b_3$ are mapped to the complex-valued modulation symbol x and $\gamma = 1/\sqrt{10}$. In the case of 64QAM modulation, $N_{\text{modulation order}} = 6$ and six information bits $b_0 b_1 b_2 b_3 b_4 b_5$ are mapped to the modulation symbol x with $\gamma = 1/\sqrt{42}$. The constellation format for various modulation schemes is illustrated in Figure 9.74.

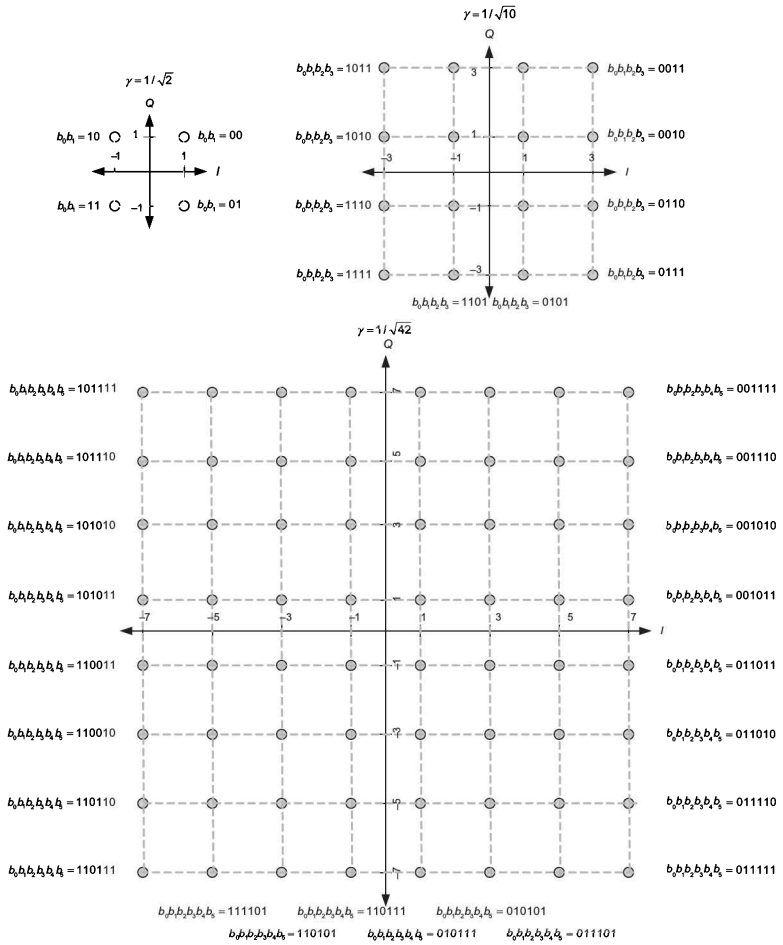


FIGURE 9.74

QPSK, 16QAM, and 64QAM constellations [3].

Figures 9.75 and 9.76 show the performance of the M-ary QAM and M-ary PSK signals in an AWGN channel in terms of probability of error as a function of SNR per bit for different values of modulation order M. It can be seen that for a given target error rate, the required SNR per bit increases with increasing modulation order, i.e., higher modulation orders can only be used in good channel conditions. The latter suggests that in poor channel conditions, more robust modulation schemes must be used.

9.11.4 Channel coding, multiplexing, and interleaving

The baseband modulation schemes supported in the downlink and uplink of LTE are QPSK, 16QAM, and 64QAM. While all three modulation schemes are supported in the downlink in all UE categories, the use of 64QAM in the uplink is only supported in the higher UE categories, and lower UE categories only support QPSK and 16QAM. The channel coding scheme for transport blocks in LTE is turbo coding similar to that of UMTS, with a basic code rate of $R = 1/3$, two eight-state constituent encoders and a contention-free quadratic permutation polynomial (QPP)[18] turbo interleaver [4]. Trellis termination is performed by taking the tail bits from the shift register feedback after all information bits are encoded. The tail bits are padded after the encoding of information bits. Before the turbo coding, transport blocks are segmented into octet-aligned segments with a maximum information block size of 6144 bits. Error detection is performed using 24-bit CRC. Link adaptation (adaptive modulation and coding) with various modulation schemes and channel coding rates is applied to the shared data channel. The same coding and modulation is applied to all groups of resource blocks

[18]In mathematics, a permutation polynomial for a given finite ring is a polynomial that performs a permutation of the elements of the ring, i.e., the mapping Z/nZ is a one-to-one mapping. In the case of finite rings Z/nZ, such polynomials have been studied and applied in the interleaver component of error detection and correction algorithms. For the finite ring Z/nZ, one can construct quadratic permutation polynomials. This is possible, if and only if, n is divisible by p^2 for some prime number p. The construction is simple, but it can produce permutations with desirable properties. As an example, consider $f(x) = 2x^2 + x$ for the ring $Z/4Z$. One can observe that $f(0) = 0$; $f(1) = 3; f(2) = 2; f(3) = 1$, hence the polynomial defines the permutation $\begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 3 & 2 & 1 \end{bmatrix}$. Note that a ring is an algebraic structure which generalizes the main properties of the addition and the multiplication of integers, real numbers, complex numbers, and square matrices. It consists of a set together with two binary operations usually called addition and multiplication, which satisfy the following set of axioms: (1) the addition is associative and commutative, it has an identity and each element in the set has an additive inverse, thus the set is an Abelian group (or commutative group) under addition; (2) the multiplication is associative, but is not necessarily commutative, it has an identity and distributes over addition [31,32].
In general, let us assume $n = \alpha_1^{k_1}\alpha_2^{k_2}\ldots\alpha_m^{k_m}$, in which α_i is a prime number. It can be proved that any polynomial $f(x) = \beta_0 + \sum_{0<i\le M}\beta_i x^i$ defines a permutation for the ring Z/nZ if and only if all the polynomials $f_{\alpha_k}(x) = \beta_{0_{\alpha_k}} + \sum_{0<i\le M}\beta_{i_{\alpha_k}} x^i$ define the permutations for all rings $Z/\alpha_m^{k_m}Z$, in which $\beta_{i_{\alpha_k}}$ is the remainder of β_i modulo $\alpha_i^{k_i}$ [31,32].

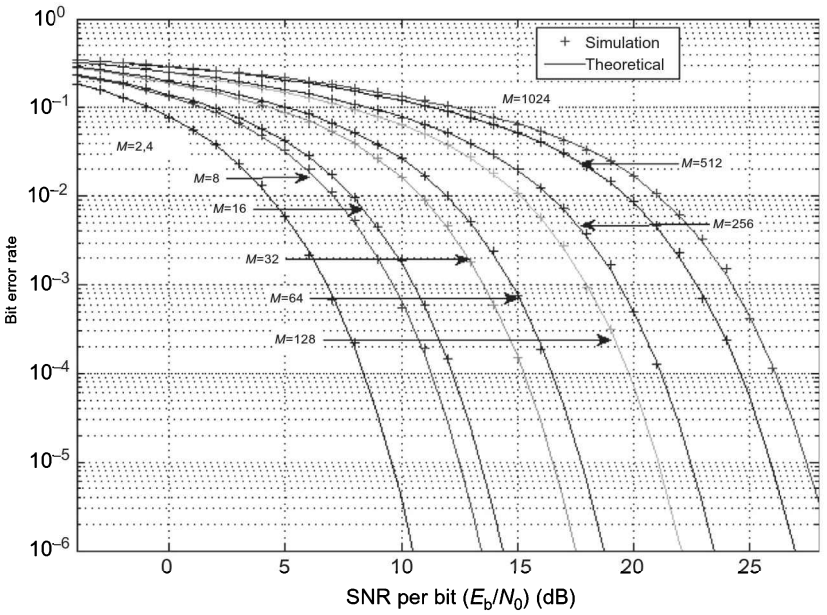


FIGURE 9.75

Performance of *M*-ary QAM in an AWGN channel [141].

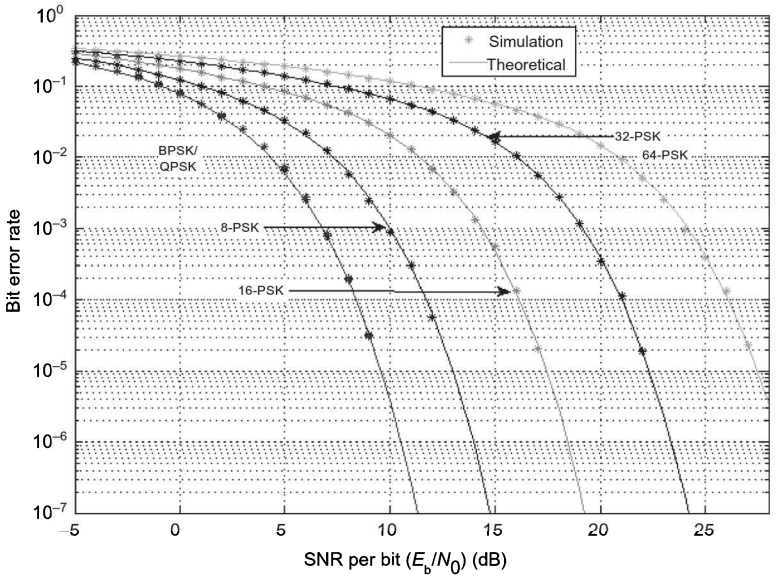


FIGURE 9.76

Performance of coherent *M*-ary PSK over an AWGN channel [141].

belonging to the same MAC PDU scheduled for one user within one TTI and within a single stream.

The user data and control information from the MAC sub-layer are encoded prior to transmission of the transport blocks over the radio link. The channel coding in E-UTRA is a combination of error detection, error correction, rate-matching, interleaving and transport channel, or control information mapping to the appropriate physical channels. As shown in Figure 9.77, the channel coding procedure for BCH transport blocks is different from that of the DL-SCH, PCH, and MCH. Depending on the type of the transport channel, a different coding scheme and coding rate is applied as given in Table 9.17. The first stage of the channel coding process is to append 16/24 bit CRC to the input transport block.

In the case of BCH, the information (i.e., MIB) is received by the channel coding unit in the form of one transport block every 40 ms. A 16-bit CRC is calculated over the entire transport block and appended to the input transport block. Let $(a_0, a_1, \ldots, a_{N_A-1})$ denote the bits in a BCH transport block delivered to the physical layer. The parity bits are denoted by $(p_0, p_1, \ldots, p_{N_P-1})$ where $N_P = 16$.

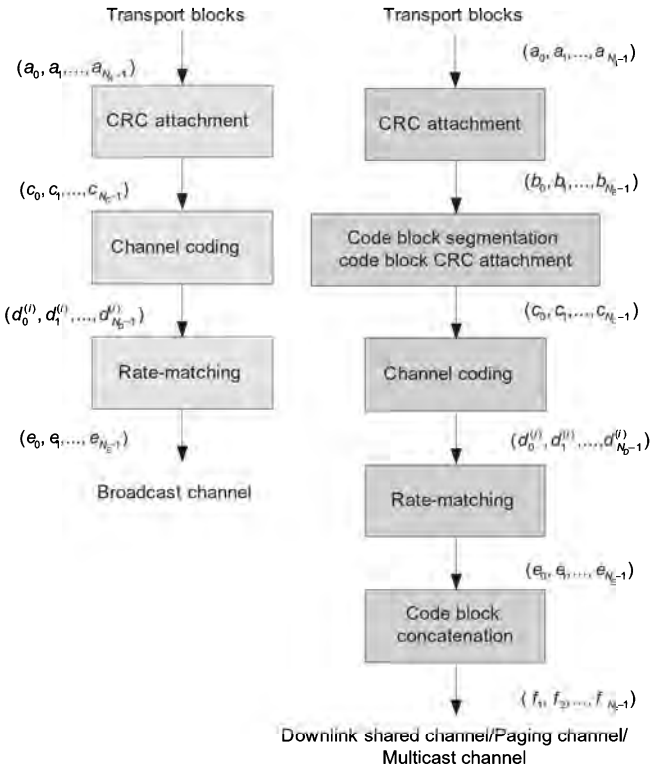


FIGURE 9.77

Channel coding procedure for downlink transport blocks [4].

Table 9.17 Channel Coding Scheme for Various Transport Channels/Control Information in E-UTRA [4]

Channel/ Information	Channel/ Information Types	Coding Scheme	Coding Rate	CRC Size (Bits)
Transport channels	Uplink shared channel (UL-SCH)	Turbo coding	1/3	24
	Downlink shared channel (DL-SCH)			24
	Paging channel (PCH)			24
	Multicast channel (MCH)			24
	Broadcast channel (BCH)	Tail-biting convolutional coding	1/3	16
Control Information	Downlink control information (DCI)	Tail-biting convolutional coding	1/3	16
	Control format indicator (CFI)	Block code	1/16	–
	HARQ indicator (HI)	Repetition code	1/3	–
	Uplink control information (UCI)	Block code	Variable	16
		Tail-biting convolutional coding	1/3	

Following the calculation of the CRC bits, they are scrambled according to the eNB multi-antenna configuration with binary sequence $(x_0, x_1, \ldots, x_{15})$ to form the output sequence $(c_0, c_1, \ldots, c_{N_C-1})$, where the first A bits of the latter sequence are formed by the input transport block and the last 16 bits are the masked CRC bits. The masking sequence is given as follows:

$$(x_0, x_1, \ldots, x_{15}) = \begin{cases} (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0) & N_{TX} = 1 \\ (1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1) & N_{TX} = 2 \\ (0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1) & N_{TX} = 4 \end{cases} \quad (9.49)$$

where N_{TX} denotes the number of antenna ports in the eNB that are used for transmission of the PBCH. The CRC-attached information bits denoted by $(c_0, c_1, \ldots, c_{N_C-1})$ are processed through a tail-biting convolutional coder. The channel-coded bits, denoted by $(d_0^{(i)}, d_1^{(i)}, \ldots, d_{N_D-1}^{(i)})$ where $i = 0, 1, 2$ is the coded stream number to sub-block interleaver and $N_D = N_C$ is the number of bits on the ith coded stream, are fed to the rate-matching block. The use of tail-biting convolutional coding necessitates that the initial value of the shift register of the encoder be set to the values corresponding to the last 6 information bits in the input stream so that the initial and final states of the shift register are the same.

The rate-matching function for channel-coded transport channels is defined per coded block, and consists of interleaving of the three coded streams

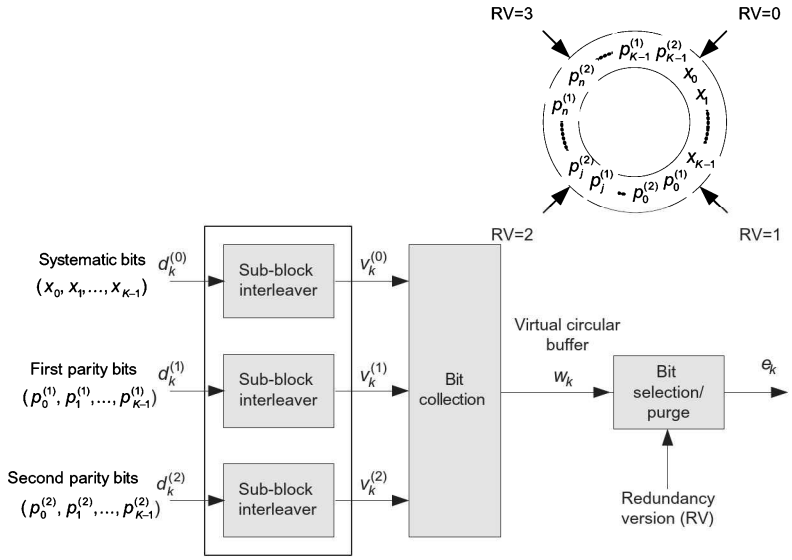


FIGURE 9.78

Rate-matching process of the coded information [4,15].

$d_k^{(0)}$,$d_k^{(1)}$, and $d_k^{(2)}$ followed by the collection of bits and generation of a virtual circular buffer as depicted in Figure 9.78.

The rate-matching block generates an output bit stream with the desired code rate. As the number of bits available for transmission depends on the available resources, the rate-matching algorithm is capable of producing any arbitrary code rate. The three bit streams from the channel coding module are interleaved followed by bit collection to generate a circular buffer. Bits are selected and pruned from the circular buffer to create an output bit stream with the desired code rate, as shown in Figure 9.78. The three sub-block interleavers used in the rate-matching block are identical. Interleaving is a technique to reduce the impact of burst errors on a signal so that consecutive bits of data will not be corrupted. The sub-block interleaver reshapes the coded bit sequence, row-by-row, to form a matrix with $C_{\text{sub-block}} = 32$ columns and $R_{\text{sub-block}}$ rows, where $R_{\text{sub-block}}$ is determined by finding the minimum integer that satisfies the following equation: $R_{\text{sub-block}} C_{\text{sub-block}} \leq N_D$. If $R_{\text{sub-block}} C_{\text{sub-block}} > N_D$, then N_{NULL} null bits are appended to the front of the encoded input sequence such that $R_{\text{sub-block}} C_{\text{sub-block}} = N_D + N_{\text{NULL}}$. For $d_k^{(0)}$ and $d_k^{(1)}$ bit streams, an inter-column permutation is performed on the matrix to reorder the columns as follows: 0, 16, 8, 24, 4, 20, 12, 28, 2, 18, 10, 26, 6, 22, 14, 30, 1, 17, 9, 25, 5, 21, 13, 29, 3, 19, 11, 27, 7, 23, 15, and 31 for turbo-coded transport blocks, and 1, 17, 9, 25, 5, 21, 13, 29, 3, 19, 11, 27, 7, 23, 15, 31, 0, 16, 8, 24, 4, 20, 12, 28, 2, 18, 10, 26, 6, 22, 14, and 30 for transport blocks coded by convolutional coder. For the $d_k^{(0)}$ and $d_k^{(1)}$ bit

streams, the output of the block interleaver is the bit sequence readout column-by-column from the inter-column permutated matrix to create a stream $K' = R_{\text{sub-block}} C_{\text{sub-block}}$ bits long. For the $d_k^{(2)}$ bit stream, the elements in the matrix are permutated separately based on the permutation pattern shown above, but modified to create a permutation which is a function of $(R_{\text{sub-block}}, C_{\text{sub-block}}, k, K')$. This creates three interleaved bit-stream mapping $d_k^{(0)} \rightarrow v_k^{(0)}; d_k^{(1)} \rightarrow v_k^{(1)}; d_k^{(2)} \rightarrow v_k^{(2)}$. The bit collection stage creates a virtual circular buffer by combining the three interleaved encoded bit streams. The block interleaver output $v_k^{(1)}$ and $v_k^{(2)}$ are combined by successive interlacing of their values. This combination is then appended to the end of $v_k^{(0)}$ to create the circular buffer w_k as illustrated in Figure 9.79. Interlacing allows equal levels of protection for each parity sequence [4].

Following the bit collection procedure, the bits are selected and trimmed from the virtual circular buffer to create an output sequence length which meets the desired code rate. The HARQ error correction scheme is incorporated into the rate-matching algorithm of LTE/LTE-Advanced. For any desired code rate, the coded bits are sent out serially from the circular buffer from a starting location, given by the HARQ redundancy version (see Figure 9.78), wrapping around to the beginning of the buffer, if the end of the buffer is reached. The null bits are discarded. Different redundancy versions, and hence starting points, allow for the retransmission of selected data. Being able to select different starting points enables the two main methods of recovering data at the receiver in the HARQ process: (1) chase combining where retransmissions contain the same data and parity bits; (2) incremental redundancy where retransmissions contain different information so the receiver gains more knowledge upon each retransmission. The rate-matched code blocks are now concatenated back together. This is done by sequentially concatenating the rate-matched blocks together to create the output of the channel coding process $(f_1, f_2, \ldots, f_{F-1})$, as shown in Figure 9.77.

In the case of DL-SCH, PCH, and MCH transport channels, the CRC calculation and addition to the information block is similar to BCH, except that 24-bit CRC is used and no CRC masking is done since in this case, the multi-antenna configuration is part of DCI. As shown in Figure 9.80, the input bits to the code block segmentation are denoted by $(b_0, b_1, \ldots, b_{N_B-1})$, where $N_B = N_A + 24$. In LTE/LTE-Advanced, a minimum code block size of 40 bits and maximum

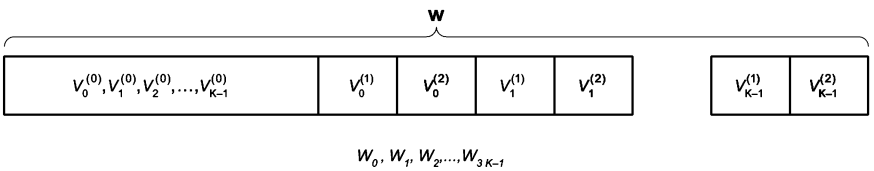


FIGURE 9.79

Structure and content of the virtual circular buffer [160].

code block size of 6144 bits are specified so that the block sizes are compatible with the block sizes supported by the turbo interleaver. If the length of the input block N_B is greater than the maximum code block size, the input block is segmented. When the input block is segmented, it is segmented into $N_C = \lceil N_B/6120 \rceil$ code blocks. Each code block is appended with a 24-bit CRC to the end. Padding bits are added to the front of each block so that code block sizes match a set of permissible turbo interleaver block sizes. If no segmentation is necessary, only one code block is generated. If N_B is less than 40 bits, then padding bits (zeros) are added to the beginning of the code block to attain a total of 40 bits. The code blocks then undergo turbo coding.

A PCCC turbo coder with two recursive convolutional encoders and a contention-free QPP interleaver is used as a channel coder in LTE/LTE/Advanced for traffic channels as shown in Figure 9.81.

The outputs of the encoder are three streams $d_k^{(0)}, d_k^{(1)}$, and $d_k^{(2)}$ in order to achieve a code rate of 1/3. As shown in Figure 9.81, the input to the first constituent encoder is the input bit stream to the turbo coding block. The input to the second constituent encoder is the output of the QPP interleaver, which is a permutated version of the input sequence. There are two output sequences from each encoder which are the systematic bits (x_k, x'_k) and parity bits (z_k, z'_k). Only one of the systematic bit sequences x_k is used as the output, as the other one x'_k is simply a permutated version of the former systematic sequence. The transfer function for each constituent encoder is given by $G(D) = [1, g_1(D)/g_0(D)]$. The first

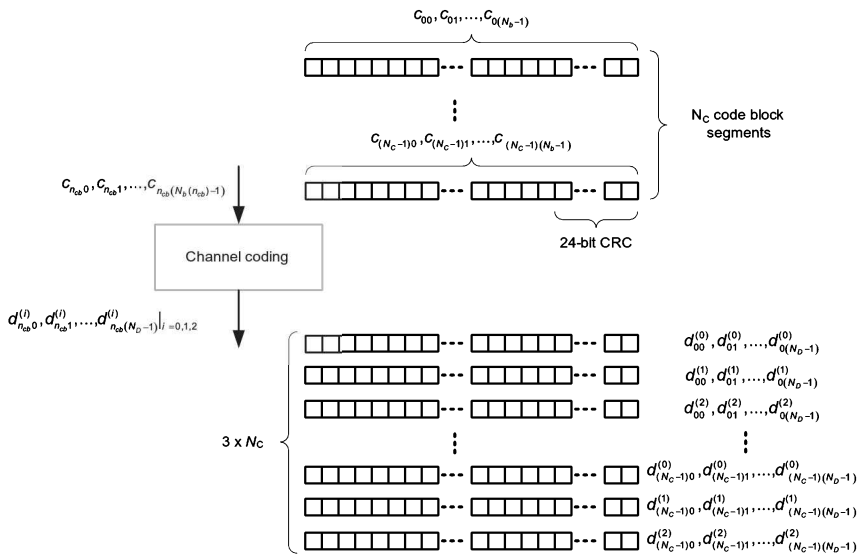


FIGURE 9.80

Code block segmentation and CRC attachment for DL-SCH [4,158,160].

element in the latter equation represents the systematic output transfer function, and the second element denotes the recursive convolutional output transfer function, where $g_0(D) = 1 + D^2 + D^3; g_1(D) = 1 + D + D^3$. The output for each sequence can be calculated using the transfer function.

The encoder is initialized with all zeros. If the code block to be encoded is the zeroth code block and the number of padding bits $N_{\text{Padding}} > 0$, then the encoder input is set to zero for $k = 0, 1, \ldots, N_{\text{Padding}} - 1$, and $d_k^{(0)}, d_k^{(1)}, k = 0, 1, \ldots, N_{\text{Padding}} - 1$ are set to null at the output. In a conventional convolutional coder, the coder is driven to an all zero state upon termination by appending zeros to the end of the input data stream. Since the decoder is aware of the start and end state of the encoder, it can decode the data. Driving a recursive coder to an all zeros state using this method is not possible. To overcome this problem trellis termination is used where, upon termination, the tail bits are fed back to the input of each encoder using a switch (see Figure 9.81). The first three tail bits are used to terminate each encoder.

The role of the interleaver is to spread the information bits such that in the event of an error, the two codes are affected differently allowing data to still be recoverable. As shown in Figure 9.81, the input bits to the turbo code QPP interleaver are denoted by $c_0, c_1, \ldots, c_{N_C-1}$, where N_C is the number of input bits. The output bits from the turbo code QPP interleaver are denoted by $c'_0, c'_1, \ldots, c'_{K-1}$. The output of the interleaver is a permutated version of the input data $c'_i = c_{\Phi(i)}, i = 0, 1, \ldots, K - 1$ where $\Phi(i) = (f_1 i + f_2 i^2) \bmod K$. The coefficients f_1

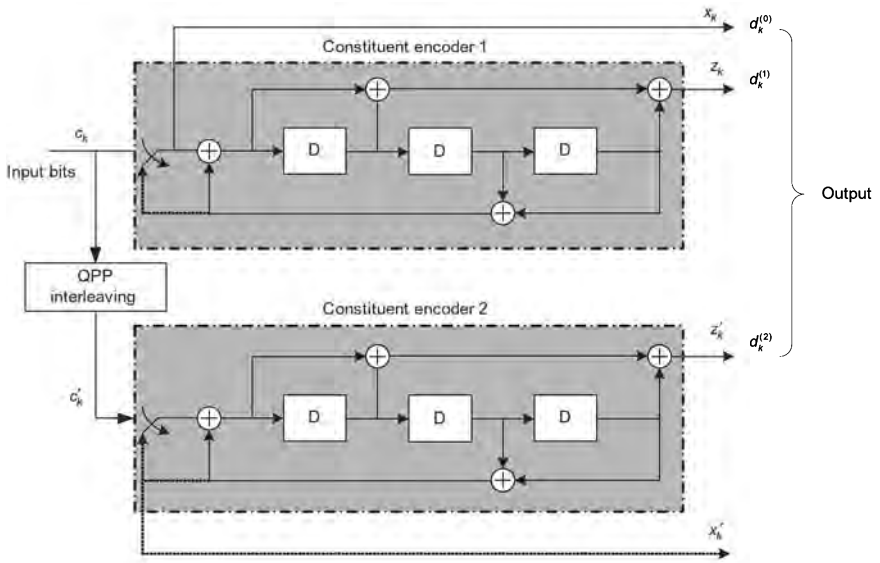


FIGURE 9.81

Structure of an LTE turbo coder [4].

and f_2 are chosen depending on the length of the input data. As an example, if $K = 40$ then $f_1 = 3$, $f_2 = 10$, consequently one can find the permutation pattern as $\Phi(i) = \{0, 13, 6, 19, 12, 25, 18, 31, 24, 37, 30, 3, 36, 9, 2, 15, 8, 21, 14, 27, 20, 33, 26, 39, 32, 5, 38, \ldots\}$. Following processing at the rate-matching block, as explained earlier and shown in Figure 9.82, the rate-matched code blocks are sequentially concatenated to create the output of the channel coding process as $f_k, k = 0, 1, \ldots, N_F - 1$ (Figure 9.83). The sequence of coded bits corresponding to one transport block following code block concatenation is referred to as one codeword. In the case of multiple transport blocks per TTI, the transport block to codeword mapping is performed depending on the DCI Format [3,4].

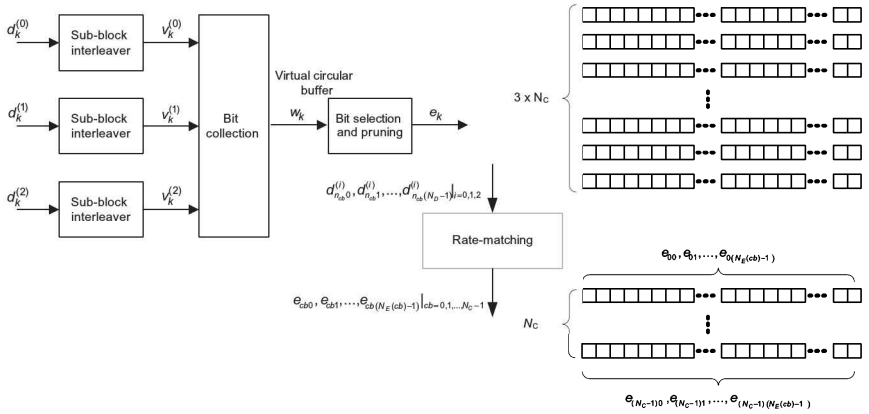


FIGURE 9.82

Illustration of the rate-matching process [158].

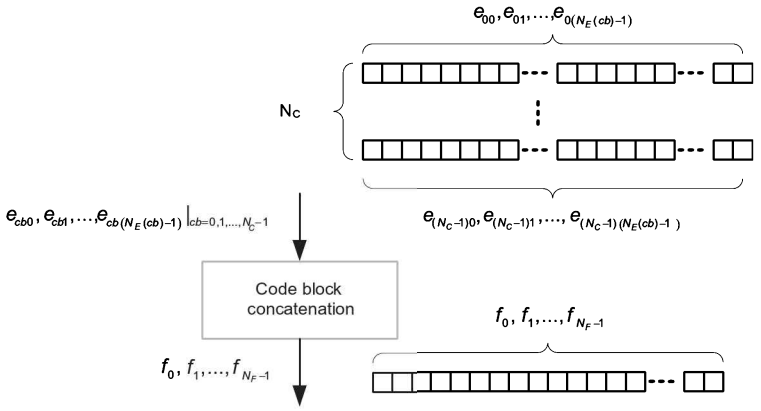


FIGURE 9.83

Illustration of the block concatenation process [158].

9.12 Downlink physical channel processing

This section describes the general physical layer processing structure through which physical channels are processed. As shown in Figure 9.84, data arrives at the coding unit in the form of a maximum of two transport blocks in each TTI. The transport blocks are appended with a CRC which is further scrambled with user identity. If the size of the transport blocks is larger than the maximum coding block size permitted by the physical layer, the code blocks are segmented, and a new CRC is calculated and affixed to each code block. Channel coding and rate-matching are later performed, and the code blocks are concatenated to create the codewords that are the input to the physical channel processing units. In order to generate the baseband signal corresponding to a downlink physical channel, the following processing steps are taken in sequence: (1) scrambling of coded bits in each of the codewords to be transmitted on a physical channel; (2) modulation of scrambled bits to generate complex-valued modulation symbols; (3) mapping of the complex-valued modulation symbols to one or several transmission layers; (4) precoding of the complex-valued modulation symbols on each layer for transmission on the antenna ports; (5) mapping of complex-valued modulation symbols for each antenna port to resource elements; and (6) generation of a complex-valued time-domain OFDM signal for each antenna port. Figure 9.84 depicts the entire physical channel processing from the point that one transport block is delivered to the physical layer by the MAC sub-layer to the point where the $I-Q$ components of the baseband signal to be transmitted on each antenna port are generated.

One or two coded transport blocks (or codewords) can be transmitted simultaneously on PDSCH depending on the precoding scheme used. The codewords undergo scrambling, modulation, layer mapping, precoding and resource element mapping procedures as shown in Figure 9.84. In the following, the function of each block in Figure 9.84 will be described. The specific details about physical layer processing pertaining to physical channels will be described in the next sections.

The block of coded bits $(b_0^l, b_1^l, \ldots, b_{(N_c^l-1)}^l)$ corresponding to codeword $l = 0, 1$, where N_c^l is the number of bits in codeword l transmitted on a physical channel in one subframe are scrambled prior to modulation, resulting in a block of scrambled

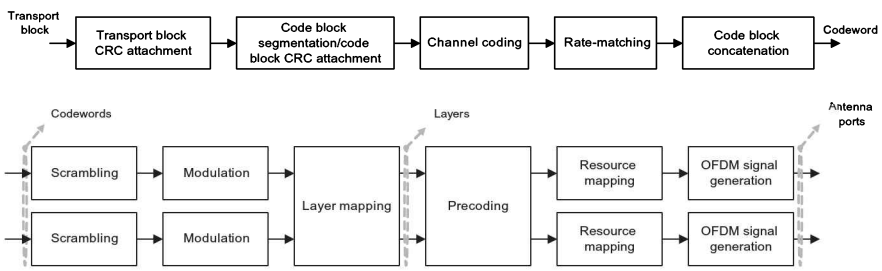


FIGURE 9.84

Overview of downlink physical channel processing in LTE/LTE-Advanced [3,4].

bits $(\hat{b}_0^l, \hat{b}_1^l, \ldots, \hat{b}_{(N_c^l-1)}^l)$ in which $\hat{b}_i^l = (b_i^l + c^l(i))\text{mod } 2$. The scrambling sequence $c^l(i)$ was defined in previous sections. The scrambling sequence generator is initialized at the start of each subframe with the value of c_{init} whose definition depends on the transport channel type as follows: $c_{\text{init}} = n_{\text{RNTI}}2^{14} + l2^{13} + \lfloor n_s/2 \rfloor 2^9 + N_{\text{ID}}^{\text{cell}}$ for PDSCH, and $c_{\text{init}} = \lfloor n_s/2 \rfloor 2^9 + N_{\text{ID}}^{\text{MBSFN}}$ for PMCH. In the expression for PDSCH scrambling sequence initialization, n_{RNTI} corresponds to the RNTI associated with the PDSCH transmission to a specific user. The maximum number of codewords that can be transmitted in one downlink subframe is two.

For each codeword, the block of scrambled bits $(\hat{b}_0^l, \hat{b}_1^l, \ldots, \hat{b}_{(N_c^l-1)}^l)$ are modulated using one of the modulation schemes supported in the downlink, i.e., QPSK, 16QAM, 64QAM, resulting in a block of complex-valued modulation symbols $(d_0^l, d_1^l, \ldots, d_{(N_s^l-1)}^l)$. The complex-valued modulation symbols for each of the codewords are mapped to one or several layers. The complex-valued modulation symbols $(d_0^l, d_1^l, \ldots, d_{(N_s^l-1)}^l)$ for each codeword are mapped to the layers $\mathbf{x}_i = (x_i^0, x_i^1, \ldots, x_i^{N_l-1})^{\text{T}}$, $i = 0, 1, \ldots, N_s^l - 1$ where N_l is the number of layers and N_s^l is the number of modulation symbols per layer.

It will be discussed later in this chapter that LTE Rel-8/9 supports multi-antenna schemes in both downlink and uplink directions. In the downlink direction, up to four transmit antennas may be used with the maximum number of codewords of two, irrespective of the number of antennas. Advanced antenna techniques such as spatial multiplexing and beamforming were supported in the downlink to increase peak data rates and capacity. In downlink SU-MIMO, the eNB may transmit one or more layers to the same UE to take advantage of capacity gains associated with using multiple transmit/receive antennas in rich scattering environments. The UE provides the channel quality feedback and the preferred number of layers as well as the precoding matrix index to the eNB. In downlink MU-MIMO, the eNB attempts to increase capacity by transmitting to multiple users using its multiple transmit antennas at the same time and also using the same time-frequency resources. The UEs must feedback channel state information so that the eNB can schedule the UEs with sufficient channel separation together. The transmitted data is precoded to maximize throughput for users and minimize inter-user interference. In uplink direction, only MU-MIMO was supported, i.e., there is only one modulated symbol stream per UE to be received by the eNB; however, multiple UEs may transmit on the same time–frequency resources. Considering the specified UE categories for LTE Rel-8/9, one can expect two transmit antenna in the downlink and one transmit antenna in the uplink to be the baseline configuration for early LTE deployments. LTE-Advanced extended the MIMO capabilities of LTE Rel-8/9 to eight transmit antennas in the downlink and four transmit antennas in the uplink as well as new SU-MIMO and MU-MIMO modes in the downlink and uplink.

When using transmit diversity, the layer mapping is done with only one codeword, and the number of layers N_l is equal to the number of antenna ports used for

transmission of the corresponding physical channel. In the case of transmit diversity with two layers, even-indexed modulation symbols are mapped to layer 0 and odd-indexed modulation symbols are mapped to layer 1. In the case of transmit diversity with four layers, the input symbols are mapped to layers sequentially as shown in Figure 9.85. If the total number of input symbols is not divisible by four ($N_s^0 \bmod 4 \neq 0$), then two null symbols are appended to the end of the symbol sequence. This creates a total number of symbols which is an integer multiple of four, as the original number of symbols will always be an integer multiple of two [3,160].

For spatial multiplexing, the mapping of complex-valued modulation symbols to spatial layers depends on the number of codewords and the number of layers, as illustrated in Figures 9.85 and 9.86 for single codeword and dual codeword layer mapping, respectively. In this case, the number of layers N_l is less than or equal to the number of antenna ports used for transmission of the corresponding physical channel. The case of a single codeword mapped to multiple layers is only applicable when the number of CRSs is four, or when the number of UE-specific reference signals is two or larger. If only one layer is used, then $x_i^0 = d_i^0$ for $i = 0, 1, \ldots, N_s^l - 1$. When one codeword is mapped to two layers, even-indexed modulation symbols are mapped to layer 0 and odd-indexed modulation symbols are mapped to layer 1. If two codewords are mapped to two layers, then codeword 0 is mapped to layer 0 and codeword 1 is mapped to layer 1. In the case of three layers, codeword 0 is mapped to layer 0, even symbols within codeword 1 are mapped to layer 1, and odd symbols within codeword 1 are mapped to layer 2. In the case of four layers, two codewords must be used. In this case, the even symbols of codeword 0 are mapped to layer 0, and the odd symbols are mapped to layer 1. The even symbols of codeword 1 are mapped to layer 2, and the odd symbols are mapped to layer 3. The above layer mapping schemes for spatial multiplexing are illustrated in Figures 9.85 and 9.86. Note that LTE Rel-10 supports downlink spatial multiplexing up to eight layers. The mapping of modulation symbols to 5, 6, 7, or 8 layers with two codewords is simply the extension of the above symbol mapping method [3].

The precoding block in Figure 9.84 receives input complex-valued vectors $\mathbf{x}_i = (x_i^0, x_i^1, \ldots, x_i^{N_l-1})^{\mathrm{T}}$, with $i = 0, 1, \ldots, N_s^l - 1$ and N_s^l denoting the number of modulation symbols per layer, from the layer-mapping block and generates a block of complex-valued vectors $\mathbf{y}_i = (y_i^0, y_i^1, \ldots, y_i^{N_P-1})^{\mathrm{T}}$ to be mapped to resource elements on each antenna port, where $i = 0, 1, \ldots, N_s^p - 1$, N_s^p denotes the number of symbols per antenna port, N_p is the number of antenna ports used for transmission of the corresponding physical channel, and y_i^p represents the ith symbol for antenna port p. For transmission on a single antenna port, the precoding function is defined as $y_i^p = x_i^0$ where $p \in \{0, 4, 5, 7, 8\}$ is the number of the single-antenna ports used for transmission of the physical channel, $i = 0, 1, \ldots, N_s^p - 1$, and $N_s^p = N_s^l$ [3].

The precoding function for spatial multiplexing using antenna ports associated with CRSs is only used in conjunction with layer mapping and supports two or four antenna ports. The set of antenna ports is given by $p \in \{0, 1\}$ or $p \in \{0, 1, 2, 3\}$.

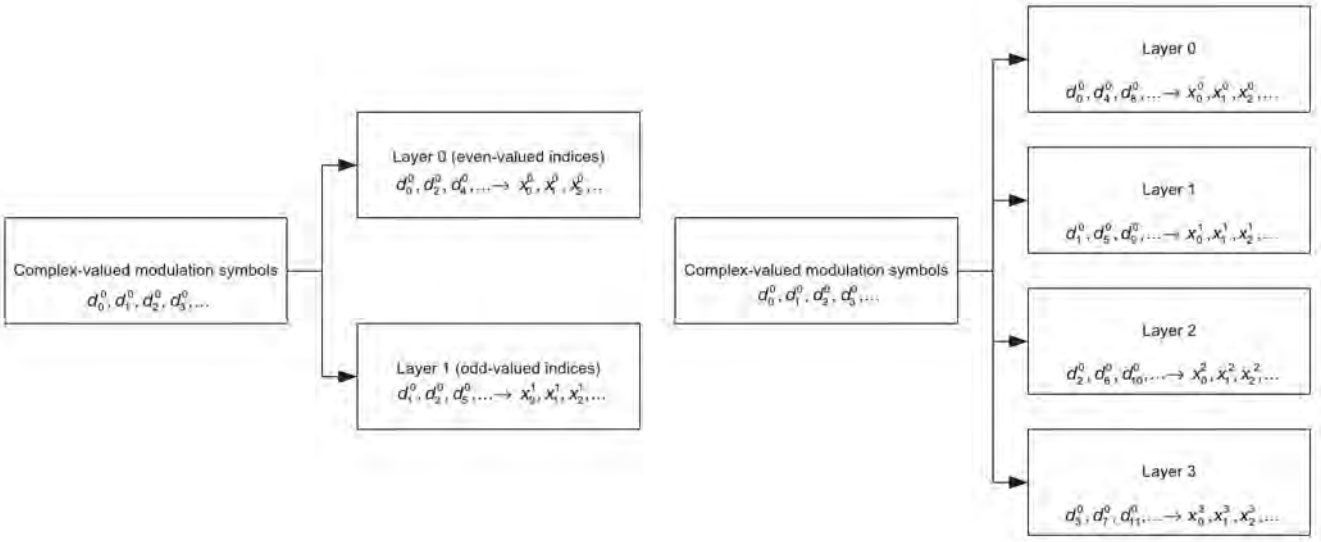


FIGURE 9.85

Layer mapping for 2 and 4 layer transmit diversity (single codeword) [3,160].

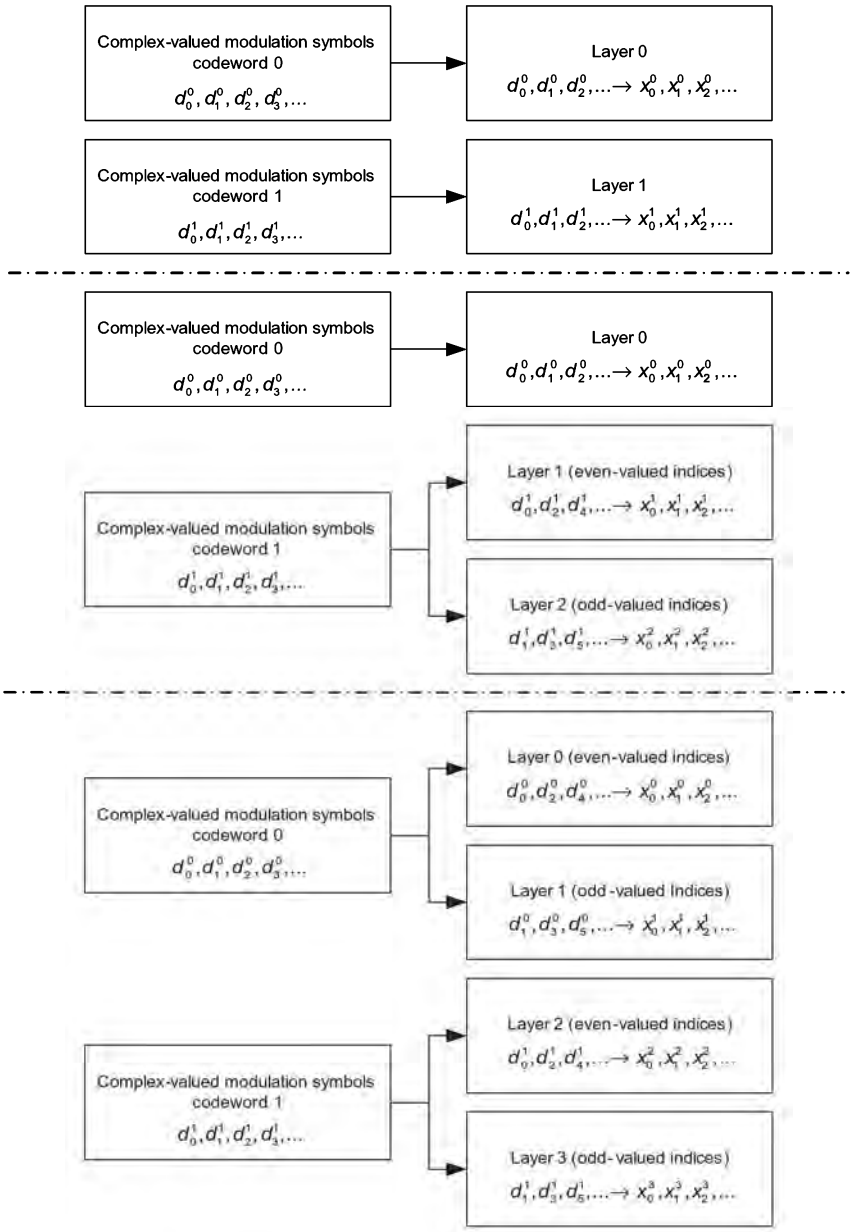


FIGURE 9.86

Layer mapping for spatial multiplexing with 2, 3, and 4 layers (dual codeword) [3,160].

The precoding function for spatial multiplexing without cyclic delay diversity (CDD) is defined as $(y_i^0, y_i^1, \ldots, y_i^{N_p-1})^T = \mathbf{W}_i(x_i^0, x_i^1, \ldots, x_i^{N_l-1})^T$ where the precoding matrix $\mathbf{W}_i$ is of size $N_p \times N_l$ and $i = 0, 1, \ldots, N_s^p - 1$, and $N_s^p = N_s^l$. The values of $\mathbf{W}_i$ are selected among the precoder elements in the codebook configured and stored in the eNB and the UE. The eNB can further restrict the precoder selection in the UE to a subset of the elements in the codebook.

In the case of precoding using large-delay CDD, the precoder for spatial multiplexing is defined as $(y_i^0, y_i^1, \ldots, y_i^{N_p-1})^T = \mathbf{W}_i\mathbf{D}_i\mathbf{U}(x_i^0, x_i^1, \ldots, x_i^{N_l-1})^T$, where the precoding matrix $\mathbf{W}_i$ is of size $N_p \times N_l$ and $i = 0, 1, \ldots, N_s^p - 1$ and $N_s^p = N_s^l$. The diagonal matrix $\mathbf{D}_i$ that is an $N_l \times N_l$ matrix which defines CDD, and the square matrix $\mathbf{U}$ is given in Table 9.18 for different numbers of transmission layers N_l. Similar to the previous scenario, the values of $\mathbf{W}_i$ are selected among the precoder elements in the codebook configured and stored in the eNB and the UE. The eNB can limit the precoder selection in the UE to a subset of the elements in the codebook.

Precoding for transmit diversity is only used in combination with layer mapping. The precoding operation for transmit diversity is defined for two and four antenna ports. For transmission on two antenna ports $p \in \{0, 1\}$, the output vector $\mathbf{y}_i = (y_i^0, y_i^1)^T$ with $i = 0, 1, \ldots, N_s^p - 1$ is defined as follows:

$$\begin{bmatrix} y_{2i}^0 \\ y_{2i}^1 \\ y_{2i+1}^0 \\ y_{2i+1}^1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & \mathrm{j} & 0 \\ 0 & -1 & 0 & \mathrm{j} \\ 0 & 1 & 0 & \mathrm{j} \\ 1 & 0 & -\mathrm{j} & 0 \end{bmatrix} \begin{bmatrix} \mathrm{Re}(x_i^0) \\ \mathrm{Re}(x_i^1) \\ \mathrm{Im}(x_i^0) \\ \mathrm{Im}(x_i^1) \end{bmatrix} \tag{9.50}$$

where $i = 0, 1, \ldots, N_s^l - 1$ and $N_s^p = 2N_s^l$.

For transmission on four antenna ports $p \in \{0, 1, 2, 3\}$ and for the purpose of CSI reporting based on antenna ports $p \in \{0, 1, 2, 3\}$ or $p \in \{15, 16, 17, 18\}$, the precoding matrix $\mathbf{W}_i$ is defined based on Householder codebook[19] $\mathbf{W}_n = \mathbf{I}_4 - 2\mathbf{u}_n\mathbf{u}_n^H/\mathbf{u}_n^H\mathbf{u}_n$ where $\mathbf{I}_4$ is a 4×4 identity matrix and the vector $\mathbf{u}_n$ is defined in Table 9.18.

[19]In linear algebra, a Householder transformation is defined as a linear transformation that performs a reflection about a plane or hyper-plane containing the origin. The reflection about the hyper-plane can be defined by a unit vector $\mathbf{u}_n$ which is orthogonal to the hyper-plane. The reflection of a vector $\mathbf{x}$ about this hyper-plane is defined as $\mathbf{x} - 2\langle\mathbf{u}_n, \mathbf{x}\rangle\mathbf{u}_n = \mathbf{x} - 2\mathbf{u}_n(\mathbf{u}_n{}^H\mathbf{x})$ where $\mathbf{u}_n$ is given as a column unit vector. This is a linear transformation given by the Householder matrix $\mathbf{\Lambda} = \mathbf{I} - 2\mathbf{u}_n\mathbf{u}_n{}^H$ in which $\mathbf{I}$ denotes the identity matrix. The Householder matrix has the following properties [31,32]:

1. It is a Hermitian matrix $\mathbf{\Lambda} = \mathbf{\Lambda}^H$
2. It is a unitary matrix $\mathbf{\Lambda}^{-1} = \mathbf{\Lambda}^H$ hence it is involutary $\mathbf{\Lambda}^2 = \mathbf{I}$. An involutary function is a function $f(x)$ that is its own inverse function, i.e., $f(f(x)) = x, \forall x$.
3. The eigenvalues of the Householder matrix are ± 1.
4. The determinant of a Householder reflector is $\det(\mathbf{\Lambda}) = -1$ given that the determinant of a matrix is the product of its eigenvalues.

Table 9.18 Codebook for Spatial Multiplexing with and without CDD [3]

N_l	U	D_i	W_i
2	$\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 1\\ 1 & e^{-j2\pi/2}\end{bmatrix}$	$\begin{bmatrix}1 & 0\\ 0 & e^{-j2\pi i/2}\end{bmatrix}$	$N_{TX}=2$: $\mathbf{W}_i=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ $N_{TX}=4$: $\mathbf{W}_n=\mathbf{I}_4-2\mathbf{u}_n\mathbf{u}_n^H/\mathbf{u}_n^H\mathbf{u}_n$ $n=12,13,14,15$
3	$\frac{1}{\sqrt{3}}\begin{bmatrix}1 & 1 & 1\\ 1 & e^{-j2\pi/3} & e^{-j4\pi/3}\\ 1 & e^{-j4\pi/3} & e^{-j8\pi/3}\end{bmatrix}$	$\begin{bmatrix}1 & 0 & 0\\ 0 & e^{-j2\pi i/3} & 0\\ 0 & 0 & e^{-j4\pi i/3}\end{bmatrix}$	$\mathbf{W}_n=\mathbf{I}_4-2\mathbf{u}_n\mathbf{u}_n^H/\mathbf{u}_n^H\mathbf{u}_n$ $n=12,13,14,15$
4	$\frac{1}{2}\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & e^{-j2\pi/4} & e^{-j4\pi/4} & e^{-j6\pi/4}\\ 1 & e^{-j4\pi/4} & e^{-j8\pi/4} & e^{-j12\pi/4}\\ 1 & e^{-j6\pi/4} & e^{-j12\pi/4} & e^{-j18\pi/4}\end{bmatrix}$	$\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & e^{-j2\pi i/4} & 0 & 0\\ 0 & 0 & e^{-j4\pi i/4} & 0\\ 0 & 0 & 0 & e^{-j6\pi i/4}\end{bmatrix}$	$\mathbf{W}_n=\mathbf{I}_4-2\mathbf{u}_n\mathbf{u}_n^H/\mathbf{u}_n^H\mathbf{u}_n$ $n=12,13,14,15$

Precoding for spatial multiplexing using antenna ports with UE-specific reference signals is only applicable in conjunction with layer mapping, and supports up to eight antenna ports where the set of antenna ports is given by $p = 7, 8, \ldots, N_l + 6$. For transmission on $N_p = N_l$ antenna ports, the precoding operation is defined by $(y_i^7, y_i^8, \ldots, y_i^{N_l+6})^{\mathrm{T}} = (x_i^0, x_i^1, \ldots, x_i^{N_l-1})^{\mathrm{T}}$, where $i = 0, 1, \ldots, N_s^p - 1$ and $N_s^p = N_s^l$.

For each antenna port used for transmission of the physical channels, the block of complex-valued symbols $y_0^p, \ldots, y_{N_s^p-1}^p$ must be sequentially mapped (starting with y_0^p) to resource elements (k, l) whose coordinates satisfy the following criteria: (1) these resource elements are in the PRBs corresponding to the designated VRBs; (2) these resource elements are not used for transmission of PBCH, synchronization signals, CRSs, MBSFN reference signals, or UE-specific reference signals; (3) the UE assumes that these resource elements are not used for transmission of CSI-RSs and the DCI associated with the downlink transmission using the C-RNTI or semi-persistent C-RNTI; and (4) index l in the first slot in a subframe satisfies $l \geq l_{\mathrm{DataStart}}$, where $l_{\mathrm{DataStart}}$ is the starting OFDM symbol index for corresponding downlink transmission. Therefore, for each antenna port used for transmission of PDSCH, the block of complex-valued symbols y_i^p are mapped in sequential order to resource elements that are not occupied by the PCFICH, PHICH, PDCCH, PBCH or synchronization, and reference signals. The number of resource elements involved in this mapping is controlled by the number of resource blocks allocated to PDSCH. The symbols are mapped by increasing the sub-carrier index and mapping all available resource elements within allocated resource blocks for each OFDM symbol, as shown in Figure 9.87.

9.12.1 Physical control format indicator channel

The PCFICH carries information about the number of OFDM symbols used for transmission of PDCCHs in a subframe. In other words, the PCFICH carries a control format indicator (CFI) which indicates the number of OFDM symbols, which is typically 1, 2, or 3, except for system bandwidth less than 5 MHz, used for transmission of control channel information in each subframe. In principle, the UE can intuitively deduce the value of the CFI without using PCFICH, e.g., by multiple attempts to blindly decode the control channels assuming each possible number of symbols, but this would result in significant additional processing complexity. Three different CFI values are used in LTE/LTE-Advanced. In order to make the CFI sufficiently robust, each codeword is 32 bits long and is mapped to 16 resource elements using QPSK modulation. These 16 resource elements are organized in groups of four resource elements, known as REGs. The resource elements occupied by reference signals are not included within the REGs, which means that the total number of REGs in a given OFDM symbol depends on whether cell-specific reference signals are present. The concept of REGs (i.e., mapping in groups of four resource elements) is also used for PHICH and

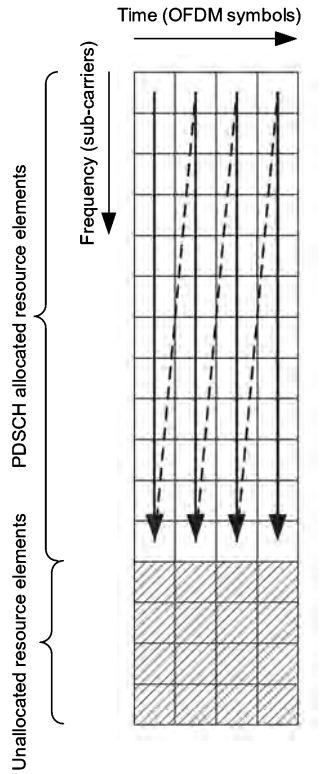


FIGURE 9.87

Resource element mapping [160].

PDCCH control channels. The PCFICH is transmitted on the same set of antenna ports as PBCH. The transmit diversity is applied if more than one antenna port is used. In order to achieve frequency diversity, the four REGs carrying the PCFICH are distributed across the frequency-domain. This is done according to a predefined pattern in the first OFDM symbol of each downlink subframe (Figure 9.88), so that the UEs can always locate the PCFICH information, which is a prerequisite for detection and decoding of the other control channels. To minimize the inter-cell interference, as well as the possibility of confusion with PCFICH information transmitted from neighboring cells, a cell-specific frequency offset is applied to the positions of the PCFICH resource elements. The offset value depends on the physical cell identity, which is derived from the primary and secondary synchronization signals. In addition, a cell-specific scrambling sequence is applied to the CFI codewords, so that the UE can preferentially receive PCFICH from the desired cell site.

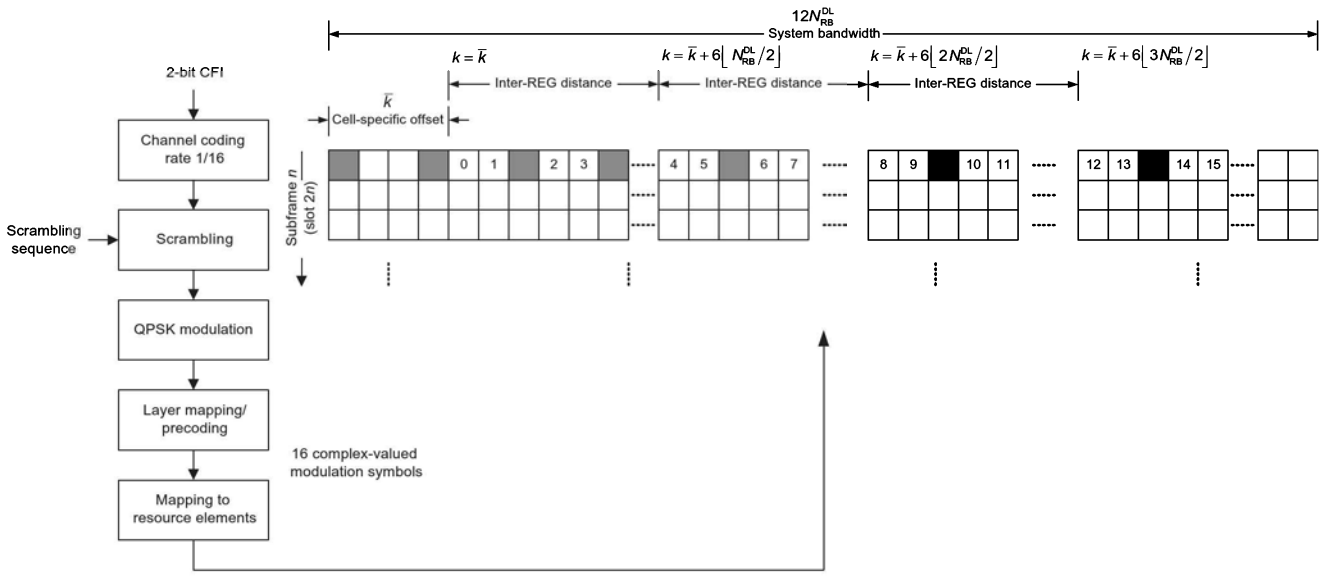


FIGURE 9.88

PCFICH physical layer processing [15,162].

As shown in Figure 9.88, the block of coded bits $(b_0, \ldots, b_{31})$ containing CFI information (i.e., 2 bits of information encoded at the rate 1/16) transmitted in one subframe are scrambled with a cell-specific sequence prior to modulation, resulting in a block of scrambled bits $(\tilde{b}_0, \ldots, \tilde{b}_{31})$, where $\tilde{b}_i = (b_i + c(i)) \bmod 2$. The scrambling sequence $c(i)$ defined in previous sections is initialized with $c_{\text{init}} = \left(\lfloor n_s/2 \rfloor + 1\right)\left(2N_{\text{ID}}^{\text{cell}} + 1\right)2^9 + N_{\text{ID}}^{\text{cell}}$ at the start of each subframe. The block of scrambled bits $(\tilde{b}_0, \ldots, \tilde{b}_{31})$ is modulated using QPSK modulation, resulting in a block of complex-valued modulation symbols $(d_0, \ldots, d_{15})$ which are mapped to spatial layers using a transmit diversity transmission mode with $N_s^0 = 16$, i.e., 16 complex-valued modulation symbols per layer and are precoded, resulting in a block of vectors $\mathbf{y}_i = (y_i^0, y_i^1, \ldots, y_i^{N_p-1})^{\text{T}}$, $i = 0, \ldots, 15$, where y_i^p represents the signal for antenna port p and $p = 0, \ldots, N_p - 1$. The number of antenna ports for cell-specific reference signals $N_p \in \{1, 2, 4\}$. The PCFICH is transmitted on the same set of antenna ports that are used for transmission of PBCH. The mapping to resource elements is defined in terms of quadruplets of complex-valued symbols. Let $Z_i^P = (y_{4i}^P, y_{4i+1}^P, y_{4i+2}^P, y_{4i+3}^P)$ denote symbol quadruplet i for antenna port p. For each of the antenna ports, symbol quadruplets are mapped in increasing order of index i to four resource-element groups in the first OFDM symbol in a downlink subframe (assuming 12 sub-carriers per resource block) in the following manner:

$$\begin{aligned}
&Z_0^P \quad \text{is mapped to the resource-element represented by} \quad k = \bar{k} \\
&Z_1^P \quad \text{is mapped to the resource-element represented by} \quad k = \bar{k} + 6\lfloor N_{\text{RB}}^{\text{DL}}/2 \rfloor \\
&Z_2^P \quad \text{is mapped to the resource-element represented by} \quad k = \bar{k} + 6\lfloor 2N_{\text{RB}}^{\text{DL}}/2 \rfloor \\
&Z_3^P \quad \text{is mapped to the resource-element represented by} \quad k = \bar{k} + 6\lfloor 3N_{\text{RB}}^{\text{DL}}/2 \rfloor
\end{aligned} \tag{9.51}$$

where the additions are modulo $12N_{\text{RB}}^{\text{DL}}$, $\bar{k} = 6(N_{\text{ID}}^{\text{cell}} \bmod 2N_{\text{RB}}^{\text{DL}})$ and $N_{\text{ID}}^{\text{cell}}$ is the physical-layer cell identity. The above procedure is illustrated in Figure 9.88.

9.12.2 Physical broadcast channel

The essential system information which allows the other channels in the cell to be configured and operated is carried by PBCH. Therefore, the reliability and coverage of the PBCH is crucial for the successful operation of LTE/LTE-Advanced systems. The system information is divided into two categories: (1) the master information block as given in Table 9.19, which consists of a limited number of the most frequently transmitted parameters essential for initial access to the cell which is carried in PBCH; and (2) the other system information blocks (SIBs) which are multiplexed with unicast data and are transmitted in PDSCH. The design requirements for PBCH are as follows:

- The UE must be able to detect the PBCH without prior knowledge of the system bandwidth.
- The PBCH must have negligible impact on the total physical layer overhead.
- The PBCH must be reliably received by the cell-edge UEs.
- The PBCH must be decodable with low latency and complexity to reduce the impact on the UE battery life.

Table 9.19 Contents of the Master Information Block [9]

Parameter	Size (Bits)	Description
Downlink channel bandwidth	3	Selects one of the configured downlink system bandwidths in terms of the number of resource blocks (i.e., $N_{RB}^{DL} = 6, 15, 25, 50, 75, 100$)
PHICH configuration: PHICH Duration	1	Selects between two options "normal" and "extended" where normal refers to one OFDM symbol and extended refers to three OFDM symbols
PHICH configuration: PHICH Resource	2	Selects one of the values of PHICH group scaling factor $N_g \in \{1/6, 1/2, 1, 2\}$
System frame number (SFN)	8	Most significant bits of the 10-bit system frame number. The two least significant bits of the SFN are deducted by the UE from the position of the PBCH (i.e., within 40 ms PBCH TTI, the first radio frame: 00, the second radio frame: 01, the third radio frame: 10, the last radio frame: 11)
Unused bits	10	Null bits for future extensions

The overall PBCH structure is shown in Figure 9.89. The detectability of the PBCH without the UE's prior knowledge of the system bandwidth is achieved by mapping the PBCH to the center of the band using 72 sub-carriers, which corresponds to the minimum system bandwidth of six resource blocks, regardless of the actual system bandwidth. The UE will have to identify the system center frequency from the synchronization signals as described in earlier sections. Minimizing the impact on total overhead is achieved by deliberately keeping the amount of information carried by the PBCH to a minimum, since achieving stringent coverage requirements for a large quantity of data would result in a high system overhead. The size of MIB it is 24 bits and it is transmitted every 40 ms. In order to achieve reliable reception, time diversity, FEC, and antenna diversity mechanisms are used to transmit PBCH. Time diversity is introduced by spreading the transmission of the PBCH over a period of 40 ms with three retransmissions at 10 ms intervals. This significantly reduces the likelihood of missing/mis-detection of MIB due to multipath fading, interference, and noise in the radio channel, even when the mobile terminal is moving at high speeds.

The FEC coding for PBCH utilizes a tail-biting convolutional coder, considering that the number of information bits to be encoded is small and turbo codes are typically not suitable for small payloads. The basic code rate is 1/3, after which a high repetition factor is applied to the systematic and parity bits such that each MIB is coded at a very low code rate (1/48 over a 40 ms period) to provide strong error protection. Antenna diversity may be utilized at the eNB and the UE. The UE performance requirements specified for LTE/LTE-Advanced assume that all UEs can achieve a level of decoding performance proportional to dual-antenna

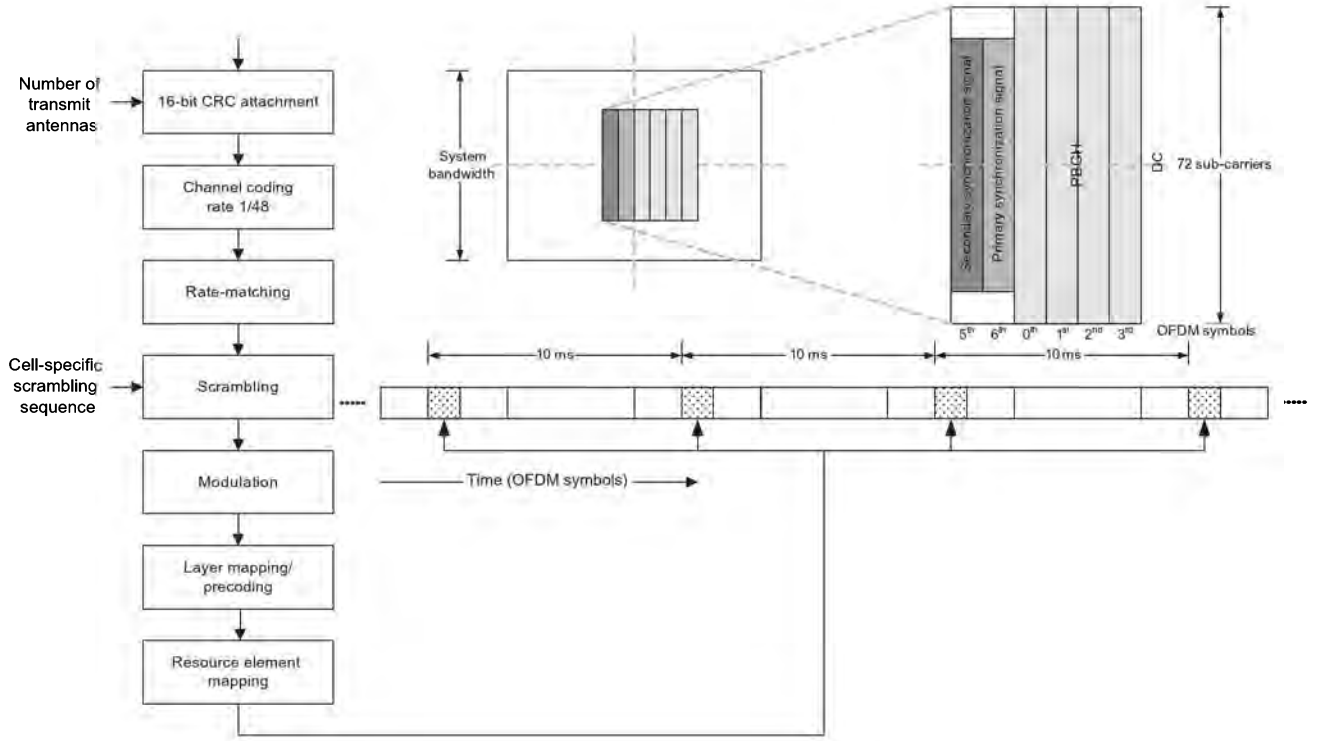


FIGURE 9.89

PBCH transmission and physical layer processing [162].

receive diversity, although it is understood that in low-frequency deployments, e.g., below 1 GHz, the advantage obtained from receive diversity is reduced due to the correspondingly higher correlation between the antennas. This enables wider or more reliable cell coverage to be achieved with fewer cell sites.

Transmit antenna diversity may be employed at the eNB to further improve coverage, depending on eNB implementation, eNB may transmit PBCH with two or four transmit antennas using an SFBC scheme. The set of resource elements designated to PBCH transmission is independent of the number of transmit antennas at the eNB, thereby any resource elements which may be used for cell-specific reference signals transmission are avoided by the PBCH, irrespective of the actual number of transmit antenna ports used by the eNB. The number of transmit antennas used by the eNB is blindly detected by the UE by decoding PBCH transmission corresponding to a different number of transmit antennas (i.e., one, two, or four). The detection of the number of transmit antennas is further facilitated by masking the CRC of each MIB with a unique masking sequence representing the number of transmit antenna ports. The 16-bit PBCH CRC masking sequence is $\langle 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0\rangle$,$\langle 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1\rangle$, or $\langle 0,1,0,1,0,1,0,1,0,1,0,1,0,1,0,1\rangle$ for one, two, or four transmit antennas, respectively.

Achieving low latency and minimal impact on UE battery life are facilitated through use of a low code rate with repetition wherein the set of coded bits are divided into four subsets, each of which is self-decodable. Each subset of the coded bits is transmitted in a different radio frame during the 40 ms transmission period, as shown in Figure 9.89. This means that if the SINR is sufficiently good to allow the UE to successfully decode the MIB in less than four radio frames, then the UE does not need to receive the other parts of the PBCH transmission in the remainder of the 40 ms period. On the other hand, if the SINR is low, the UE can take advantage of soft-combing of other parts of the MIB transmission until MIB is successfully decoded. The UE does not know in advance the timing of the 40 ms transmission interval for each MIB on the PBCH; therefore, this information is determined implicitly from the scrambling and bit positions, which are reinitialized every 40 ms. The UE can initially determine the 40 ms timing by performing four separate decodings of the PBCH using each of the four possible phases of the PBCH scrambling code and checking the CRC for each decoding.

When a UE initially attempts to select a cell by acquiring the PBCH, different approaches may be taken to perform the necessary blind decoding. A simple approach is to perform the decoding using soft combing of the PBCH instances over four radio frames, advancing a 40 ms sliding window one radio frame at a time until the sliding window aligns with the 40 ms period of the PBCH and the decoding succeeds. However, this would result in a 40−70 ms delay before the PBCH can be decoded. A faster approach would be to attempt to decode the PBCH from the first radio frame, which should be possible provided that the SINR is sufficiently high. If the decoding fails for all four possible scrambling sequences, the PBCH from the first frame could be soft-combined with the next

PBCH instances. It is evident that the latter approach may be much faster at the expense of slightly more complexity.

More specifically, the block of coded bits $(b_0, b_1 \ldots, b_{N_{\text{bit}}-1})$ are scrambled with a cell-specific scrambling sequence prior to modulation, resulting in a block of scrambled bits $(\tilde{b}_0, \tilde{b}_1, \ldots, \tilde{b}_{N_{\text{bit}}-1})$ in which $\tilde{b}_i = (b_i + c(i)) \bmod 2$ where the cell-specific scrambling sequence $c(i)$ was defined in the previous sections. The number of bits transmitted on PBCH $N_{\text{bit}} = 1920$ for a normal cyclic prefix and $N_{\text{bit}} = 1728$ for an extended cyclic prefix. The scrambling sequence is initialized with $c_{\text{init}} = N_{\text{ID}}^{\text{cell}}$ in each radio frame, provided that $n_f \bmod 4 = 0$.

The block of scrambled bits $(\tilde{b}_0, \tilde{b}_1, \ldots, \tilde{b}_{N_{\text{bit}}-1})$ is modulated using QPSK modulation, resulting in a block of complex-valued modulation symbols $(d_0, d_1, \ldots, d_{N_s-1})$. The block of modulation symbols $(d_0, d_1, \ldots, d_{N_s-1})$ is mapped to spatial layers with $N_s^0 = N_s$ and precoded, resulting in a block of vectors $\mathbf{y}_i = (y_i^0, y_i^1, \ldots, y_i^{N_p-1})^T$ where $i = 0, 1, \ldots, N_s - 1$ and y_i^p is the signal for antenna port $p = 0, 1, \ldots, N_p - 1$ and $N_p \in \{1, 2, 4\}$.

The block of complex-valued symbols $(y_0^p, y_1^p \ldots, y_{N_s-1}^p)$ for each antenna port is transmitted during four consecutive radio frames starting in the frame whose number satisfies $n_f \bmod 4 = 0$. The symbols are mapped sequentially starting with y_0^p to resource elements (k, l) that are not reserved for transmission of reference signals. The resource mapping is done in increasing order of sub-carrier index k (i.e., frequency first) and then index l (OFDM symbol number) in slot 1 of subframe 0 in each radio frame. The resource-element indices are given by $k = 6N_{\text{RB}}^{\text{DL}} - 36 + k'; k' = 0, 1, \ldots, 71$; and $l = 0, 1, \ldots, 3$ where resource elements reserved for reference signals are excluded. The mapping procedure assumes that cell-specific reference signals for antenna ports 0–3 are used irrespective of the actual antenna configuration [3].

9.12.3 Physical hybrid-ARQ indicator channel

The PHICH carries the HARQ ACK/NACK, which indicates whether the eNB has correctly received a transmission on PUSCH. The HARQ indicator is set to "0" for a positive acknowledgment and is set to "1" for a negative acknowledgment. This information is repeated in each of three BPSK symbols. Multiple PHICHs are mapped to the same set of resource elements. These constitute a PHICH group, where different PHICHs within the same PHICH group are separated through different complex-valued orthogonal Walsh sequences. Each PHICH is uniquely identified by a PHICH index, which indicates both the group and the sequence. The sequence length is four for the normal cyclic prefix (or two for the extended cyclic prefix). As the sequences are complex-valued, the number of PHICHs in a group, i.e., the number of UEs receiving their acknowledgments on the same set of downlink resource elements, can be up to twice the sequence length. A cell-specific scrambling sequence is applied. A 3× code repetition factor is applied for improved robustness, resulting in three instances of the orthogonal Walsh code being transmitted for each ACK or NACK. The error

rate on the PHICH is intended to be in the order of 0.01 for ACKs, and as low as 0.0001 for NACKs. The resulting PHICH construction, including repetition and orthogonal spreading, is shown in Figure 9.90.

The PHICH duration, in terms of the number of OFDM symbols used in the time-domain, is typically one or three OFDM symbols, which is configurable and the eNB broadcasts the configuration information via the PBCH (see Table 9.19). Since the PHICH cannot extend into the PDSCH transmission region, the duration configured for the PHICH puts a lower limit on the size of the control channel region at the start of each subframe. Each of the three instances of the orthogonal code of a PHICH transmission is mapped to a REG on one of the first three OFDM symbols of each subframe, so that each PHICH is partly transmitted on each of the available OFDM symbols. This mapping is illustrated in Figure 9.90 for an example PHICH duration. The PBCH also provides the number of PHICH groups that are configured in the cell, which enables the UEs to derive the resource elements in the control region where the PDCCHs are mapped.

In order to obviate the need for additional signaling to indicate which PHICH carries the ACK/NACK response for each PUSCH transmission, the PHICH index is implicitly associated with the index of the lowest uplink resource block used for the corresponding PUSCH transmission. This relationship is such that adjacent PUSCH resource blocks are associated with PHICHs in different PHICH groups, to enable some degree of load balancing. However, this mechanism alone is not sufficient to enable multiple UEs to be allocated in the same resource blocks for a PUSCH transmission, as occurs in the case of uplink MU-MIMO, in this case, different cyclic shifts of the uplink DM-RSs are configured for the different UEs which are allocated the same time–frequency PUSCH resources, and the same cyclic shift index is then used to shift the PHICH allocations in the downlink so that each UE receives its ACK/NACK on a different PHICH. The above procedures are illustrated in Figure 9.90 [13,14].

Each PHICH group consists of 12 complex-valued symbols which require three REGs and can carry up to four (extended cyclic prefix) or eight (normal cyclic prefix) ACK/NACKs, which are identified within a PHICH group by an orthogonal code. Each PHICH group is identified by a PHICH group number and one of the eight or four orthogonal spreading codes. The starting location of the first REG of the PHICH group is derived from the physical cell identity. The other two groups are equally spaced from the first group.

The PHICH may be configured with a normal or extended duration. In a normal duration, all three REGs of a PHICH group are mapped to the first OFDM symbol. In an extended duration, the three REGs are mapped to three different OFDM symbols, as shown in Figure 9.91. The extended duration PHICH mapping provides time diversity. The duration of the PHICH is signaled via RRC and carried in the PHICH configuration information element.

As mentioned earlier, multiple PHICHs are mapped to the same set of resource elements to create a PHICH group, where PHICHs within the same

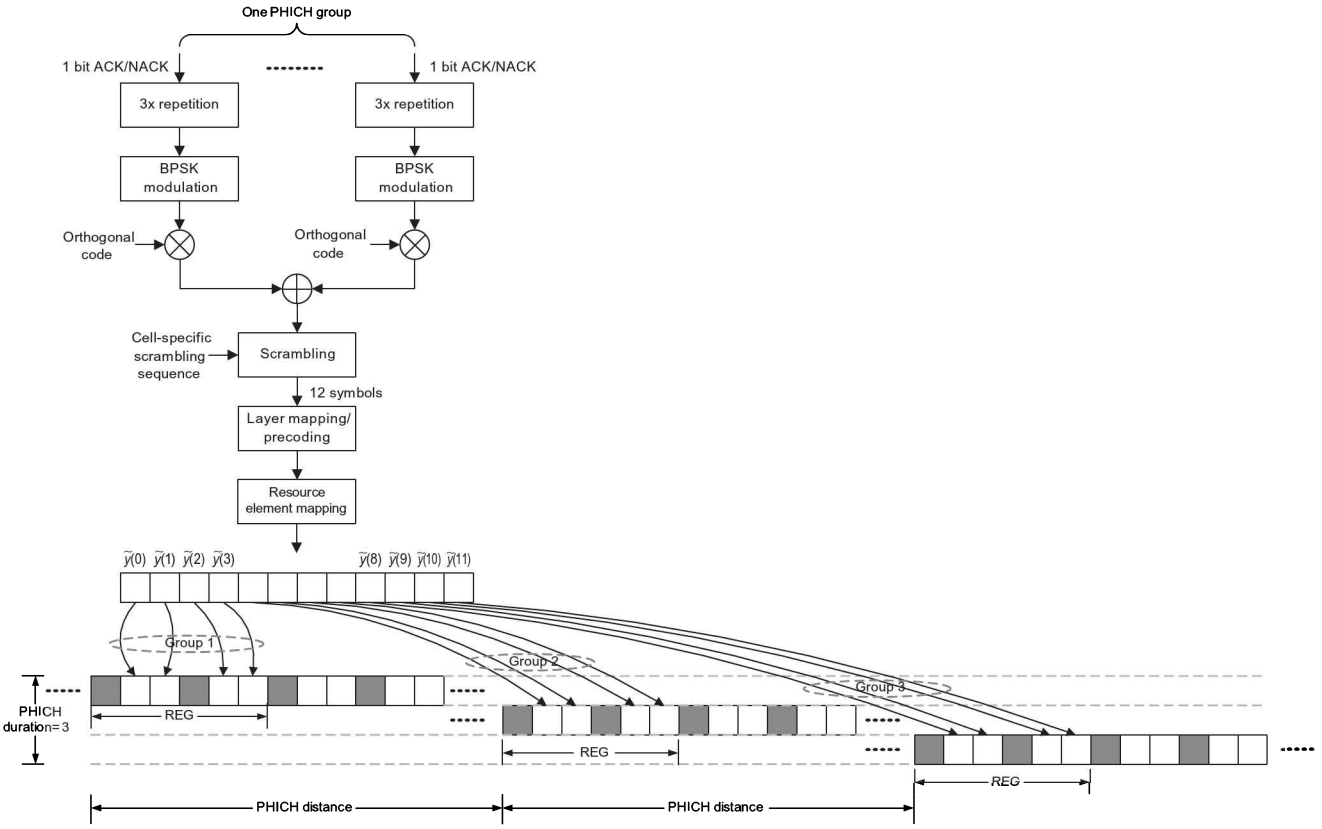


FIGURE 9.90

PHICH physical layer processing and mapping in time and frequency for an FDD system with normal cyclic prefix [15,162].

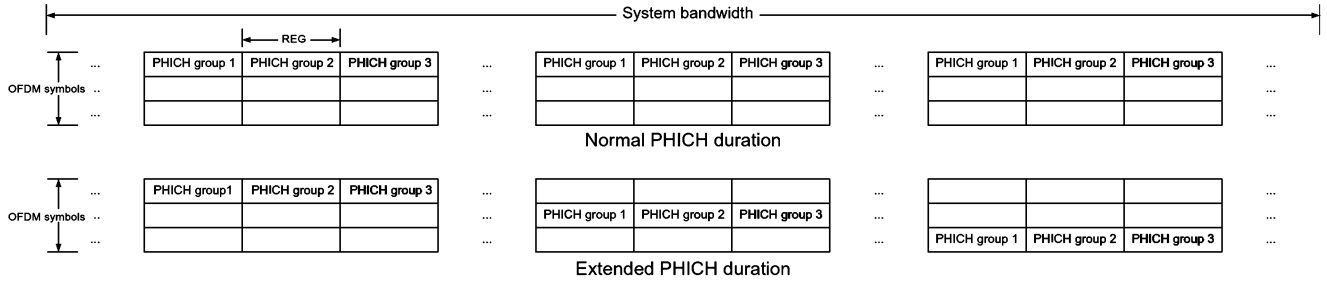


FIGURE 9.91

Illustration of normal and extended PHICH mapping.

PHICH group are separated by different orthogonal sequences. A PHICH resource is identified by the index pair $(n_{\text{PHICH}}^{\text{group}}, n_{\text{PHICH}}^{\text{sequence}})$, where $n_{\text{PHICH}}^{\text{group}}$ denotes the PHICH group number and $n_{\text{PHICH}}^{\text{sequence}}$ is the orthogonal sequence index within the group. For frame structure type 1, the number of PHICH groups $n_{\text{PHICH}}^{\text{group}}$ is constant in all subframes and is given by $N_{\text{PHICH}}^{\text{group}} = \lceil N_g(N_{\text{RB}}^{\text{DL}}/8) \rceil$ for a normal cyclic prefix and $N_{\text{PHICH}}^{\text{group}} = 2\lceil N_g(N_{\text{RB}}^{\text{DL}}/8) \rceil$ for an extended cyclic prefix, in which $N_g \in \{1/6, 1/2, 1, 2\}$ is a parameter configured by higher layers and transmitted in the form of two bits in the PBCH (see Table 9.19). The group index $n_{\text{PHICH}}^{\text{group}}$ assumes integer values between 0 and $N_{\text{PHICH}}^{\text{group}} - 1$. For frame structure type 2, the number of PHICH groups may vary from one downlink subframe to another and is given by $m_i N_{\text{PHICH}}^{\text{group}}$, where the value of parameter $m_i = 0, 1, 2$ depends on the TDD subframe configuration, e.g., $m_i = 1$ for the second subframe in TDD frame configuration 1. The group index $n_{\text{PHICH}}^{\text{group}}$ in a downlink subframe with non-zero PHICH resources assumes integer values between 0 and $m_i N_{\text{PHICH}}^{\text{group}} - 1$.

One may ask at this point how the UEs can find the PHICH corresponding to their earlier PUSCH transmissions. The UE searches for the PHICH in the fourth subframe $(n + 4)$ after an uplink transport block is transmitted in uplink subframe n. The PHICH group and sequence indices are calculated based on the lowest PRB index allocated for PUSCH transmission and the cyclic shift for the PUSCH DM-RS, both of which are associated with the uplink transmission and are provided to the UE in DCI format 0 or DCI format 4. The PHICH resource (N_g) is received in the MIB or in the PHICH configuration information element via dedicated RRC signaling. The ith UE selects its PHICH group and orthogonal sequence index $(n_{\text{PHICH}}^{\text{group}}(i), n_{\text{PHICH}}^{\text{sequence}}(i))$ according to the following rules in an FDD system [13]:

$$\begin{aligned} 1 - n_{\text{PHICH}}^{\text{group}}(i) &= (I_{\text{lowest}} + n_{\text{DM-RS}}) \bmod N_{\text{PHICH}}^{\text{group}} \\ 2 - n_{\text{PHICH}}^{\text{sequence}}(i) &= \left(\lfloor I_{\text{lowest}}/N_{\text{PHICH}}^{\text{group}} \rfloor + n_{\text{DM-RS}} \right) \bmod (2N_{\text{SF}}^{\text{PHICH}}) \end{aligned} \tag{9.52}$$

where $n_{\text{DM-RS}}$ denotes the PUSCH DM-RS cyclic shift signaled in DCI format 0 or 4, and I_{lowest} is the lowest PRB index allocated to the first slot of the PUSCH transmission. In the case of a TDD system with subframe configuration 0 and PUSCH transmission in subframes 4 or 9, the PHICH group index calculated in Eq. (9.51) must be incremented by $N_{\text{PHICH}}^{\text{group}}$ [13]. For a Rel-10 UE using two transport blocks within a subframe, the parameter I_{lowest} is incremented by one when calculating the PHICH group and sequence number for the second transport block.

As shown in Figure 9.90, the block of bits $(b_0, b_1, \ldots, b_{N_{\text{bit}}-1})$ transmitted on one PHICH in one subframe are modulated using BPSK modulation, resulting in a block of complex-valued modulation symbols $(z_0, z_1, \ldots, z_{N_s-1})$ where $N_s = N_{\text{bit}}$ given the type of modulation. The block of modulation symbols $(z_0, z_1, \ldots, z_{N_s-1})$ are symbol-by-symbol multiplied by an orthogonal

scrambling sequence, resulting in a sequence of scrambled modulation symbols $(d_0, d_1 ..., d_{N_{SF}^{PHICH} N_s - 1})$ where:

$$d_i = \underbrace{w(i \bmod N_{SF}^{PHICH})}_{\substack{\text{Orthogonal sequence symbol} \\ \text{with index } n_{PHICH}^{sequence}}} \quad \underbrace{(1-2c(i))}_{\substack{\text{Cell-specific scrambling} \\ \text{sequence symbol}}} \quad \underbrace{z(\lfloor i/N_{SF}^{PHICH} \rfloor)}_{\substack{\text{Modulated HARQ} \\ \text{indicator symbol}}} \quad i = 0, 1, \ldots, N_{SF}^{PHICH} N_s - 1 \tag{9.53}$$

The value of the spreading factor is $N_{SF}^{PHICH} = 4$ or 2 for a normal or extended cyclic prefix type, respectively [3]. The scrambling sequence $c(i)$ is a cell-specific scrambling sequence which was defined in the previous sections and is initialized by $c_{init} = (\lfloor n_s/2 \rfloor + 1)(2N_{ID}^{cell} + 1)2^9 + N_{ID}^{cell}$ at the beginning of each subframe. The orthogonal spreading sequence $(w_0, w_1, \ldots, w_{N_{SF}^{PHICH}-1})$ is defined in Table 9.20 where the sequence index $n_{PHICH}^{sequence}$ corresponds to the PHICH number within the PHICH group.

Since REGs contain four resource elements, each able to carry one modulation symbol, the blocks of scrambled symbols are aligned to create blocks of four symbols (Figure 9.92). In the case of a normal cyclic prefix, each of the original complex-valued modulated symbol is represented by four scrambled symbols, thus no alignment is required. In the case of an extended cyclic prefix, each of the original complex-valued modulated symbols is represented by two scrambled symbols. To create blocks of four symbols, zeros are added before or after blocks of two scrambled symbols, depending on whether the PHICH index is odd or even. This allows two groups to be combined during the resource mapping stage and mapped to one REG. Groups of four symbols are formed as illustrated in Figure 9.92.

The block of symbols $(d_0, d_1 \ldots, d_{N_{SF}^{PHICH} N_s - 1})$ are first aligned with REG size, and are then mapped to spatial layers and precoded, yielding a block of vectors $\mathbf{y}_i = (y_i^0, y_i^1, \ldots, y_i^{N_p - 1})^T$ where $i = 0, 1, \ldots, N_{SF}^{PHICH} N_s - 1$, and y_i^p represents the signal of antenna port $p = 0, 1, \ldots, N_p - 1$. The number of antenna ports with CRSs is chosen as $N_p \in \{1, 2, 4\}$. The layer mapping and precoding operation depend on the cyclic prefix type and the number of antenna ports used for transmission of the PHICH. The PHICH is transmitted on the same set of antenna ports that are used for transmission of the PBCH.

The number of OFDM symbols used to carry PHICH is configurable by PHICH duration. The PHICH duration is either normal or extended. A normal PHICH duration allows PHICH to present only in the first OFDM symbol of the subframe. In general, an extended PHICH duration allows PHICH to be located in the first three OFDM symbols of the subframe, with some exceptions. Under those exceptions, PHICH may present in the first two OFDM symbols of the subframe. Those exceptions are as follows: (1) within subframe 1 and 6 when frame structure type 2 (TDD) is used; and (2) within MBSFN subframes.

Table 9.20 Orthogonal Spreading Sequence for PHICH Processing [3]

Sequence Index	$(w_0, w_1, \ldots, w_{N_{SF}^{PHICH}-1})$	
$n_{PHICH}^{sequence}$	Normal Cyclic Prefix $N_{SF}^{PHICH} = 4$	Extended Cyclic Prefix $N_{SF}^{PHICH} = 2$
0	[+1 +1 +1 +1]	[+1 +1]
1	[+1 −1 +1 −1]	[+1 −1]
2	[+1 +1 −1 −1]	[+j +j]
3	[+1 −1 −1 +1]	[+j −j]
4	[+j +j +j +j]	–
5	[+j −j +j −j]	–
6	[+j +j −j −j]	–
7	[+j −j −j +j]	–

Given that CFI carried by PCFICH configures the number of OFDM symbols that are used for the control region, or alternatively the size of the PDSCH region, the extended PHICH duration must not exceed the size of the control region. For example, when using an extended PHICH in subframe zero of frame structure type 1 (FDD) with 10 MHz bandwidth, the first three OFDM symbols will contain the PHICH. Therefore, the CFI must be set to 3 so that the PDSCH is not mapped to OFDM symbols 0, 1, or 2, and thus does not collide with the PHICH. As shown in Figure 9.92, the corresponding elements of each PHICH sequence are summed to create the sequence for each PHICH group where the sum is over all PHICHs in the PHICH group, and $y_i^{(p)}(n)$ represents the symbol sequence from the ith PHICH in the PHICH group. As shown in Figure 9.92, the PHICH groups are mapped to PHICH mapping units, i.e., $\bar{y}_n^p = \sum_i y_i^p(n)$ where $y_i^p(n)$ is the nth element within the ith PHICH group on antenna port p. For a normal cyclic prefix, the mapping of PHICH group m to PHICH mapping unit m' is defined as $\tilde{y}_{m'}^p(n) = \bar{y}_m^p(n)$, where $m' = m = 0, 1, ..., m_i N_{PHICH}^{group} - 1$ for frame structure type 1, and $m' = m = 0, 1, \ldots, m_i N_{PHICH}^{group} - 1$ for frame structure type 2. In other words, the PHICHs are mapped to REGs using PHICH mapping units $\tilde{y}_{m'}^p(n)$, where m' is the index of the mapping unit. For a normal cyclic prefix, each PHICH group is mapped to a PHICH mapping unit. In the case of an extended cyclic prefix, two PHICH groups are mapped to one PHICH mapping unit. Due to the location of the padding zeros added during resource group alignment, when two consecutive groups are added, the zeros of one group overlap with the data of the other group.

Each mapping unit contains 12 symbols. In order to map these 12 symbols to REGs, the mapping units are split into three groups of four symbols, or

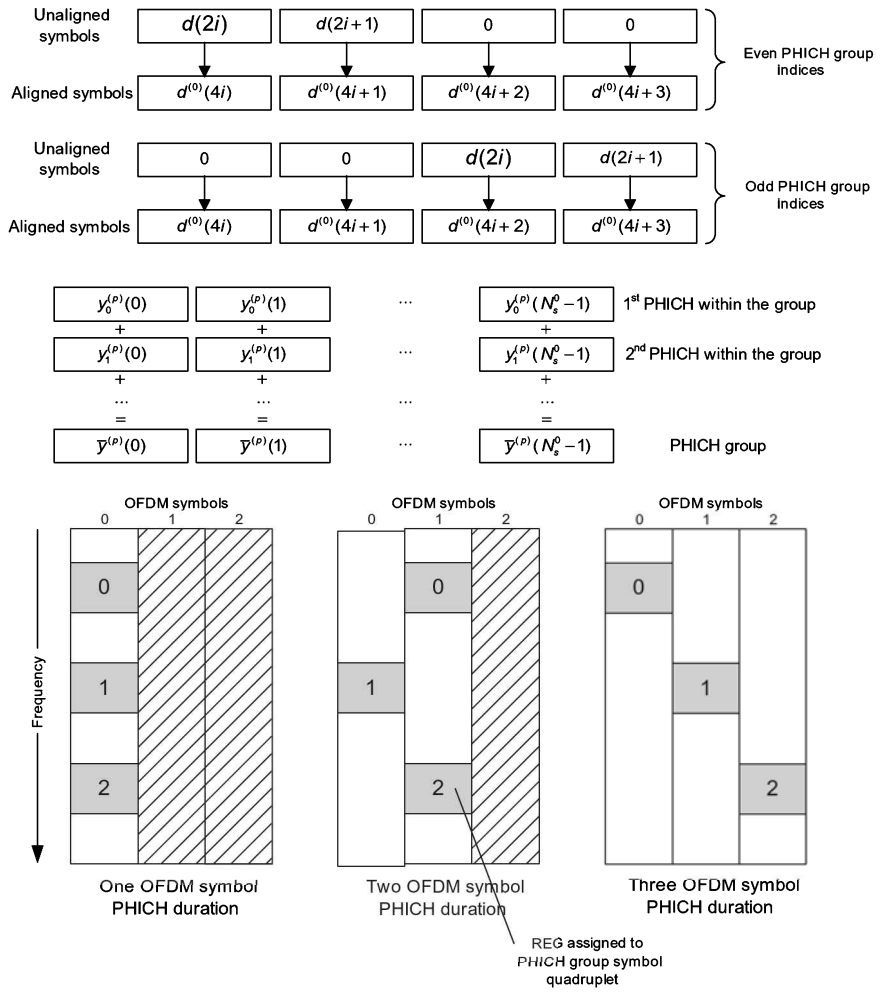


FIGURE 9.92

Illustration of PHICH symbol alignment, sequence combining, and mapping to REGs for normal and extended PHICH [160].

quadruplets. Each of the three symbol quadruplets $z^p(i) = (\tilde{y}^p(4i), \tilde{y}^p(4i+1), \tilde{y}^p(4i+2)$, and $\tilde{y}^p(4i+3))$ is mapped to REG (k'_i, l'_i) such that PHICH is spread over all available OFDM symbols and resource blocks. The OFDM symbol index l'_i is set so that adjacent quadruplets are spread among the available OFDM symbols as illustrated in Figure 9.90. The sub-carrier index k'_i of the REG is based on the cell identity $N_{\mathrm{ID}}^{\mathrm{cell}}$ and is chosen to spread the three symbol quadruplets over the entire system bandwidth.

9.12.4 Physical downlink control channel

The physical downlink control channel is used to carry the DCI, which contains downlink/uplink scheduling decisions and power-control commands. More specifically, the DCI includes downlink scheduling assignments, PDSCH resource allocation, transport format, HARQ related information, and control information related to spatial multiplexing. The downlink scheduling assignment further includes a command for power control of the PUCCH that is used for transmission of HARQ ACK/NACK in response to downlink scheduling assignments. The DCI further includes uplink scheduling grants, PUSCH resource indication, transport format, and other HARQ related information. The uplink scheduling grant also consists of a command for power control of PUSCH. Furthermore, it carries power-control commands for a set of terminals as complementary to the commands included in the scheduling assignments/grants. It must be noted that different types of control information correspond to different DCI message sizes. For example, supporting spatial multiplexing with non-contiguous allocation of resource blocks in the frequency-domain requires larger scheduling message in comparison to an uplink grant, only allowing frequency-contiguous allocations. The DCI is subsequently categorized to different formats where a format corresponds to a certain message size, content, and usage. The LTE/LTE-Advanced DCI formats are summarized in Table 9.21. The actual message size depends on, among other factors, the cell bandwidth, thus for larger bandwidths, a larger number of bits is required to indicate the resource-block allocation. The number of CRSs in the cell and whether cross-carrier scheduling is configured will also affect the size of most DCI formats. Hence, a given DCI format may have different sizes depending on the overall configuration of the cell.

9.12.4.1 PDCCH physical layer processing

As mentioned earlier, each PDCCH carries a message known as the DCI, which includes resource assignments and other control signaling information for a UE or a group of UEs. Each PDCCH is transmitted using one or more control channel elements (CCEs), where each CCE corresponds to nine REGs (Figure 9.93). The number of REGs that are not assigned to a PCFICH or PHICH is denoted by N_{REG}. The CCEs available in the system are numbered from 0 to $N_{CCE} - 1$, where $N_{CCE} = \lfloor N_{REG}/9 \rfloor$. The PDCCH supports multiple formats as listed in Table 9.21. A

Table 9.21 PDCCH Formats [3]

PDCCH Format	Number of CCEs (Aggregation Level)	Number of Resource Element Groups (REGs)	Number Of PDCCH Bits
0	1	9	72
1	2	18	144
2	4	36	288
3	8	72	576

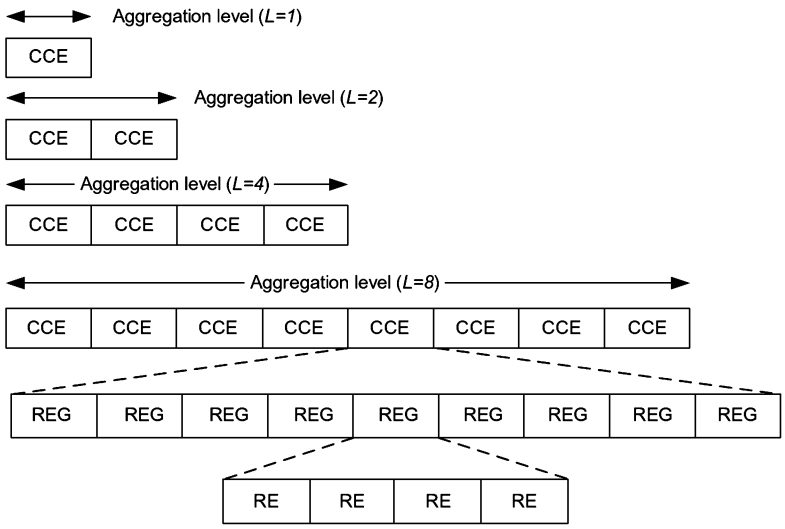


FIGURE 9.93

Illustration of various aggregation levels and CCE structure [162].

PDCCH consisting of L consecutive CCEs (aggregation level) may only start on a CCE that satisfies (n_{CCE} mod $L = 0$), where the parameter n_{CCE} denotes the CCE index. Multiple PDCCHs can be transmitted in a subframe. Four QPSK symbols are mapped to each REG. Four PDCCH formats are supported, as listed in Table 9.21.

A cyclic redundancy check is used for error detection in the DCI messages. The entire PDCCH payload is used to calculate a set of CRC parity bits. The PDCCH payload is divided by a cyclic generator polynomial to generate 16 parity bits. These parity bits are then appended to the end of the PDCCH payload. As multiple PDCCHs relevant to different UEs can be present in one subframe, the CRC is also used to specify to which UE a PDCCH is addressed. This is done by scrambling the CRC parity bits with the corresponding RNTI of the UE. The scrambled CRC is obtained by performing a bit-wise XOR operation between the 16-bit calculated PDCCH CRC and the 16-bit RNTI. Different RNTI can be used to scramble the CRC. Some examples include:

- A UE unique identifier in the RRC_CONNECTED mode (C-RNTI).
- A paging indication identifier (P-RNTI), if the PDSCH contains paging information.
- A system information identifier (SI-RNTI), if PDSCH contains system information.

When encoding a DCI format 0 payload, containing the UE UL-SCH resource allocation, and if the UE transmit-antenna selection is configured, the RNTI-scrambled CRC undergoes a bit-wise XOR operation with an antenna selection

mask. This informs the UE which transmit antenna must be used for uplink transmission. The antenna selection mask for UE antenna port 0 is (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) and for UE antenna port 1 is (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1).

As shown in Figure 9.94, the DCI message with the attached CRC goes through tail-biting convolutional coding. LTE/LTE-Advanced uses a rate 1/3 tail-biting convolutional encoder with a constraint length $K = 7$. This means that one

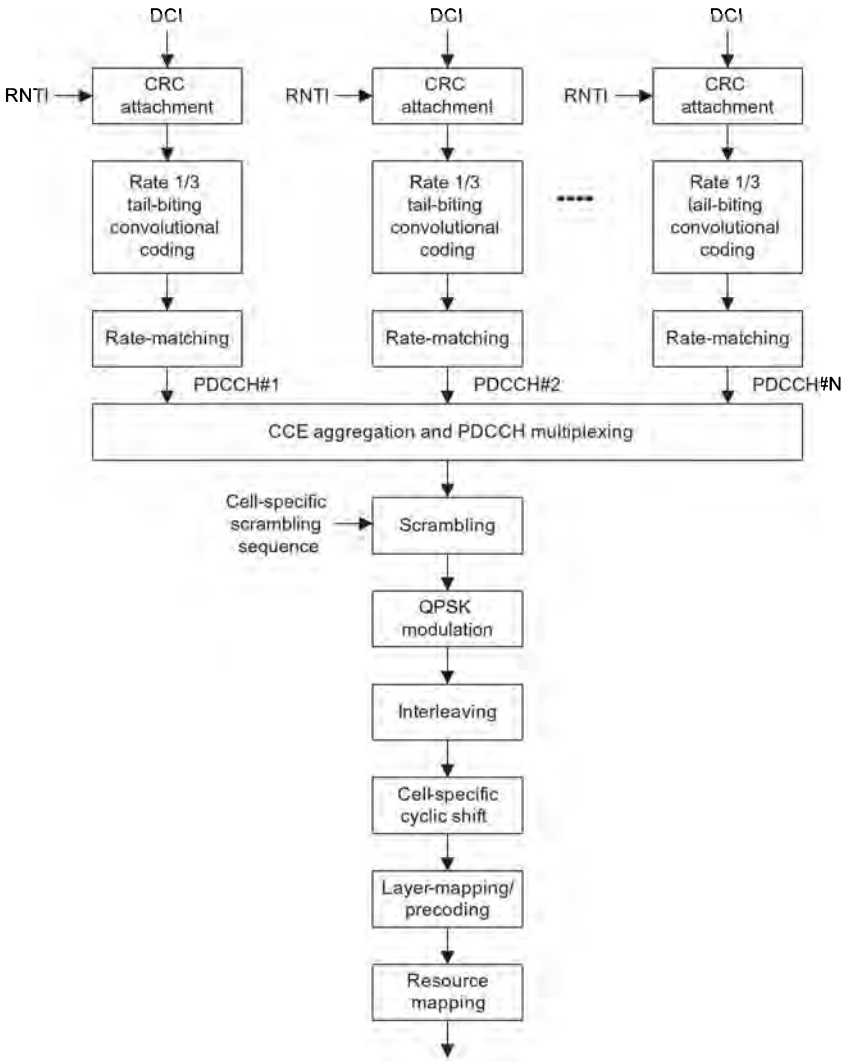


FIGURE 9.94

PDCCH physical layer processing stages [15].

in three bits of the output contain useful information while the other two add parity bits. Each output stream of the encoder is obtained by convolving the input with the impulse response of the encoder. The impulse responses are called the generator sequences of the encoder. Note that a conventional convolutional encoder initializes its internal shift register to the all-zero state and also ensures that the encoder finishes in the all-zero state, by padding the input sequence with K zeros at the end. Knowing the start and end states (i.e., all zeros) simplifies the design of the decoder (typically an implementation of the Viterbi algorithm). A tail-biting convolutional coder initializes its internal shift register to the last K bits of the current input block, rather than to the all-zero state. This means the starting and the ending states are the same, without the need to zero pad the input block. As the overhead of terminating the encoder has been eliminated, the output block contains fewer bits than a conventional convolutional coder. The drawback is that the decoder becomes more complicated, since the initial state is unknown; however, it does know the start and end states are the same. The rate-matching block creates an output bitstream with the desired code rate. The three bit streams from the convolutional encoder are interleaved, followed by bit collection to create a virtual circular buffer.

The three sub-block interleavers used in the rate-matching block are identical. Interleaving is a technique to reduce the impact of burst errors on a signal to ensure that consecutive data bits will not be corrupted. The bit collection stage creates a virtual circular buffer by concatenating the three interleaved encoded bit streams. The bits are then selected and pruned from the circular buffer to create an output sequence length which meets the desired code rate. The coded DCI messages for each control channel are multiplexed and scrambled before undergoing QPSK modulation, layer mapping, and precoding. The blocks of coded bits for each control channel are multiplexed in order to create a block of data as shown in Figure 9.96, where N_{bit}^{i} is the number of bits in the ith control channel and N_{PDCCH} is the number of control channels.

As shown in Figure 9.94, the block of coded and rate-matched bits $(b_0^i, b_1^i, \ldots, b_{N_{\text{bit}}^i-1}^i)$ on each of the control channels awaiting transmission in a subframe are multiplexed, resulting in a block of multiplexed bits $(b_0^0, b_1^0, \ldots, b_{N_{\text{bit}}^0-1}^0, b_0^1, b_1^1, \ldots, b_{N_{\text{bit}}^1-1}^1, \ldots, b_0^{N_{\text{PDCCH}}-1}, b_1^{N_{\text{PDCCH}}-1}, \ldots, b_{N_{\text{bit}}^{N_{\text{PDCCH}}-1}-1}^{N_{\text{PDCCH}}-1})$, where N_{bit}^i is the number of bits in the ith DCI, and N_{PDCCH} denotes the number of PDCCHs transmitted in one subframe. The block of multiplexed bits $(b_0^0, b_1^0, \ldots, b_{N_{\text{bit}}^0-1}^0, b_0^1, b_1^1, \ldots, b_{N_{\text{bit}}^1-1}^1, \ldots, b_0^{N_{\text{PDCCH}}-1}, b_1^{N_{\text{PDCCH}}-1}, \ldots, b_{N_{\text{bit}}^{N_{\text{PDCCH}}-1}-1}^{N_{\text{PDCCH}}-1})$ is scrambled with a cell-specific sequence prior to modulation (i.e., the multiplexed block of bits undergoes a bit-wise XOR operation with a cell-specific scrambling sequence), resulting in a block of scrambled bits $(\tilde{b}_0, \tilde{b}_1, \ldots, \tilde{b}_{N_{\text{total}}-1})$, in which $\tilde{b}_i = b_i \oplus c(i)$. The scrambling function provides more robustness through inter-cell interference randomization. When a UE descrambles the received bit stream with a known cell-specific scrambling sequence, interference from other cells will be descrambled

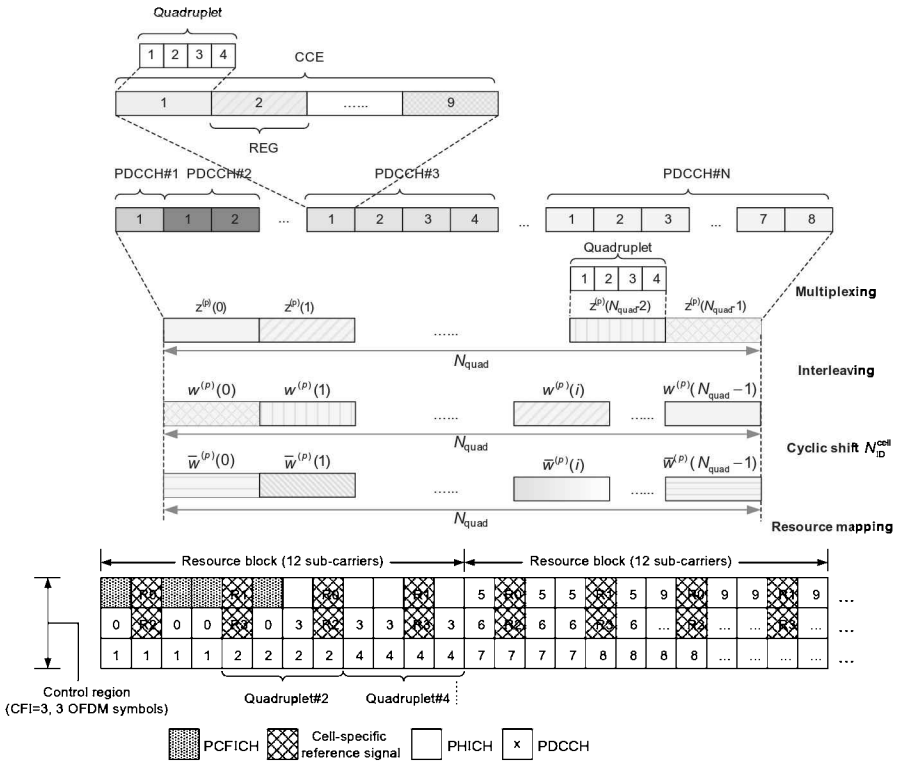


FIGURE 9.95

Example breakdown of CCE processing [162].

incorrectly and appear as uncorrelated noise. The cell-specific scrambling sequence $c(i)$ was defined in the previous sections and is initialized with $c_{\text{init}} = \lfloor n_s/2 \rfloor 2^9 + N_{\text{ID}}^{\text{cell}}$ at the start of each subframe. It must be noted that the CCE number n corresponds to bit array $(b_{72n}, b_{72n+1}, \ldots, b_{72n+71})$. If necessary, null elements are inserted in the block of bits prior to scrambling to ensure that the PDCCHs start at the permissible CCE positions, and to ensure that the length $N_{\text{total}} = 8N_{\text{REG}} \geq \sum_{i=0}^{N_{\text{PDCCH}}-1} N_{\text{bit}}^i$ of the scrambled block of bits matches the amount of REGs not assigned to the PCFICH or PHICH.

The block of scrambled bits $(\tilde{b}_0, \tilde{b}_1, \ldots, \tilde{b}_{N_{\text{total}}-1})$ is modulated using QPSK modulation, resulting in a block of complex-valued modulation symbols $(d_0, d_1, \ldots, d_{N_s-1})$. The block of modulated symbols $(d_0, d_1, \ldots, d_{N_s-1})$ are further mapped to $N_s^0 = N_s$ spatial layers and precoded, yielding the block of precoded vectors $\mathbf{y}_i = (y_i^0, y_i^1, \ldots, y_i^{N_p-1})^T$ with $i = 0, \ldots, N_s - 1$. The latter block of symbols are mapped to resources on the antenna ports used for transmission of PDCCH, where y_i^p represents the signal for antenna port p. Note that the PDCCH is transmitted on the same set of antenna ports that were used for PBCH transmission.

The mapping to resource elements are defined by operations on quadruplets of complex-valued symbols, as shown in Figure 9.95. Let $z_i^p = (y_{4i}^p, y_{4i+1}^p, y_{4i+2}^p, y_{4i+3}^p)$ denote the ith symbol quadruplet for antenna port p. The block of quadruplets $(z_0^p, z_1^p, \ldots, z_{N_{\text{quad}}-1}^p)$ where $N_{\text{quad}} = N_s/4$ are permuted (interleaved), resulting in a permuted quadruplet vector $(w_0^p, w_1^p, \ldots, w_{N_{\text{quad}}-1}^p)$. The permutation is performed similar to sub-block interleaving with the following exceptions: (1) the input and output to the interleaver are defined by symbol quadruplets instead of bits; and (2) interleaving is performed on symbol quadruplets instead of bits by substituting the bits with symbol quadruplets. The null elements at the output of the interleaver must be removed when constructing the permuted vector $(w_0^p, w_1^p, \ldots, w_{N_{\text{quad}}-1}^p)$. The block of interleaved quadruplets $(w_0^p, w_1^p, ..., w_{N_{quad}-1}^p)$ is then cyclically shifted, resulting in $(\overline{w}_0^p, \overline{w}_1^p, ..., \overline{w}_{N_{\text{quad}}-1}^p)$ where $\overline{w}_i^p = w_{(i+N_{\text{ID}}^{\text{cell}}) \bmod N_{\text{quad}}}^p$. The mapping of the block of quadruplets $(\overline{w}_0^p, \overline{w}_1^p, \ldots, \overline{w}_{N_{\text{quad}}-1}^p)$ is defined in terms of resource-element groups. Each symbol-quadruplet is sequentially mapped to an unallocated REG starting with REG $(k' = 0, l' = 0)$ such that symbol quadruplet $\overline{w}_{m'}^p$ is mapped to REG m'. The REG symbol index l' is then incremented until all the REGs at sub-carrier index $k' = 0$ have been allocated, provided that $l' < N_{\text{CFI}}$. The REG sub-carrier index k' is then incremented and the process is repeated until $k' < 12N_{\text{RB}}^{\text{DL}}$. This mapping continues until all symbol quadruplets have been allocated to REGs. The mapping is illustrated in Figure 9.95 for an example resource grid. Four transmit antenna ports and a control region size of three OFDM symbols are used to create the grid. The mapping is performed along symbols first, and then along frequency until the entire symbol quadruplets are mapped, in order to cause no significant power imbalance among the PDCCH symbols. If symbol-quadruplets were only mapped to the first OFDM symbol among OFDM symbols dedicated to the control region on a particular subframe, it would negatively impact the PA operation and performance.

Figure 9.97 shows an example mapping of CCEs in a subframe. In this example, REGs $(k' = 0, l' = 0)$ and $(k' = 6, l' = 0)$ were already allocated to PCFICH and PHICH, respectively, thus no symbol quadruplets are assigned to them. The symbol-quadruplets are first mapped to REG $(k' = 0, l' = 1)$, followed by REG $(k' = 0, l' = 2)$ and so on. The symbol-quadruplets are mapped in ascending order of index l'. As there are no REGs with $k' = 0$, the next REG allocated is REG $(k' = 4, l' = 2)$, as this REG has the lowest value of index l' that has not already been allocated. This process is repeated until all symbol quadruplets are assigned

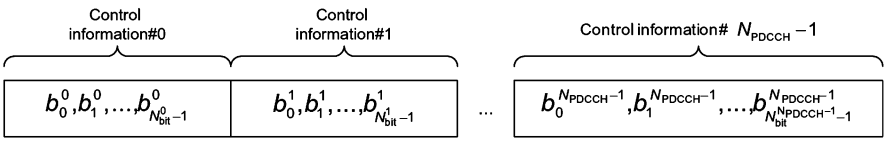


FIGURE 9.96

Illustration of PDCCH multiplexing function [160].

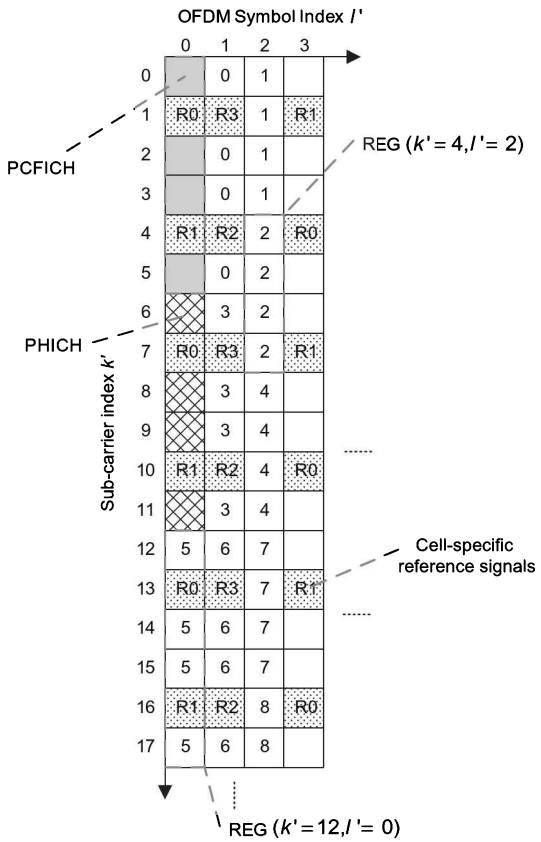


FIGURE 9.97

Example mapping of CCEs in the control region of a subframe [160].

to the available REGs [160]. Therefore, the CCE mapping process can be summarized as follows. Each symbol-quadruplet is sequentially mapped to an unallocated REG starting with REG $(k' = 0, l' = 0)$. This is described as symbol quadruplet $\overline{w}^{p}_{m'}$ being mapped to REG m'. The REG symbol index k' is then incremented until all the REGs at sub-carrier index $l' = 0$ have been allocated. The REG sub-carrier index k' is then incremented and the process is repeated. This mapping continues until all symbol quadruplets have been allocated REGs. The mapping is shown in Figure 9.97 for an example resource grid. Four transmit antenna ports and control region size of three OFDM symbols are used to create the grid.

9.12.4.2 DCI formats

One PDCCH carries one DCI message with one of the formats given in Table 9.21. Given that multiple terminals can be scheduled simultaneously on both downlink and uplink, there must be a possibility to transmit multiple

scheduling messages within each subframe. Each scheduling message is transmitted on a separate PDCCH, and consequently there are typically multiple simultaneous PDCCH transmissions within each cell. Furthermore, to support different radio channel conditions, link adaptation can be used, where the transmission format of the PDCCH (i.e., coding and modulation scheme, multi-antenna configuration) is adapted to cope with the severity of the radio channel conditions. For carrier aggregation, scheduling assignments/grants are transmitted individually per component carrier. Depending on the purpose of DCI message, different DCI formats are defined as follows [4]:

- Format 0: Transmission of uplink grants for PUSCH.
- Format 1: Transmission of PDSCH assignments with a single codeword.
- Format 1A: DCI Format 1A is used for compact signaling of resource assignments in single-codeword PDSCH transmissions using any PDSCH transmission mode. It is also used to allocate a dedicated preamble signature to a UE for the purpose of contention-free random access.
- Format 1B: DCI Format 1B is used for compact signaling of resource assignments in PDSCH transmissions using rank-1 closed-loop precoding. The information transmitted is the same as DCI Format 1A with the addition of an indicator for the precoding vector applied to PDSCH transmission.
- Format 1C: Transmission of PDSCH assignment using a very compact format. When format 1C is used, the PDSCH transmission is constrained to use QPSK modulation.
- Format 1D: Same as Format 1B with additional information for power offset (i.e., instead of one of the precoding vector indicator bits, there is a single bit to indicate whether a power offset is applied to the data symbols to indicate whether the transmission power is shared between two UEs) and PDSCH assignments for multi-user MIMO.
- Format 2: PDSCH assignments for closed-loop MIMO operation.
- Format 2A: PDSCH assignments for open-loop MIMO operation.
- Format 2B: PDSCH assignments for dual-layer beamforming.
- Format 2C: DCI Format 2C was introduced in Rel-10 and is used for signaling resource assignments for PDSCH transmissions using closed-loop single-user or multi-user MIMO with up to eight layers.
- Format 2D: DCI Format 2D was introduced in Rel-11 and is used for CoMP operations in conjunction with transmission mode 10.
- Format 3: Transmission of transmit power control (TPC) commands for multiple users and for PUCCH and PUSCH with 2-bit power adjustments.
- Format 3A: Transmission of TPC commands for multiple users and for PUCCH and PUSCH with 1-bit power adjustments.
- Format 4: DCI Format 4 was introduced in Rel-10 and is used for signaling resource grants for PUSCH when the UE is configured in PUSCH transmission mode 2 for uplink single-user MIMO. The information transmitted is similar to DCI Format 0, with the addition of MCS and new data indicator for a second transport block (TB) and precoding data.

Multiple UEs can be scheduled in one subframe; therefore, multiple DCI messages can be sent using multiple PDCCHs. The description, usage, and detailed information fields of various DCI formats for LTE/LTE-Advanced are given in Table 9.22.

As shown in Table 9.22, the number of bits required for resource allocation/assignment depends on the system bandwidth; therefore, the size of that information field also varies with the system bandwidth. The payload size for each DCI format including information bits and CRC are summarized in Table 9.23 for the E-UTRA supported system bandwidths. In addition, padding bits are added (if necessary) to ensure that DCI Formats 0 and 1A are of the same size, even in the case of different uplink and downlink bandwidths, in order to avoid additional complexity at the UE receiver. Furthermore, the padding bits may be added to ensure that DCI Formats 3 and 3A have the same size as DCI Formats 0 and 1A, to circumvent additional complexity at the UE receiver; to avoid potential ambiguity in identifying the correct PDCCH location; to ensure that DCI Format 1 has a different size than that of DCI Formats 0/1A, so that these formats can be easily distinguished at the UE receiver; to make sure that Format 4 has a different size from DCI Formats 1/2/2A/2B/2C, so that this format can be easily distinguished. As given in Table 9.22, in LTE Rel-10, some optional features were added that may be configured in some of the DCI formats; those bit counts are not included in Table 9.23. The DCI formats 1B/1D/2/2A/4 payload size calculations are based on two antenna ports [13]. The new DCI format 2D which has been added in Rel-11 includes the following information fields: Carrier indicator (0 or 3 bits); Resource allocation header for resource allocation type 0 or type 1 (1 bit), note that if downlink bandwidth is less than or equal to 10 resource blocks, there is no resource allocation header and resource allocation type 0 is assumed; Resource block assignment type 0/type 1 (variable number of bits depending on the system bandwidth); TPC command for PUCCH (2 bits); Downlink Assignment Index (TDD only) for uplink–downlink subframe configurations 1–6 (2 bits); HARQ process number (3 bits for FDD), (4 bits for TDD); Antenna port(s), scrambling identity and number of layers (3 bits); SRS request (0 or 1 bit) for TDD only; Modulation and coding scheme (5 bits), New data indicator (1 bit); Redundancy version (2 bits) for transport blocks 1 and 2; PDSCH resource element Mapping and Quasi Collocation Indicator (2 bits); HARQ-ACK resource offset (note that this field is present when this DCI format is carried by ePDCCH and is not present when this format is carried by PDCCH) (2 bits). The latter 2 bits are set to 0 when this format is carried by ePDCCH on a secondary cell. If both transport blocks are enabled; transport block 1 is mapped to codeword 0 and transport block 2 is mapped to codeword 1. In case one of the transport blocks is disabled; the transport block to codeword mapping is further specified [4].

9.12.4.3 Common and UE-specific search spaces and blind decoding

The PDCCH region consists of a number of CCEs which can be allocated to a PDCCH. The PDCCHs can be flexibly mapped to CCEs based on certain criteria.

Table 9.22 Descriptions, Usage, and Information Fields of Various DCI Formats [4,13]

DCI Format	Usage	Release	Information Fields	Size (Bits)	Description
Format 0	Scheduling of PUSCH in one uplink cell	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		10	Format 0/format 1A flag	1	Flag to differentiate between Format 0 and Format 1A.
		8	Frequency hopping flag	1	PUSCH frequency hopping flag for resource allocation type 0. This bit is used as the MSB of the resource block allocation in resource allocation type 1.
		8	Resource block allocation	7–13 variable	Resource block assignment/allocation. (In resource allocation type 0: 5, 7, 9, 11, 12, or 13 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth). (In resource allocation type 1: 6, 8, 10, 12, 13, or 14 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		8	MCS and redundancy version	5	Modulation, coding scheme and redundancy version.
		8	New data indicator (NDI)	1	New data indicator.
		8	Transmit power control (TPC)	2	PUSCH TPC command.
		8	Cyclic shift for uplink DM-RS	3	Cyclic shift for uplink DM-RS (different look up tables are used for LTE Rel-8/9 and LTE Rel-10 UEs based on this value.
		10	Cyclic shift for uplink DM-RS and OCC index		

(*Continued*)

Table 9.22 (Continued)

DCI Format	Usage	Release	Information Fields	Size (Bits)	Description
		8	Aperiodic CQI request	1	Instructs the UE to provide a periodic CQI report.
		10	CSI request	1 or 2	Instructs the UE to send a CSI report, where one-bit case refers to a single cell and two-bit case applies to a UE which is configured with more than one downlink cell.
		10	SRS request	0 or 1	Instructs the UE to transmit an SRS for uplink channel measurements.
		10	Resource allocation type	1	Signals the type of uplink resource allocation.
		8	Uplink index (TDD only)	2	For TDD configuration 0, this field is the uplink index. For TDD configuration 1–6, this field is the downlink assignment index.
		8	Downlink assignment index (DAI) (TDD only)		
Format 1	Transmission of PDSCH assignments with a single codeword	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		8	Resource allocation header	0 or 1	Resource allocation header selects between resource allocation type 0 or type 1 (only exists if downlink bandwidth is greater than 10 RBs).
		8	Resource block allocation	6–25 variable	Resource block assignment/allocation (6, 8, 13, 17, 19, or 25 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		8	MCS	5	Modulation and coding scheme.

		8	HARQ process number	3 (FDD) 4 (TDD)	HARQ process number.
		8	New data indicator	1	HARQ new data indicator.
		8	Redundancy version	2	HARQ redundancy version.
		8	TPC command for PUCCH	2	PUCCH TPC command.
		8	Downlink assignment index (TDD only)	2	For TDD configuration 0, this field is not used. For TDD configurations 1–6, this field is interpreted as the DAI.
Format 1A	Compact signaling of resource assignments for single-codeword PDSCH transmissions for any PDSCH transmission mode	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		8	Format 0/format 1A flag	1	Flag to differentiate between Format 0 and Format 1A.
		8	Localized/ distributed VRB assignment flag	1	Selects between 0 (localized) or 1 (distributed) resource blocks.
		8	Resource block allocation	5–13 variable	Resource block assignment/allocation (5, 7, 9, 11, 12, or 13 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		8	MCS	5	Modulation and coding scheme.
		8	HARQ process number	3 (FDD) 4 (TDD)	HARQ process number.
		8	New data indicator	1	HARQ new data indicator.

(*Continued*)

Table 9.22 (Continued)

DCI Format	Usage	Release	Information Fields	Size (Bits)	Description
		8	Redundancy version	2	HARQ redundancy version.
		8	TPC command for PUCCH	2	PUCCH TPC command.
		8	Downlink assignment index (TDD only)	2	For TDD configuration 0, this field is not used. For TDD configurations 1–6, this field is interpreted as DAI.
		10	SRS request	0 or 1	Instructs the UE to transmit SRS for uplink channel measurements.
Format 1B	Compact signaling of resource assignments for PDSCH transmissions using closed-loop precoding with rank-1 transmission	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		8	Localized/ distributed VRB assignment flag	1	Selects between 0 (localized) or 1 (distributed) resource blocks.
		8	Resource block allocation	5–13 variable	Resource block assignment/allocation (5, 7, 9, 11, 12, or 13 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		8	MCS	5	Modulation and coding scheme.
		8	HARQ process number	3 (FDD) 4 (TDD)	HARQ process number.
		8	New data indicator	1	HARQ new data indicator.
		8	Redundancy version	2	HARQ redundancy version.

		8	TPC command for PUCCH	2	PUCCH TPC command.
		8	Transmitted PMI (TPMI)	2 (2 antenna ports) 4 (4 antenna ports)	Downlink PMI information, i.e., information about the codebook used for precoding PDSCH (codebook-based closed-loop MIMO).
		8	PMI confirmation for precoding	1	Indicates whether PDSCH has been precoded using the codebook indicated by TPMI or by the PMI sent by the UE.
		8	Downlink assignment index (TDD only)	2	For TDD configuration 0, this field is not used. For TDD configuration 1–6, this field is the downlink assignment index.
Format 1C	Transmission of PDSCH assignment using a very compact format	8	Gap value	0 or 1	DCI format 1C always signals resource allocation type 2 using a distributed mapping. Smaller bandwidths always use “gap1” and larger bandwidths greater than 10 MHz can select between “gap1” and “gap2.”
		8	Resource block allocation	3–9 variable	Resource block assignment/allocation (3, 5, 7, 7, 8, or 9 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		8	Transport block size	5	DCI format 1C does not use the general TBS table, rather it uses a smaller TBS table.
Format 1D	Same as format 1B with additional information for power offset (i.e., instead of one of the precoding vector indicator	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of “0” implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		8	Localized/ distributed VRB assignment flag	1	Selects between 0 (localized) or 1 (distributed) resource blocks.
		8		5–13 variable	

(*Continued*)

Table 9.22 (Continued)

DCI Format	Usage	Release	Information Fields	Size (Bits)	Description
	bits, there is a single bit to indicate whether a power offset is applied to the data symbols to indicate whether the transmission power is shared between two UEs) and PDSCH assignments for multi-user MIMO		Resource block allocation		Resource block assignment/allocation (5, 7, 9, 11, 12, or 13 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		8	MCS	5	Modulation and coding scheme.
		8	HARQ process number	3 (FDD) 4 (TDD)	HARQ process number.
		8	New data indicator	1	HARQ new data indicator.
		8	Redundancy version	2	HARQ redundancy version.
		8	TPC command for PUCCH	2	PUCCH TPC command.
		8	Transmitted PMI (TPMI)	2 (2 antenna ports) 4 (4 antenna ports)	Downlink PMI information, i.e., information about the codebook used for precoding PDSCH (codebook-based closed-loop MIMO).
		8	Downlink power offset	1	Downlink power offset is used in MU-MIMO to inform the UE of its downlink transmit power relative to single-user transmit power.
		8	Downlink assignment index (TDD only)	2	For TDD configuration 0, this field is not used. For TDD configuration 1–6, this field is the downlink assignment index.
Format 2	PDSCH assignments for closed-loop	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.

MIMO operation	8	Resource allocation header	0 or 1	Resource allocation header selects between resource allocation type 0 or type 1 (only exists if downlink bandwidth is greater than 10 RBs).
	8	Resource block allocation	6–25 variable	Resource block assignment/allocation (6, 8, 13, 17, 19, or 25 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
	8	TPC command for PUCCH	2	PUCCH TPC command.
	8	HARQ process number	3 (FDD) 4 (TDD)	HARQ process number.
	8	Transport block to codeword swap flag	1	Defines the relationship between transport blocks 1 and 2 relative to codewords 0 and 1. If set to “0,” codeword 0 is generated from transport block 1 and codeword 1 is generated from transport block 2.
	8	MCS1	5	Modulation and coding scheme for transport block 1.
	8	New data indicator 1	1	HARQ new data indicator for transport block 1.
	8	Redundancy version 1	2	HARQ redundancy version for transport block 1.
	8	MCS2	5	Modulation and coding scheme for transport block 2.
	8	New data indicator 2	1	HARQ new data indicator for transport block 2.
	8	Redundancy version 2	2	HARQ redundancy version for transport block 2.
	8	Precoding information	3 (2-antenna ports) 6 (4-antenna ports)	Precoding information.
	8	Downlink assignment index (TDD only)	2	For TDD configuration 0, this field is not used. For TDD configuration 1–6, this field is the downlink assignment index.

(*Continued*)

Table 9.22 (Continued)

DCI Format	Usage	Release	Information Fields	Size (Bits)	Description
Format 2A	PDSCH assignments for open-loop MIMO operation	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		8	Resource allocation header	0 or 1	Resource allocation header selects between resource allocation type 0 or type 1 (only exists if downlink bandwidth is greater than 10 RBs).
		8	Resource block allocation	6–25 variable	Resource block assignment/allocation (6, 8, 13, 17, 19, or 25 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		8	TPC command for PUCCH	2	PUCCH TPC command.
		8	HARQ process number	3 (FDD) 4 (TDD)	HARQ process number.
		8	Transport block to codeword swap flag	1	Defines the relationship between transport blocks 1 and 2 relative to codewords 0 and 1. If set to "0," codeword 0 is generated from transport block 1 and codeword 1 is generated from transport block 2.
		8	MCS1	5	Modulation and coding scheme for transport block 1.
		8	New data indicator 1	1	HARQ new data indicator for transport block 1.
		8	Redundancy version 1	2	HARQ redundancy version for transport block 1.
		8	MCS2	5	Modulation and coding scheme for transport block 2.
		8	New data indicator 2	1	HARQ new data indicator for transport block 2.

		8	Redundancy version 2	2	HARQ redundancy version for transport block 2.
		8	Precoding information	0 (2 antenna ports) 2 (4 antenna ports)	Precoding information.
		8	Downlink assignment index (TDD only)	2	For TDD configuration 0, this field is not used. For TDD configuration 1–6, this field is the downlink assignment index.
Format 2B	PDSCH assignments for dual-layer beamforming	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		10	SRS request (TDD only)	0 or 1	Instructs the UE to transmit SRS for uplink channel measurements.
		9	Resource allocation header	0 or 1	Resource allocation header selects between resource allocation type 0 or type 1 (only exists if downlink bandwidth is greater than 10 RBs).
		9	Resource block allocation	6–25 variable	Resource block assignment/allocation (6, 8, 13, 17, 19, or 25 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		9	TPC command for PUCCH	2	PUCCH TPC command.
		9	HARQ process number	3 (FDD) 4 (TDD)	HARQ process number.
		9	Scrambling ID	1	Scrambling identity provides support for MU-MIMO by allowing separate users to reuse the same antenna port with different SCIDs.
		9	MCS1	5	Modulation and coding scheme for transport block 1.

(*Continued*)

Table 9.22 (Continued)

DCI Format	Usage	Release	Information Fields	Size (Bits)	Description
		9	New data indicator 1	1	HARQ new data indicator for transport block 1.
		9	Redundancy version 1	2	HARQ redundancy version for transport block 1.
		9	MCS2	5	Modulation and coding scheme for transport block 2.
		9	New data indicator 2	1	HARQ new data indicator for transport block 2.
		9	Redundancy version 2	2	HARQ redundancy version for transport block 2.
		9	Downlink assignment index (TDD only)	2	For TDD configuration 0, this field is not used. For TDD configuration 1–6, this field is the downlink assignment index.
Format 2C	Signaling resource assignments for PDSCH for closed-loop single-user or multi-user MIMO with up to 8 layers	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		10	SRS request (TDD only)	0 or 1	Instructs the UE to transmit SRS for uplink channel measurements.
		10	Resource allocation header	0 or 1	Resource allocation header selects between resource allocation type 0 or type 1 (only exists if downlink bandwidth is greater than 10 RBs).
		10	Resource block allocation	6–25 variable	Resource block assignment/allocation (6, 8, 13, 17, 19, or 25 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		10	TPC command for PUCCH	2	PUCCH TPC command.

		10	HARQ process number	3 (FDD) 4 (TDD)	HARQ process number.
		10	Antenna ports, scrambling ID and layers	3	Provides a pointer to a look-up table where the UE can find the number of layers, antenna ports, and the SCID. Scrambling identity provides support for MU-MIMO by allowing separate users to reuse the same antenna port with different SCIDs.
		10	MCS1	5	Modulation and coding scheme for transport block 1.
		10	New data indicator 1	1	HARQ new data indicator for transport block 1.
		10	Redundancy version 1	2	HARQ redundancy version for transport block 1.
		10	MCS2	5	Modulation and coding scheme for transport block 2.
		10	New data indicator 2	1	HARQ new data indicator for transport block 2.
		10	Redundancy version 2	2	HARQ redundancy version for transport block 2.
		10	Downlink assignment index (TDD only)	2	For TDD configuration 0, this field is not used. For TDD configuration 1−6, this field is the downlink assignment index.
Format 3	Transmit power control (TPC) commands for multiple users and for PUCCH and PUSCH with 2-bit power adjustment	8	TPC commands	20−28 variable	TPC commands for PUCCH and PUSCH are distinguished based the RNTI that is used to scramble CRC (20, 22, 24, 26, 26, or 28 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).

(*Continued*)

Table 9.22 (Continued)

DCI Format	Usage	Release	Information Fields	Size (Bits)	Description
Format 3A	Transmit power control (TPC) commands for multiple users and for PUCCH and PUSCH with one-bit power adjustment	8	TPC commands	21–28 variable	TPC commands for PUCCH and PUSCH are distinguished based the RNTI that is used to scramble CRC (21, 22, 25, 27, 27, or 28 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
Format 4	Signaling resource grants for PUSCH when the UE is configured in PUSCH transmission mode 2 for uplink single-user MIMO	10	Carrier indicator	0 or 3	Identifies the RF carrier to be used as the serving cell. A value of "0" implies the primary cell is used as the serving cell. This field only applies when carrier aggregation is enabled.
		10	Resource block allocation	6–25 variable	Resource block assignment/allocation (6, 7, 10, 12, 13, or 14 bits corresponding to 1.4, 3, 5, 10, 15, or 20 MHz system bandwidth).
		10	TPC Command for Scheduled PUSCH	2	Interpretation of this command depends on whether absolute or accumulated power control is used.
		10	Cyclic shift for DM-RS and OCC		Provides a pointer to a look-up table at the UE for generation of DM-RS and OCC.
		10	Downlink Assignment Index (TDD Only)	2	For TDD configuration 0, this field is not used. For TDD configuration 1–6, this field is the downlink assignment index.

10	CSI request	1 or 2	Instructs the UE to send a CSI report. 2 bits are used when the UE is configured with more than one downlink cell; otherwise, 1 bit is used.
10	SRS request	2	Instructs the UE to transmit SRS for uplink channel measurement and selects between SRS configurations
10	Resource allocation type	1	Selects the type of uplink resource allocation (type 0 or type 1).
10	Precoding information and number of layers	3 (2 Antenna Ports) 6 (4 Antenna Ports)	Provides a pointer to a look-up table at the UE to obtain the number of layers and transmitted PMI (TPMI).
10	MCS1	5	Modulation and coding scheme for transport block 1.
10	New data indicator 1	1	HARQ new data indicator for transport block 1.
10	MCS2	5	Modulation and coding scheme for transport block 2.
10	New data indicator 2	1	HARQ new data indicator for transport block 2.

Table 9.23 DCI Formats' Payload Sizes Including 16-bit CRC and Padding Bits [13,14]

DCI Format	FDD System Payload Size (Bits)						TDD System Payload Size (Bits)					
	1.4 MHz	3 MHz	5 MHz	10 MHz	15 MHz	20 MHz	1.4 MHz	3 MHz	5 MHz	10 MHz	15 MHz	20 MHz
Format 0	37	38	41	43	43	44	39	41	43	45	46	47
Format 1	35	39	43	47	49	55	38	43	46	50	52	58
Format 1A	37	38	41	43	43	44	39	41	43	45	46	47
Format 1B	38	41	43	44	45	46	41	43	45	47	49	49
Format 1C	24	26	28	29	30	31	24	26	28	29	30	31
Format 1D	38	41	43	44	45	46	41	43	45	47	49	49
Format 2	47	50	55	59	61	67	50	53	58	62	64	70
Format 2A	44	47	52	57	58	64	47	50	55	59	61	67
Format 2B	44	47	52	57	58	64	49	51	57	61	62	68
Format 2C	46	49	54	58	61	66	50	53	58	62	64	70
Format 3	37	38	41	43	43	44	37	38	41	43	43	44
Format 3A	37	38	41	43	43	44	37	38	41	43	43	44
Format 4	46	48	50	52	53	54	48	49	52	54	55	56

The common and UE-specific PDCCHs are mapped to CCEs differently where each type has a specific set of search spaces associated with it. Each search space consists of a group of consecutive CCEs which can be allocated to a PDCCH, known as PDCCH candidate. The CCE aggregation level is given by the PDCCH format and determines the number of PDCCH candidates in a search space. The number of candidates and size of the search space for each aggregation level is given in Table 9.24.

The UE is only informed of the number of OFDM symbols which construct the control region of a subframe, and is not provided with the location of its corresponding PDCCH(s). The UE finds its PDCCH by monitoring a set of PDCCH candidates in every subframe. This process is referred to as blind decoding. The UE unmasks each control candidate's CRC using its C-RNTI. If no CRC error is detected, the UE considers that it as a successful decoding attempt and reads the control information within the candidate PDCCH. The serving eNB determines a PDCCH format to be transmitted to the UE, creates an appropriate DCI and attaches a CRC. The CRC is then masked with an RNTI corresponding to a specific user or users. If the PDCCH is meant for a specific UE, the CRC will be masked with a unique UE identifier, e.g., a cell-RNTI (C-RNTI). If the PDCCH contains paging information, the CRC will be masked with a paging indication identifier, i.e., paging-RNTI (P-RNTI). If the PDCCH contains system information, a system information identifier, i.e., a system information-RNTI (SI-RNTI), will be used to mask the CRC.

With the possibilities of different RNTIs, PDCCH candidates, DCI and PDCCH formats, a significant number of attempts may be required to successfully decode the PDCCH. To overcome this complexity, the UE first tries to blindly decode the first CCE in the control channel candidate set of a subframe. If the blind decoding fails, the UE tries to blindly decode the first 2, 4, then 8 CCEs sequentially, where the starting location is fixed for a common search space and is given by a hash function for a UE-specific search space. The UE first tries to decode PDCCHs in the common search space before trying in the UE-specific search space. When searching the common search space, it iterates for only two

Table 9.24 PDCCH Candidates Monitored by a UE [5]

Search Space $S_k^{(L)}$				
Type	Aggregation Level L	Size (Number of CCEs)	Number of PDCCH Candidates N_L	DCI Format
UE-specific	1	6	6	0, 1/1A/1B/1D, 2/2A/2B/2C
	2	12	6	
	4	8	2	
	8	16	2	
Common	4	16	4	1, 1A, 1C, 3, 3A
	8	16	2	

aggregation levels, i.e., 4 and 8, and tries to decode all PDCCH candidates for all possible common search space DCI formats (see Table 9.24). The UE-specific search is carried out on four aggregation levels, i.e., 1, 2, 4, and 8. If no CRC error is detected during a decoding attempt, the UE considers it a successful decoding and reads the decoded DCI message.

A perfect PDCCH scheduler would be able to continuously allocate resources to each UE that is chosen by the downlink and uplink schedulers up to the point where all resources for the PDCCH, PDSCH, and PUSCH simultaneously become exhausted. However, because of the mechanism used for allocating PDCCH resources, it is possible that we can no longer allocate PDCCH resources to a UE even if such resources are available. In such cases, the corresponding UE is said to be blocked, while the eNB is still able to serve subsequent UEs. As more UEs are blocked, the performance of the scheduler decreases because such UEs cannot be served. Hence, the blocking probability, together with the utilization efficiency of bandwidth and power resources, are suitable metrics to measure the performance of the system. We assume that the number of PDCCH resources (i.e., CCEs) and total PDCCH power are known for each subframe. Multiple CCEs (1, 2, 4, or 8) can be aggregated to achieve different coding rates. The aggregation level required for a particular UE depends on the channel conditions of the UE, and the size of the scheduling grant that must be used for the UE. Each UE is allocated a candidate set of PDCCH channels for each possible aggregation level. The UE must then conduct blind decoding of each of these possibilities for each of the feasible message formats to determine if it was allocated PDSCH or PUSCH resources. The starting point for the candidate set for each aggregation level varies for each UE in each subframe. However, the new starting point is always known by the eNB and the UE. For each scheduled UE, the eNB must determine a suitable aggregation level, and then must use an unallocated channel from the corresponding candidate set to signal the location and format of the transmitted data.

Let us summarize some underlying definitions concerning search spaces before proceeding to more detailed descriptions.

- *Common search space*: A common search space consists of the common control information and is monitored by all UEs in a cell. The number of CCE aggregation levels supported by the common search space is limited to four and eight, compared to the UE-specific search space where four CCE aggregation levels are possible. This reduces the burden on the UE for decoding common control information compared to decoding UE-specific control information. The common search space is used to carry important initial information including paging information, system information, and random access procedures. When searching the common search space, the decoder always starts decoding from the first CCE. This restriction further simplifies the common search space design. The decoding is done on every possible PDCCH candidate set for a given PDCCH format until it successfully decodes the PDCCH which can be present in the common search space. All

UEs monitor the common search space, which may carry paging, RACH response, system information, and uplink TPC commands. The common search space always corresponds to CCEs 0–15, i.e., four decoding candidates on aggregation level 4 (CCEs 0–3, 4–7, 8–11, 12–15), and two decoding candidates on aggregation level 8 (CCEs 0–7, 8–15). The common search space can be used for any PDCCH signaling not restricted to common PDCCH, it can be used to resolve blocking, as well. The UE may attempt a maximum of 12 blind decodings on the common search space.

- *UE-specific search space*: The UE-specific search space carries control information specific to a particular UE and is monitored by at least one UE in a cell. Unlike the common search space, the starting location of the UE-specific search space may vary for each subframe or UE. The starting location of the UE-specific search space is determined in every subframe using a hash function. In the UE-specific search space, the UE finds its PDCCH by monitoring a set of PDCCH candidates (i.e., a set of consecutive CCEs on which the PDCCH could be mapped) in every subframe. If no CRC error is detected when the UE uses its RNTI (i.e., a 16-bit value also referred to as C-RNTI) to unmask the CRC on a PDCCH, the UE determines that the PDCCH carries its own control information. The PDCCH candidate sets correspond to different PDCCH formats. There are four PDCCH formats, i.e., 0, 1, 2, or 3. If the UE fails to decode any PDCCH candidates for a given PDCCH format, then it attempts to decode candidates for other PDCCH formats. This process is repeated for possible PDCCH formats until all designated PDCCHs that can be present in the UE-specific search space are successfully decoded. A PDCCH can be transmitted on an aggregation of one or several CCEs depending on the target reliability and channel conditions. Each PDCCH carries one DCI message. The UE performs blind decoding on the received CCEs, assuming all four different aggregations of CCEs (Figure 9.98). The UE-specific search space indicates the starting offset for blind decoding; different UEs may have different offsets.

In LTE/LTE-Advanced, each scheduling grant is defined based on fixed size CCEs. Four different CCE aggregation levels are defined for the transmission of a control channel (see Table 9.21). The eNB scheduler schedules the PDCCH that should be mapped to the CCE size by applying different coding rates. For example, if the LTE scheduler uses DCI Format 1A which has a length of 28 bits, then a 16-bit CRC is added to the 28 bit information element, resulting in 44 bits. The 44 bits can be mapped to different CCE sizes depending on the desired coding rate, as follows (see Table 9.21):

- PDCCH Format 0: (CCE size = 72 bits) 44 bits are converted to 72 bits using coding rate 44/72.
- PDCCH Format 1: (CCE size = 144 bits) 44 bits are converted to 144 bits using coding rate 44/144.

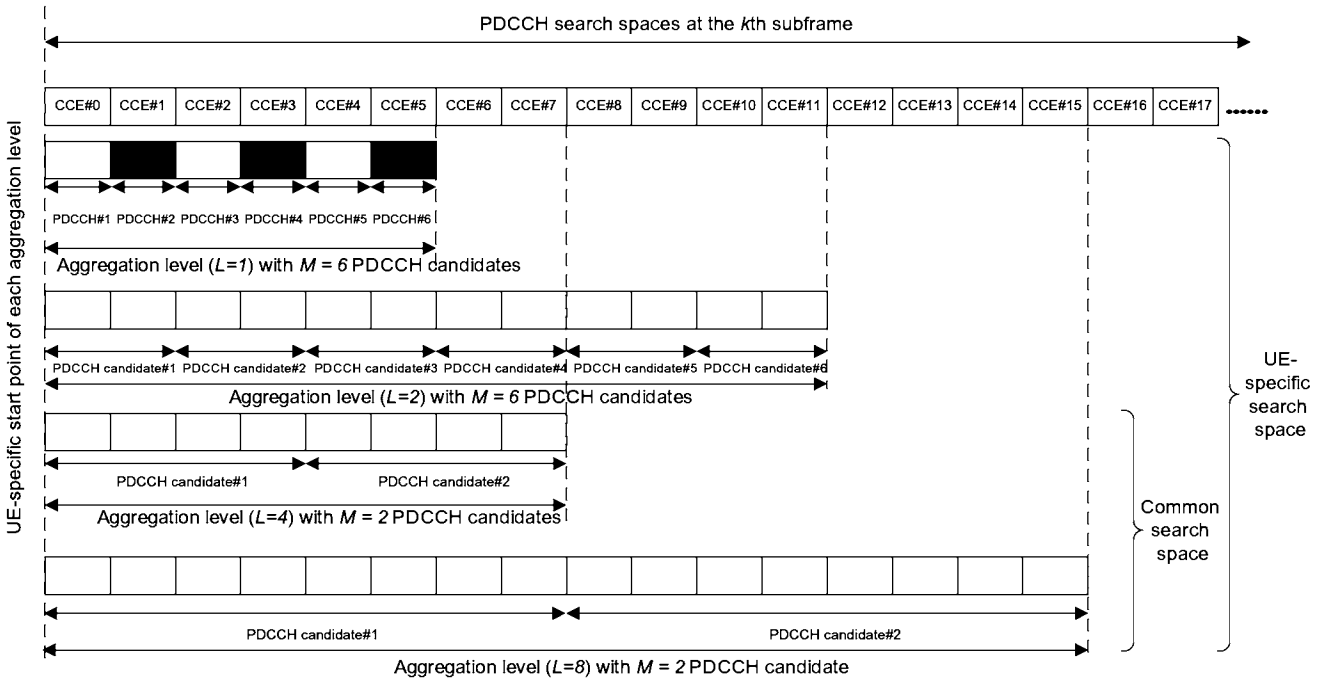


FIGURE 9.98

UE-specific/common search space (UE-specific search space starting offset is calculated based on RNTI) [15,162].

- PDCCH Format 2: (CCE size = 288 bits) 44 bits are converted to 288 bits using coding rate 44/288.
- PDCCH Format 3: (CCE size = 576 bits) 44 bits are converted to 576 bits using coding rate 44/576.

Note that the combination of PDCCH formats and CCE aggregation levels with coding rates greater than 3/4 is not supported. There are a maximum of 32 blind decoding attempts in the UE-specific search space. Assuming two payload sizes per aggregation level, the decoding attempts per payload size include six decoding attempts of 1-CCE aggregation, six decoding attempts of 2-CCE aggregation, two decoding attempts of 4-CCE aggregation, and two decoding attempts of 8-CCE aggregation. If the bandwidth is limited, then only some candidates may be available as the PDCCH region will be truncated. A PDCCH can be mapped to any candidate within its suitable search space as long as the allocated CCEs within the candidate do not overlap with a PDCCH already allocated.

The control region of each serving cell consists of a set of CCEs indexed from 0 to $N_{\mathrm{CCE}}^{k}-1$, where N_{CCE}^{k} denotes the total number of CCEs in the control region of the kth subframe. The UE monitors a set of PDCCH candidates on one or more activated serving cells configured for the UE for control information in every non-DRX subframe, where monitoring implies attempting to decode each of the PDCCHs in the set according to all the monitored DCI formats. The set of PDCCH candidates that a UE monitors are defined in terms of search spaces, where a search space S_k^L with aggregation level $L \in \{1,2,4,8\}$ is defined by a set of PDCCH candidates. The starting point of the UE-specific search space to be monitored by the UE is determined through a hashing function.[20] For each serving cell on which the PDCCH is monitored, the starting point of the CCEs corresponding to the mth PDCCH candidate in search space S_k^L is given by $k' = [(Y_k + m') \mod \lfloor N_{\mathrm{CCE}}^{k}/L \rfloor]L + i$, where Y_k is defined as $Y_k = (AY_{k-1}) \mathrm{mod}\, D$ (i.e., the integer Y_k changes in each subframe, but it is deterministically determined from the value in the previous subframe) for the UE-specific search space S_k^L with aggregation level L or $Y_k = 0$ for common search space, and aggregation levels $L = 4$ or 8. In the latter equation $Y_{-1} = n_{\mathrm{RNTI}} \neq 0, A = 39827, D = 65537$, $k = \lfloor n_s/2 \rfloor$, and n_s is the slot number within a radio frame, and $i = 0, 1, \cdots, L-1$. For the UE-specific search space and the serving cell on which the PDCCH is monitored, if the monitoring UE is configured with a carrier indicator field (CIF), then $m' = m + M_L N_{\mathrm{CIF}}$, where N_{CIF} is the CIF value; otherwise $m' = m$, where $m = 0, 1, \ldots, M_L - 1$, and M_L is the number of PDCCH candidates to monitor in the given search space. In the case of common search space, it is assumed that

[20] A hash function is an algorithm that maps large data sets of variable length to smaller data sets of a fixed length. Hash functions are typically used to accelerate table lookup or data comparison tasks such as finding items in a database, detecting duplicates or similar records in a large file, etc. [31].

$m' = m$. The following rules are applied to the UEs that monitor PDCCH candidates while supporting carrier aggregation [5]:

- The UE monitors one common search space at each of the aggregation levels 4 and 8 on the primary cell.
- A UE, which is not configured with a CIF, monitors one UE-specific search space at each of the aggregation levels 1, 2, 4, and 8 on each activated serving cell.
- A UE configured with a CIF monitors one or more UE-specific search spaces at each of the aggregation levels 1, 2, 4, and 8 on one or more activated serving cells configured through higher layer signaling.
- The common and UE-specific search spaces on the primary cell may overlap.
- A UE that is configured with the CIF associated with monitoring a PDCCH on serving cell monitors the PDCCH whose CRC may have been scrambled with C-RNTI in the UE-specific search space of the serving cell.
- A UE configured with the CIF associated with monitoring the PDCCH on the primary cell monitors the PDCCH whose CRC may have been scrambled with SPS C-RNTI in the UE-specific search space of the primary cell.
- The UE monitors the common search space for the PDCCH without CIF.
- For the serving cell on which the PDCCH is monitored, if the UE is not configured with a carrier aggregation, then it is required to monitor the UE-specific search space for the PDCCH without CIF; otherwise, the UE is required to monitor the UE-specific search space for the PDCCH with CIF.
- A UE is not expected to monitor the PDCCH of a secondary cell, if it is configured to monitor the PDCCH with CIF corresponding to that secondary cell in another serving cell. For the serving cell on which the PDCCH is monitored, the UE monitors PDCCH candidates at least for the same serving cell.
- A UE configured to monitor PDCCH candidates whose CRCs are scrambled with C-RNTI or SPS C-RNTI, with a common payload size and with the same first CCE index n_{CCE}, but with different sets of DCI information fields in the common and UE-specific search spaces on the primary cell, is required to assume that only the PDCCH in the common search space is transmitted by the primary cell.
- A UE configured to monitor PDCCH candidates in a given serving cell with a given DCI format size with CIF, and CRC scrambled with C- RNTI, where the PDCCH candidates may have one or more possible values of CIF for the given DCI format size, can assume that a PDCCH candidate with the given DCI format size may be transmitted in the given serving cell in any UE-specific search space corresponding to any of the possible values of CIF for the given DCI format size.

The aggregation levels defining the search spaces are listed in Table 9.24. The DCI formats that the UE is expected to monitor depend on the configured transmission mode in each serving cell. For the common search spaces, Y_k is set to 0 for the two aggregation levels $L = 4$ and $L = 8$.

9.12.5 Enhanced PDCCH

In LTE Rel-8 downlink control channel design, the PDCCH capacity is constrained by the number of available CCEs and the aggregation level associated with each scheduled PDCCH. Since the size of the time-division-multiplexed control region in each subframe is limited to three OFDM symbols, there might be some scenarios where the eNB exhausts the resources for PDCCH transmission. With the increased use of DM-RS, DCI format 2C will be more often used. DCI format 2C should use at least two CCEs for most UEs, since using only one CCE results in high coding rates (e.g., 0.81 for a 10 MHz FDD system). Moreover, for the UEs at the cell-edge configured for the CoMP mode, a large number of CCEs is needed for the PDCCH to maintain coverage, because the PDCCH cannot use CoMP to improve cell-edge performance. Furthermore, LTE Rel-11 downlink MIMO enhancement for heterogeneous network deployments requires increased PDCCH capacity and the use of virtual cell identities to overcome the limitation of the existing LTE technology in new deployment scenarios related to CoMP, HetNet, etc., where several low-power nodes or subsets of a macro-cell may share the same physical cell identity with the macro-cell.

Sharing the same cell identity by the neighboring nodes would increase intra-cell interference and degrade the channel estimation performance and coherent detection of data and control channels, as well as limit the use of advanced closed-loop SU-MIMO/MU-MIMO techniques. Also, the use of the same identifier by subordinate nodes would prevent transmission/reception node identification and switching. An example case is a non-uniform network deployment comprising a macro-node with several low-power RRHs using the same physical cell identifier as the macro-node. In such a case, there is one common PDCCH for the macro eNB and all RRHs, thus the capacity of the current control region might be too small for sending all PDCCHs without collision.

The use of different scrambling sequences initialized with unique cell identifiers would achieve interference randomization and help the UE to separate data and control channels transmitted by different transmission nodes. Since the reference signals (e.g., CRS, DM-RS, CSI-RS) are also scrambled with cell-specific sequences to allow the UE to conduct various channel estimations and mobility measurements on different transmission nodes, the use of unique identifiers would facilitate these functions, as well as node selection and switching in homogeneous and heterogeneous networks.

Several justifications for the design of a new downlink control structure were discussed in LTE Rel-11 contributions that are summarized below [149–155]:

- Limited PDCCH capacity: The existing PDCCH structure was designed to provide control signaling and resource assignments for UEs in a single transmission point per macro-cell deployment. Under CoMP scenario 4 (see Chapter 12), several RRHs having the same cell identity are deployed in a single macro-cell. In such a scenario, the assignments corresponding to all RRHs may have to be simultaneously transmitted. This can cause a significant

increase in the number of resource assignments and shortage of CCEs per subframe. In CoMP scenario 3, each RRH has a different cell identity relative to the associated macro-cell; however, all nodes share the same time/frequency resources, which may result in significant intra-cell interference. Therefore, the detection performance of the PDCCH in subframes with strong interference may be unsatisfactory. Although the almost-blank subframe concept is supported in LTE Rel-10, it requires the network node to reduce the activity (e.g., transmission power) in the almost-blank subframe, leading to inefficient spectral utilization. Hence, the PDCCH enhancements were necessary for this scenario, as well.

- Increased resource assignment payload size: Since the initial Rel-8 PDCCH design, several improvements have been made to the standard. In particular, new transmission modes, based on UE-specific reference signals, with support for MU-MIMO were designed. It is anticipated that transmission mode 9 will be widely used in future deployments. One of the DCI formats used with transmission mode 9 is DCI format 2C. With its large payload size, the use of DCI format 2C may result in fewer resource assignments (i.e., fewer CCEs per subframe) based on the existing PDCCH structure.
- Interference avoidance/coordination on the control channels: The enhancement of MIMO performance through improved CSI feedback for certain deployment scenarios, especially the case of four transmit antennas in a cross-polarized downlink configuration, in both homogeneous and heterogeneous cases, was of special interest in LTE Rel-11. In such environments, neighboring transmission nodes may interfere with each other. The existing mechanisms in earlier releases of LTE may not be sufficient for robust transmission of control channels in dense and diverse deployments. In that case, it might be beneficial to be able to perform interference avoidance/coordination by orthogonalizing the users in neighboring cells.
- Higher user throughput: Due to capacity limits of the PDCCH, some data resources may not be allocated in a timely manner. One reason for the limitation is the use of the hashing function to map CCEs in the control region. The hashing function (in some cases) may cause two candidate sets to collide, especially when an aggregation level higher than one is chosen for each UE. As a result, the number of assignments that can be transmitted on the PDCCH will be limited, reducing the overall user throughput and increasing transmission latency.
- Use of soft frequency partitioning and fractional frequency reuse with PDCCH: The time-division-multiplexed structure of PDCCH was designed based on a single-frequency partition in each slot/subframe and assuming a frequency reuse factor of one. While hard frequency partitioning is very undesirable from a network deployment and frequency planning perspective, the soft frequency partitioning combined with fractional-frequency reuse techniques has been shown to provide more robustness and reliability for transmission of control signaling and data, due to improved interference mitigation. The use of frequency-division-multiplexing of data and control regions would allow separate power control for each channel type.

- MU-MIMO and beamforming of control channels: The existing structure of the PDCCH relies on CRS for channel estimation and coherent detection. It is shown that CRS has more overhead and is less effective when using closed-loop precoding techniques, beamforming, and MU-MIMO. The beamforming of the control channels can significantly improve the performance of the cell-edge users. The use of DM-RS instead of CRS would facilitate the use of beamforming for the control channels.
- Control overhead reduction: The resource allocation granularity is one OFDM symbol for the PDCCH, thus it amounts to approximately 7% overhead per OFDM symbol. In the frequency-domain, the control channel overhead can be significantly lower if the granularity is defined as one PRB. For instance, in a 10 MHz system, the resource allocation granularity (L1/L2 overhead per PRB) is 2% and 1% in a 20 MHz system. Therefore, the design of the ePDCCH can result in more efficient resource utilization for the control channel.

To overcome the above limitations, an enhanced physical downlink control channel (ePDCCH) was introduced in 3GPP LTE Rel-11. In LTE Rel-11, the ePDCCH can be used to convey scheduling assignments. The ePDCCH is transmitted by aggregating one or several eCCEs. Each eCCE consists of multiple enhanced REGs (eREGs). In an FDD system, the number of eREGs per eCCE $N_{\text{eCCE}}^{\text{eREG}}$ is four for subframes with normal cyclic prefix, and eight for subframes with an extended cyclic prefix. In a TDD system, there are four eREGs per eCCE for subframes with a normal cyclic prefix and special subframe configurations 3, 4, and 5. In addition, there are eight eREGs per eCCE in TDD systems with an extended cyclic prefix and special subframe configurations 1, 2, 3, 5, and 6 [3]. The eCCEs that construct ePDCCHs are numbered from 0 to $N_{\text{ePDCCH}}^{\text{eCCE}} - 1$, where the number of eCCEs per ePDCCH $N_{\text{ePDCCH}}^{\text{eCCE}}$ depends on the ePDCCH format as given in Table 9.25.

The eREGs are used to define the mapping of enhanced control channels to resource elements. There are 16 eREGs in each PRB pair. The eREG indexing is obtained by numbering all resource elements within a PRB pair cyclically from 0 to 15 in an increasing order of first frequency then time, skipping the resource elements designated to DM-RS for antenna ports $p = \{107, 108, 109, 110\}$ for a normal cyclic prefix or $p = \{107, 108\}$ for an extended cyclic prefix. All resource elements with number i in that PRB pair constitute eREG_i (see Figure 9.99).

As shown in Figure 9.100, the mapping of the ePDCCH to resource elements within resource block pairs can be performed in a localized or distributed manner, where the mapping defines how the eCCEs are associated with eREGs and PRB pairs. It must be noted that multiplexing of ePDCCH and PDSCH within a PRB pair is not permitted. If a UE is configured to monitor multiple ePDCCHs, then there are one or two sets of PRB pairs, which the UE is required to monitor. All ePDCCH candidates in ePDCCH set S_{m} exclusively use localized or distributed mapping which is configured by higher layers. Within ePDCCH set S_{m} in subframe i, the eCCEs available for transmission of ePDCCHs are numbered from 0 to $N_{\text{eCCE}}(m, i) - 1$, where eCCE index n corresponds to eREGs numbered $(n \bmod N_{\text{RB}}^{\text{eCCE}}) + \text{j}N_{\text{RB}}^{\text{eCCE}}$ in PRB index $\lfloor n/N_{\text{RB}}^{\text{eCCE}} \rfloor$ for localized mapping, or

Table 9.25 ePDCCH Formats [3]

	Number of eCCEs per ePDCCH (N_{ePDCCH}^{eCCE})			
	Regular TDD/FDD Subframes and TDD Special Subframe Configurations 3, 4, 8 with Normal Cyclic Prefix (DCI Formats 2/2A/2B/2C/2D and $N_{RB}^{DL} \geq 25$ or Any DCI Format and $N_{ePDCCH} < 104$)		All Other Cases	
ePDCCH Format	Localized Transmission	Distributed Transmission	Localized Transmission	Distributed Transmission
0	2	2	1	1
1	4	4	2	2
2	8	8	4	4
3	16	16	8	8
4	–	32	–	16

it corresponds to eREGs numbered $\lfloor n/N_{RB}^{S_m} \rfloor + jN_{RB}^{eCCE}$ in PRB indices $(n + j\max[1, N_{RB}^{S_m}/N_{eCCE}^{eREG}])\bmod N_{RB}^{S_m}$ for distributed mapping. The parameter $j = 0, 1, \ldots, N_{eCCE}^{eREG} - 1$, N_{eCCE}^{eREG} is the number of eREGs per eCCE, and $N_{RB}^{eCCE} = 16/N_{eCCE}^{eREG}$ is the number of eCCEs per resource block pair (see Figure 9.99). The PRB pairs that form ePDCCH set S_m are numbered in ascending order from 0 to $N_{RB}^{S_m} - 1$ [3].

The physical layer processing of the ePDCCH can be described as follows. The block of bits $(b_0, b_1, \ldots, b_{N_{bit}-1})$ that is going to be transmitted on an ePDCCH in a subframe is scrambled, resulting in a block of scrambled bits $(\tilde{b}_0, \tilde{b}_1, \ldots, \tilde{b}_{N_{bit}-1})$ in which $\tilde{b}_i = [b_i + c(i)]\bmod 2$. The UE-specific scrambling sequence $c(i)$ defined in previous sections is initialized with $c_{init} = \lfloor n_s/2 \rfloor 2^9 + N_{ID}^{ePDCCH(i)}$, where index i is the ePDCCH set number, and $N_{ID}^{ePDCCH(i)}$ is provided by higher layers. The block of scrambled bits $(\tilde{b}_0, \tilde{b}_1, \ldots, \tilde{b}_{N_{bit}-1})$ is modulated using QPSK modulation, resulting in a block of complex-valued modulation symbols $(d_0, d_1, \ldots, d_{N_s-1})$. The block of complex-valued modulated symbols is mapped to a single layer and precoded according to $y_i = d_i|_{i=0,1,\ldots,N_s-1}$. The precoded block of symbols can be mapped to physical resource elements using localized or distributed mapping.

The block of complex-valued symbols $(y_0, y_1, \ldots, y_{N_s-1})$ is sequentially mapped to resource elements (k, l) starting with y_0 to the associated antenna port, provided that those physical resources are part of the eREGs assigned for transmission and are not part of a physical resource-block pair used for transmission of PBCH or synchronization signals, and are not used for CRSs, CSI-RSs for the

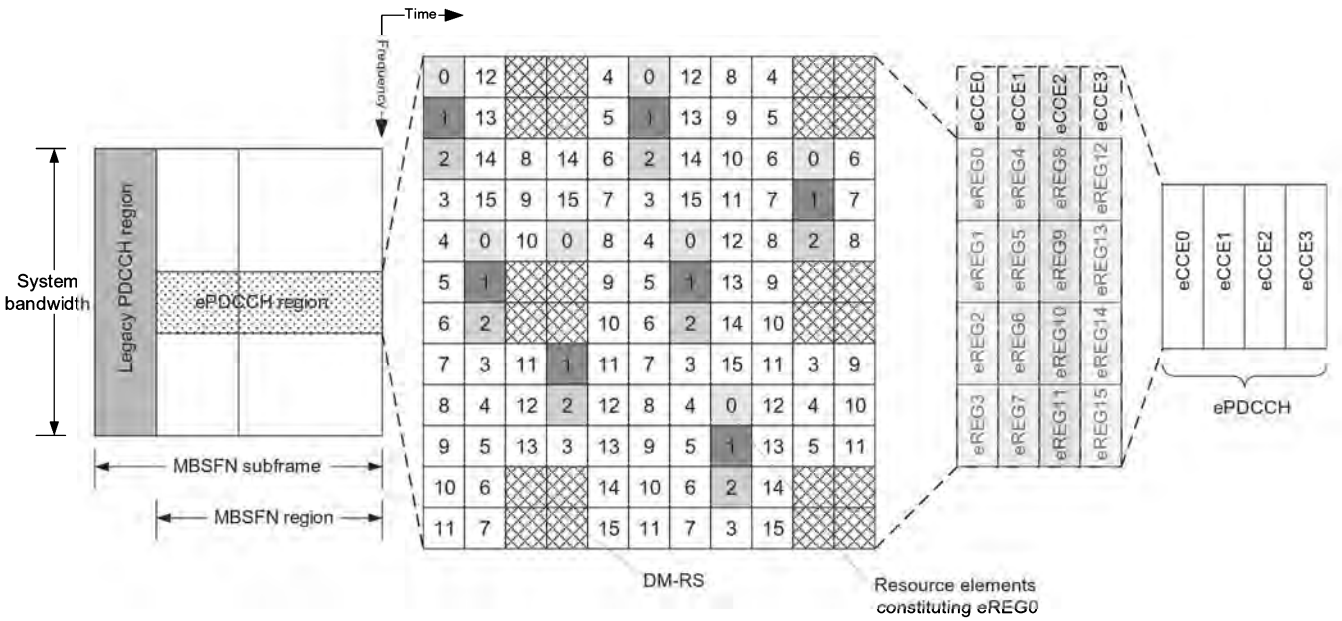


FIGURE 9.99

Structure and formation of an example ePDCCH, eCCE, and eREG [3].

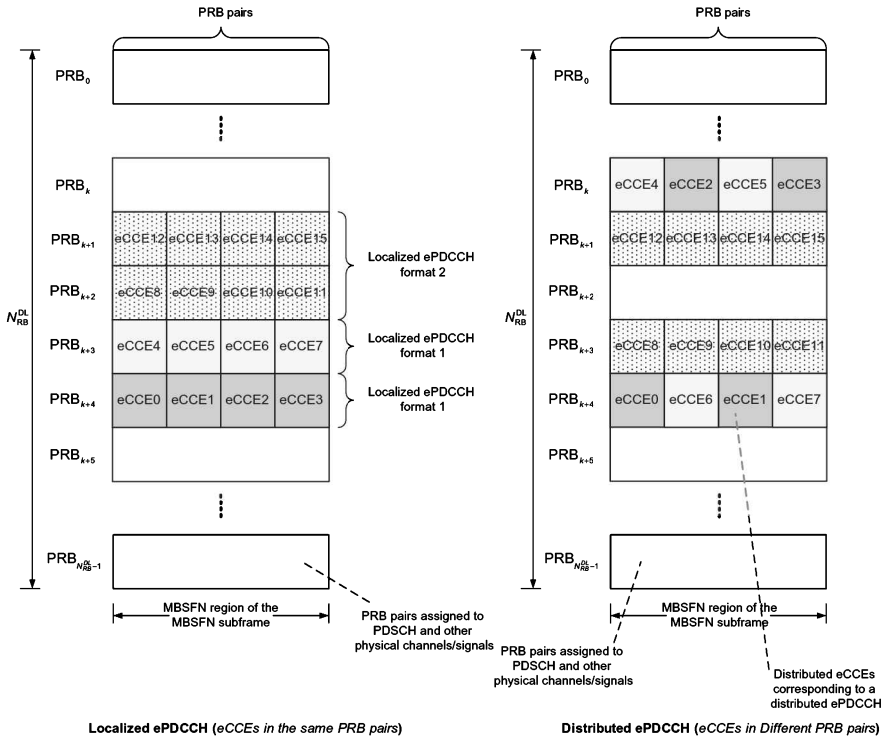


FIGURE 9.100

Structure and formation of localized and distributed ePDCCH and eCCEs [3].

specific UE; and further the index l in the first slot in a subframe stratifies the condition $l \geq l_{\text{ePDCCH}_{\text{start}}}$. The mapping to resource elements (k, l) on antenna port p is performed in increasing order of first the index k and then the index l, starting with the first slot and ending with the second slot in a subframe. For localized transmission, the single antenna port p to use for ePDCCH transmission is parameterized with parameter $n_{\text{ap}} = n_{\text{eCCE}_{\text{L}}} \bmod N_{\text{RB}}^{\text{eCCE}} + n_{\text{RNTI}} \bmod \min(N_{\text{ePDCCH}}^{\text{eCCE}}, N_{\text{RB}}^{\text{eCCE}})$, where $n_{\text{eCCE}_{\text{L}}}$ denotes the lowest eCCE index used by the referenced ePDCCH transmission in the ePDCCH set, n_{RNTI} corresponds to the RNTI associated with the ePDCCH transmission, and $N_{\text{ePDCCH}}^{\text{eCCE}}$ is the number of eCCEs per referenced ePDCCH. For FDD/TDD subframes with a normal cyclic prefix and TDD special subframe configurations 3, 4, and 8, depending on the value of parameter n_{ap} the following antenna port assignments are made [3]:

$$\begin{aligned} n_{\text{ap}} = 0 &\rightarrow p = 107 \\ n_{\text{ap}} = 1 &\rightarrow p = 108 \\ n_{\text{ap}} = 2 &\rightarrow p = 109 \\ n_{\text{ap}} = 3 &\rightarrow p = 110 \end{aligned}$$

For distributed transmission, each resource element in an eREG is associated with one out of two antenna ports in an alternating manner, where $p \in \{107, 109\}$ for a normal cyclic prefix and $p \in \{107, 108\}$ for an extended cyclic prefix.

For each serving cell, higher layer signaling can configure a UE with one or multiple ePDCCH-PRB sets for ePDCCH monitoring. Each ePDCCH-PRB set consists of a set of eCCEs indexed from 0 to $N_{\text{eCCE}}(p,k) - 1$, where $N_{\text{eCCE}}(p,k)$ denotes the number of eCCEs in ePDCCH-PRB set p in subframe k. Each ePDCCH-PRB set can be configured for either localized or distributed ePDCCH transmission. The UE is required to monitor a set of ePDCCH candidates on one or more activated serving cells configured via higher-layer signaling for control information, where monitoring implies attempting to blindly decode each of the ePDCCHs in the set according to the monitored DCI formats.

The set of ePDCCH candidates that the UE is expected to monitor are defined in terms of ePDCCH UE-specific search spaces. For each serving cell, the subframes in which the UE monitors ePDCCH UE-specific search spaces are configured via higher-layer signaling. In TDD systems using a normal cyclic prefix in downlink subframes, the UE is not expected to monitor the ePDCCH in special subframe configurations 0 and 5. A UE is not expected to monitor an ePDCCH candidate whose eCCE(s) is mapped to a PRB pair that overlaps in frequency with a transmission of either PBCH or primary/secondary synchronization signals in the same subframe.

The UE is required to monitor one common search space in every non-DRX subframe at aggregation levels 4 and 8 on the primary cell. If a UE is not configured for ePDCCH monitoring, and if it is not configured with a CIF, then it is required to monitor one PDCCH UE-specific search space at aggregation levels 1, 2, 4, 8 on each activated serving cell in every non-DRX subframe. In another case where a UE is not configured to monitor ePDCCH, but it is configured with a CIF, then it is required to monitor one or more UE-specific search spaces at aggregation levels 1, 2, 4, 8 on one or more activated serving cells that are configured through higher-layer signaling in every non-DRX subframe. Furthermore, if a UE is configured for ePDCCH monitoring on an activated serving cell, but it is not configured with CIF, then it is required to monitor one PDCCH UE-specific search space at aggregation levels 1, 2, 4, 8 on the serving cell in all non-DRX subframes where ePDCCH is not monitored on that serving cell. If a UE is configured for ePDCCH monitoring on a serving cell, and if that serving cell is activated, and if the UE is configured with a CIF, then the UE is required to monitor one or more PDCCH UE-specific search spaces at each of the aggregation levels 1, 2, 4, and 8 on that serving cell as configured by higher layer signaling in all non-DRX subframes where ePDCCH is not monitored on that serving cell. The common and PDCCH UE-specific search spaces on the primary cell may overlap. It can be concluded that in each subframe on an activated serving cell, the UE either monitors ePDCCH or PDCCH and not both of them, in order to simplify the UE search mechanism.

9.12.6 Physical downlink shared channel

The physical downlink shared channel is a downlink channel that is used for transmission of all user data, paging messages, and part of the system information which is not carried in the PBCH. The data is transmitted on the PDSCH in units of transport blocks that correspond to the MAC sub-layer PDUs. The transport blocks are received from the MAC sub-layer in each TTI where the default TTI value in LTE/LTE-Advanced is 1 ms, corresponding to one subframe duration. The maximum number of transport blocks that can be transmitted per UE per subframe is limited to two, depending on the transmission mode selected for the PDSCH for each UE. The transmission mode defines the multi-antenna configuration that is used for data transmission over the PDSCH. There are 10 transmission modes specified for LTE Rel-11 that are summarized in Table 9.26. Note that transmission mode 10 was included as part of Rel-11 enhancements to support coordinated multi-point transmission/reception; especially with multiple CSI processes configured, TM10 can be seen as enhanced TM9 with well-defined UE behavior in interference measurement via CSI interference measurement configuration [3,5].

With the exception of transmission modes 7, 8, 9, and 10, the phase reference for coherent demodulation of PDSCH is obtained from CRSs and the number of eNB antenna ports used for transmission of PDSCH is the same as the number of antenna ports used in the cell for PBCH transmission. In transmission modes 7, 8, 9, and 10, the UE-specific reference signals provide the phase reference for coherent demodulation of the PDSCH. The configured transmission mode further controls the formats of the associated downlink control signaling messages as given in Table 9.26, and the type of channel quality feedback from the UE.

The coded transport block is scrambled with a binary scrambling sequence. The sequence is based on cell identity and RNTI to make it cell- and UE-specific. The binary scrambled bits are modulated using a suitable modulation scheme considering the channel conditions and required data rate. The permissible modulation schemes for PDSCH are QPSK (2 bits/symbol), 16QAM (4 bits/symbol), or 64QAM (6 bits/symbol). Support of 64QAM is mandatory for all LTE/LTE-Advanced UE categories. The selection of modulation scheme is based on CQI feedback from the UE. The resulting codewords after modulation are mapped to spatial layers. Spatial multiplexing uses one to eight layers less than or equal to the number of antenna ports. For transmit diversity, the number of layers is equal to the number of antenna ports. The precoding matrix applies antenna weighting (phase and amplitude) for each data stream, and maps the stream to different antenna ports. In resource blocks in which UE-specific reference signals are not transmitted, PDSCH is transmitted on the same set of antenna ports used for PBCH.

All downlink resource elements are available for the PDSCH except those reserved for special purposes, such cell-specific, positioning, MBSFN, CSI, and UE-specific reference signals, synchronization signals, and control signaling. Therefore, when the UE is informed through control signaling of a resource allocation within a pair of resource blocks in a subframe, it is only the available resource elements within those resource block pairs which actually carry PDSCH

Table 9.26 Summary of PDSCH Transmission Modes in LTE/LTE-Advanced when using C-RNTI [5,13]

Release	Mode	PDSCH Transmission Scheme	DCI Format	Search Space	Channel State Information Feedback
8	1	Single-antenna transmission on antenna port 0	1	UE-specific	CQI
			1A	Common and UE-specific	
	2	Transmit diversity	1	UE-specific	CQI
			1A	Common and UE-specific	
	3	Transmit diversity	1A	Common and UE-specific	CQI and RI
		Open-loop spatial multiplexing or transmit diversity	2A	UE-specific	
	4	Transmit diversity	1A	Common and UE-specific	CQI, RI, and PMI
		Closed-loop spatial multiplexing or transmit diversity	2	UE-specific	
	5	Transmit diversity	1A	Common and UE-specific	CQI and PMI
		MU-MIMO	1D	UE-specific	
	6	Transmit diversity	1A	Common and UE-specific	CQI and PMI
		Closed-loop spatial multiplexing using single-layer transmission (closed-loop rank-1 precoding)	1B	UE-specific	
	7	Single-antenna transmission on antenna port 5 (transmission using UE-specific reference signals with a single layer)	1	UE-specific	CQI
		Single-antenna transmission using antenna port 0 or transmit diversity	1A	Common and UE-specific	

(Continued)

Table 9.26 (Continued)

Release	Mode	PDSCH Transmission Scheme	DCI Format	Search Space	Channel State Information Feedback
9	8	Single-antenna transmission using antenna port 0 or transmit diversity	1A	Common and UE-specific	CQI (PMI/RI if configured by eNB)
		Dual-layer transmission on antenna ports 7 and 8 or single-antenna transmission on antenna port 7 or 8 (transmission using UE-specific reference signals with up to two layers)	1B	UE-specific	
10	9	Single-antenna transmission on antenna port 0 or transmit diversity	1A	Common and UE-specific	CQI (PMI/RI if configured by eNB)
		Spatial multiplexing up to eight layers on antenna ports 7–14 (transmission using UE-specific reference signals with up to eight layers)	2C	UE-specific	
11	10	Coordinated multi-point transmission and reception (support CoMP operation and configure Rel-11 non-zero-power CSI-RS resources and interference measurement resources)	1A/2D	UE-specific	CQI (PMI/RI if configured by eNB)

data. The allocation of resource block pairs to PDSCH transmission for a particular UE is signaled to the UE by means of dynamic control signaling transmitted at the start of the relevant subframe using ePDCCH/PDCCH. The mapping of PDSCH resource elements in a subframe is illustrated in Figure 9.101. The modulation symbols are mapped to resource elements (k, l) in the allocated resource blocks. The mapping is performed first along frequency and then along symbols in time. The modulation symbols are mapped by incrementing the sub-carrier index k first along the allocated resource blocks. When the final resource block is reached, the symbol index l is incremented. After mapping, the OFDM modulation is performed across all sub-carriers on a symbol-by-symbol basis.

The paging control channel (PCCH) is a logical channel that is mapped to the paging channel (PCH) which is a transport channel. The PCH is transmitted via PDSCH similar to other data transmissions. When PCH is transmitted, a unique 16-bit identifier known as P-RNTI with a value of 0xFFFE is used to signal paging messages in the PDCCH. Multiple UEs may be paged simultaneously using a single paging allocation with the UE identity indicated within the paging message itself (Figure 9.102).

A UE is required to decode the corresponding PDSCH allocation in the same subframe (considering the restriction on the number of transport blocks signaled through RRC messages) upon detection of an ePDCCH/PDCCH of a serving cell with DCI format 1/1A/1B/1C/1D or 2/2A/2B/2C/2D intended for the UE in that subframe [5]. The UE may assume that PRSs are not present in resource blocks in which it is required to decode PDSCH when a PDCCH whose CRC is scrambled with SI-RNTI or P-RNTI with DCI format 1A or 1C is successfully detected. If the UE is configured with the CIF for a given serving cell, it must assume that the CIF is not present in any PDCCH of the serving cell in the common search space; otherwise, the UE must assume that for the given serving cell, the CIF is present in the ePDCCH/PDCCH located in the UE-specific search space, when the ePDCCH/PDCCH CRC is scrambled with C-RNTI or SPS C-RNTI.

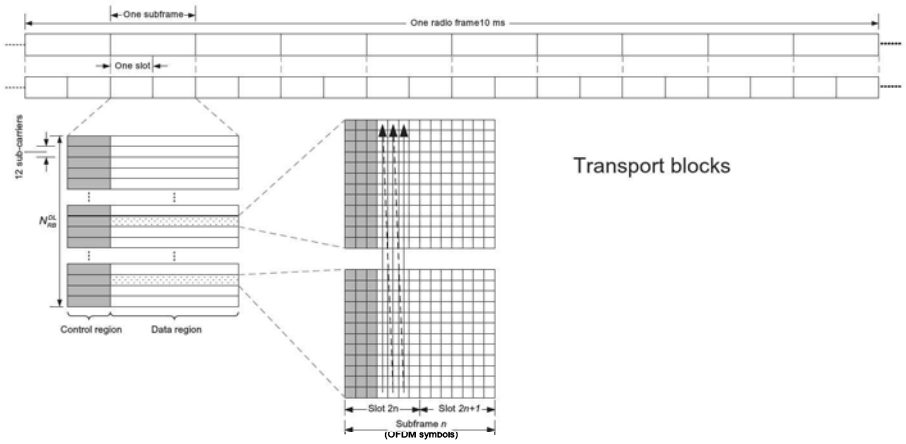


FIGURE 9.101

Illustration of PDSCH resource element mapping.

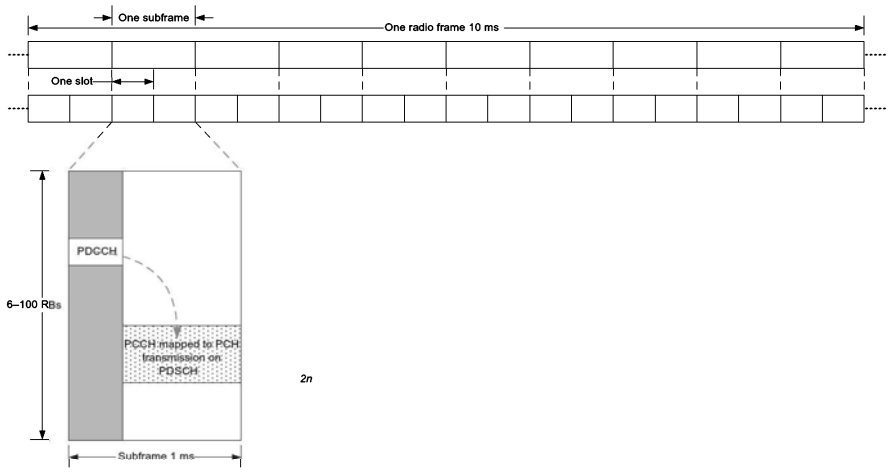


FIGURE 9.102

PCH transmission on PDSCH.

If a UE is configured by higher layers to decode the PDCCH with its CRC scrambled with SI-RNTI, the UE is required to decode PDCCH and the corresponding PDSCH according to any of the permissible combinations of DCI formats and multi-antenna schemes. The PDSCH corresponding to these PDCCHs is also scrambled with SI-RNTI. Similarly, if a UE is configured to monitor PDCCHs whose CRCs are scrambled with P-RNTI, C-RNTI, SPS C-RNTI, or RA-RNTI, the UE is required to decode the PDCCH and the corresponding PDSCH according to any of the permissible combinations of DCI formats and multi-antenna schemes. An example is given in Table 9.26 for C-RNTI [5]. Note that if the CRC of PDCCH is scrambled with broadcast identifiers such as SI-RNTI, P-RNTI, and RA-RNTI, the UE only monitors the common search space and decodes the PDSCH accordingly. For multi-user MIMO transmission of PDSCH, the UE may assume that the eNB transmission on PDSCH is on one layer. An additional power offset of $\delta_{\text{power-offset}} = -3$ dB may be applied if the downlink power offset field signaled on PDCCH with DCI format 1D is zero.

9.13 Principles of multi-antenna transmission

Design of mobile broadband wireless communication systems capable of data transmission rates in excess of 1 Gbps has been of practical interest in the past decade to enable the new generations of cellular systems and IMT-Advanced. The use of multiple antennas at the transmitter and/or receiver commonly known as the MIMO system has become a pragmatic and cost-effective approach that offers substantial gain in making 1 Gbps wireless links a reality. This section provides an

overview of the MIMO wireless technology including channel models, performance limits, coding, and transceiver design. MIMO techniques can increase system throughput and transmission reliability without increasing the required bandwidth. While most communication systems suffer in multi-path channels, a MIMO system benefits from propagation over different paths through which the signals arrive at the receiver. A MIMO system typically performs better in an indoor environment while it may not properly perform in LoS environments.

The performance improvements resulting from the use of multi-antenna systems are mainly due to array gain, diversity gain, spatial multiplexing gain, and interference reduction that may be achieved through appropriate MIMO configurations. The main advantages of the multi-antenna techniques can be summarized as follows [20]:

- Array gain can be achieved through signal processing at the transmitter and/or the receiver, and results in an increase in average SINR at the receiver due to coherent combining effect. The increased SINR further results in improved coverage and user throughput. Transmit or receive array gain requires channel knowledge in the transmitter and receiver, respectively, and further depends on the number of transmit and receive antennas. The channel knowledge in the receiver is typically available (through channel estimation based on downlink reference signals), whereas CSI in the transmitter is in general more difficult to attain.
- The signal power is faded randomly in a wireless communication channel. Diversity is an effective technique to mitigate fading in a wireless channel. The diversity schemes rely on transmitting the signal over multiple and ideally independent fading channels in time, frequency, and/or space. Spatial diversity is preferred over time or frequency diversity as it does not consume additional transmission time or bandwidth. If the $N_{RX} \times N_{TX}$ links forming the MIMO channel fade independently and the transmitted signal is properly constructed, the receiver can combine the arriving signals such that the resulting signal exhibits considerably reduced amplitude variability relative to a single-input single-output (SISO) link, and a diversity gain on the order of $N_{RX} \times N_{TX}$ can be achieved. Spatial diversity gain in the absence of channel knowledge at the transmitter can be achieved using suitably designed transmit signal which is known as space–time coding.
- MIMO channels provide a linear increase in capacity as a function of $\min(N_{RX}, N_{TX})$ without requiring additional power or transmission bandwidth. This gain, referred to as spatial multiplexing gain, is realized by transmitting independent data streams from individual transmit antennas. Note that the number of independent data streams is limited to $\min(N_{RX}, N_{TX})$. Under good channel conditions and sufficiently high SINR values, the receiver can detect different data streams, yielding a linear increase in capacity.
- Co-channel interference is generated due to frequency reuse in wireless channels. When multiple antennas are used, the differentiation between the spatial signatures of the desired signal and co-channel signals can be utilized to reduce interference. While interference cancelation requires proper knowledge of the desired signal's channel, exact knowledge of the interferers'

channels may not be necessary. The interference avoidance schemes can also be implemented at the transmitter where the goal is to minimize the interference energy sent toward the co-channel users while delivering the signal to the desired user. Interference reduction allows aggressive frequency reuse, and thereby increases multi-cell capacity. It must be noted that not all advantages of MIMO schemes can be simultaneously achieved due to conflicting demands on the spatial degrees of freedom or the number of transmit and/or receive antennas. The degree to which these conflicts can be resolved depends on the signaling scheme and transceiver design.

9.13.1 Capacity of MIMO channels

As shown in Figure 9.103, a generic MIMO system consists of a MIMO transmitter with N_{TX} transmit antennas, a MIMO receiver with N_{RX} receive antennas, and $N_{RX} \times N_{TX}$ paths or channels between transmit and receive antennas. Let $x_k(t)$ denote the transmitted signal from the kth transmit antenna at time t, then the received signal at the lth antenna can be expressed as follows:

$$y_l(t) = \sum_{k=1}^{N_{TX}} h_{lk}(t)^* x_k(t) + n_l(t) \tag{9.54}$$

where $h_{lk}(t)$ and $n_l(t)$ are the channel impulse response between the kth transmit antenna, the lth receive antenna, and the additive noise at the lth receive antenna port, respectively. The above equation can be written in the frequency-domain as:

$$Y_l(\omega) = H_{lk}(\omega)X_k(\omega) + N_l(\omega) \tag{9.55}$$

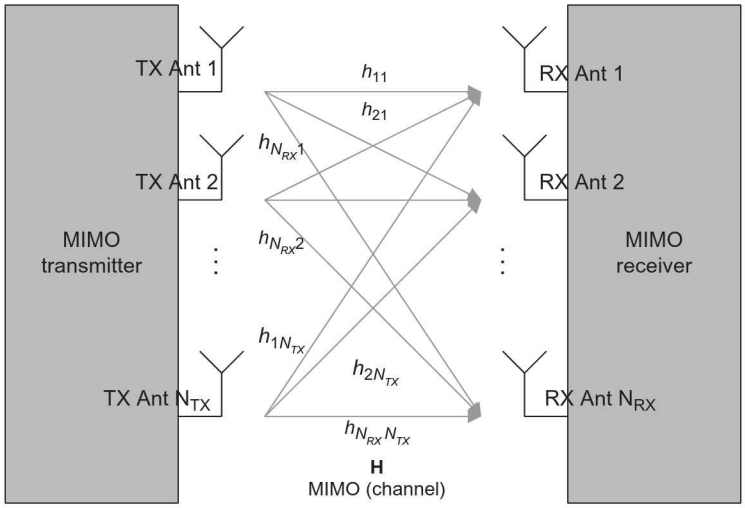


FIGURE 9.103

Illustration of the general principle of a MIMO system.

If $\mathbf{x}(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_{N_{TX}}(\omega))^{\mathrm{T}}$, $\mathbf{y}(\omega) = (Y_1(\omega),\ Y_2(\omega), \ldots, Y_{N_{RX}}(\omega))^{\mathrm{T}}$, and $\mathbf{n}(\omega) = (N_1(\omega), N_2(\omega), \ldots, N_{N_{RX}}(\omega))^{\mathrm{T}}$ denote the Fourier transform vectors of $\mathbf{x}_k(t)$, $\mathbf{y}_l(t)$, and $\mathbf{n}_l(t)$, respectively, then $\mathbf{y}(\omega) = \mathbf{H}(\omega)\,\mathbf{x}(\omega) + \mathbf{n}(\omega)$, where $\mathbf{H}(\omega)$ is an $N_{\mathrm{RX}} \times N_{\mathrm{TX}}$ channel matrix with $H_{lk}(\omega)|_{k=1,2,\ldots,N_{TX};l=1,2,\ldots,N_{\mathrm{RX}}}$ entries.

Assuming a linear time-invariant MIMO channel, the channel input–output relationship can be further described in the discrete time-domain as follows:

$$y_l(nT) = \sum_{k=1}^{N_{\mathrm{TX}}} \sum_{m=0}^{M-1} h_{lk}(mT) x_k[(n-m)T] + n_l(nT) \quad 0 \le n \le M-1;\ 1 \le l \le N_{\mathrm{RX}} \tag{9.56}$$

where $x_k(n)|_{k=1,2,\ldots,N_{\mathrm{TX}}}$ and $y_l(n)|_{l=1,2,\ldots,N_{\mathrm{RX}}}$ represent the channel input and output time-domain signals, respectively. In the case of time-varying channels, Eq. (9.54) can be written as $y_l(t) = \sum_{k=0}^{N_{\mathrm{TX}}-1} h_{lk}(t,\tau) x_k(\tau) + n_l(t)$, where $h_{lk}(t,\tau)$ denotes the time-varying impulse response of the lkth channel. The matrix form of Eq. (9.55) in the frequency-domain sampled at a single frequency ω_m can be written as follows:

$$\mathbf{y}(\omega_m) = \mathbf{H}(\omega_m)\mathbf{x}(\omega_m) + \mathbf{n}(\omega_m) \tag{9.57}$$

In an OFDM system, the signal processing is inherently performed in the frequency-domain. Furthermore, OFDM transforms a frequency-selective fading channel to a flat-fading channel when considering narrowband orthogonal sub-carriers. In such a system, the MIMO signal processing can be performed at each sub-carrier. This is the main reason for the suitability of MIMO extension to an OFDM system. When MIMO processing is done at each sub-carrier, the MIMO channel input–output relationship can be demonstrated as $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, where the channel between the transmitter and the receiver is typically modeled as a finite impulse response (FIR) filter. In this case, each tap is typically a complex-valued Gaussian random variable with exponentially decaying magnitudes. The tap delays correspond to the RMS delay spread and the channel type (e.g., low delay spread or flat fading, high delay spread or frequency selective fading). There is a new realization of the channel at every transmitted packet, if the channel remains invariant for the duration of the packet; otherwise, the variation of the channel is explicitly modeled in the signal detection. As mentioned earlier, there are $N_{\mathrm{RX}} \times N_{\mathrm{TX}}$ paths between the transmitter and the receiver where each channel is the sum of several FIR filters with different delay spreads. The channels may or may not be correlated. The MIMO schemes can be used with non-OFDM systems when the channel is modeled as flat fading such that:

$$y_l(nT) = \sum_{k=1}^{N_{TX}} h_{lk} x_k(nT) + n(nT) \tag{9.58}$$

An important question is to what extent MIMO techniques can increase the throughput and improve the reliability of the wireless communication systems.

This question can be answered by calculating the information theoretic capacity of a SISO channel, and comparing it with that of single-input multiple-output (SIMO), multiple-input single-output (MISO), and MIMO channels. For a memory-less SISO channel (i.e., one transmit and one receive antenna), the channel capacity is given by:

$$C_{\text{SISO}} = \log_2(1 + \gamma|h|^2) \tag{9.59}$$

where h is the normalized complex-valued gain/attenuation of a fixed wireless channel or that of a particular realization of a random channel, and $\gamma = E_s/N_0$ denotes the SNR at the receive antenna port. As the number of receive antennas increases, the statistics of channel capacity improve. Using N_{RX} receive antennas and one transmit antenna, a SIMO system is formed with a capacity given by (when the channel is unknown to the transmitter):

$$C_{\text{SIMO}} = \log_2(1 + \gamma \sum_{i=1}^{N_{\text{RX}}} |h_i|^2) \tag{9.60}$$

where h_i is the gain of the ith channel corresponding to the ith receive antenna. Note that increasing the value of N_{RX} results in a logarithmic increase in average channel capacity. It can be shown that knowledge of the channel at the transmitter in SIMO cases does not provide any capacity benefit. In the case of MISO or transmit diversity, where the transmitter does not typically have knowledge of the channel, the capacity is given by:

$$C_{\text{MISO}} = \log_2(1 + \frac{\gamma}{N_{\text{TX}}} \sum_{i=1}^{N_{\text{TX}}} |h_i|^2) \tag{9.61}$$

It is noted from the above equation that when the channel information is not available at the transmitter, $C_{\text{MISO}} < C_{\text{SIMO}}$. The power normalization factor N_{TX} ensures that the total transmit power is uniformly distributed among the transmit antennas. Furthermore, one notes the absence of an array gain in this case (MISO) compared to that of the receive diversity scenario where the energy of the multipath channels can be coherently combined. In addition, the MISO capacity has a logarithmic relationship with N_{TX} similar to that of a SIMO scheme. When the channel information is available at the transmitter, the capacity of a MISO system can approach that of a SIMO system [20].

The use of diversity at both transmitter and receiver sides gives rise to a MIMO system. The capacity of a MIMO system with N_{TX} transmit antennas and N_{RX} receive antennas is expressed as follows:

$$C_{\text{MIMO}} = \log_2\left(\det\left[\mathbf{I} + \frac{\gamma}{N_{\text{TX}}}\mathbf{H}\mathbf{H}^{\text{H}}\right]\right) \tag{9.62}$$

where $\mathbf{I}$ is an $N_{\text{RX}} \times N_{\text{RX}}$ identity matrix and $\mathbf{H}$ is the $N_{\text{RX}} \times N_{\text{TX}}$ channel matrix. Note that both MISO and MIMO channel capacities are based on equal power

and uncorrelated sources. It is demonstrated in the literature that the capacity of the MIMO channel increases linearly with $\min(N_{RX}, N_{TX})$ rather than logarithmically as in the case of MISO or SIMO channel capacity, since the determinant operator yields the product of non-zero eigenvalues of its channel-dependent matrix argument, each eigenvalue characterizing the SNR over a SISO eigen-channel. It will be shown later that the overall MIMO channel capacity is the sum of capacities of each of these SISO eigen-channels. The increase in capacity is dependent on properties of the channel eigenvalues. If the channel eigenvalues decay rapidly, then linear growth in capacity will not occur. However, the eigenvalues have a known limiting distribution and tend to be spaced out along the range of this distribution. Hence, it is unlikely that most eigenvalues are very small and the linear growth is indeed achieved [20].

Figures 9.104 and 9.105 show the impulse responses and the frequency responses of a large delay-spread MIMO channel (two transmit and two receive antennas) measured over the extent of a subframe. It is noted that channel characteristics vary from one path (between one transmit and one receive antenna) to another, and across time from one symbol to another. The information theoretic capacities of some MIMO channels with a different number of transmit and receive antennas are compared in Figure 9.106.

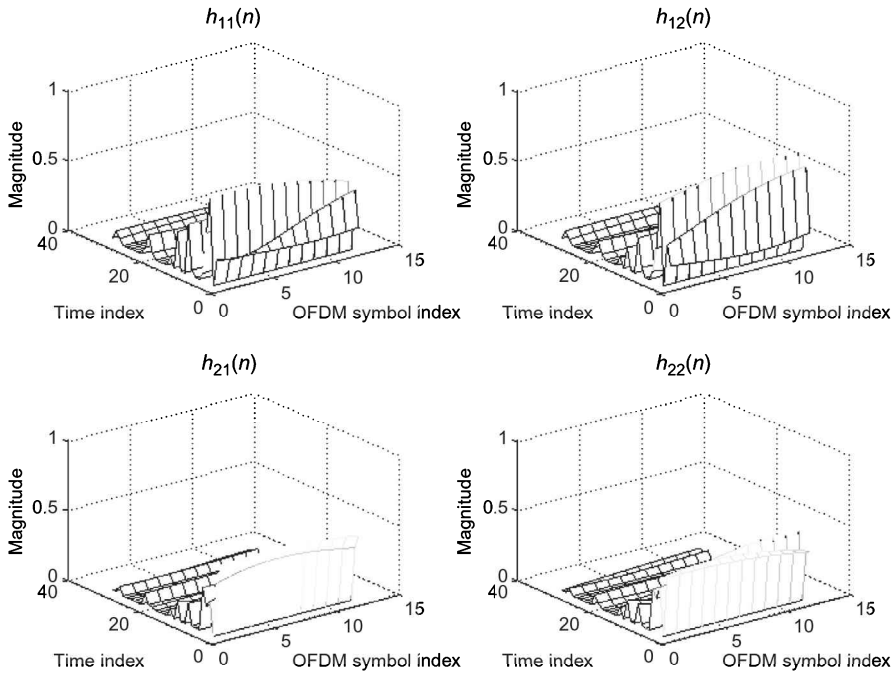


FIGURE 9.104

Example of multipath impulse responses of $N_{TX} = 2$, $N_{RX} = 2$ MIMO channel [141].

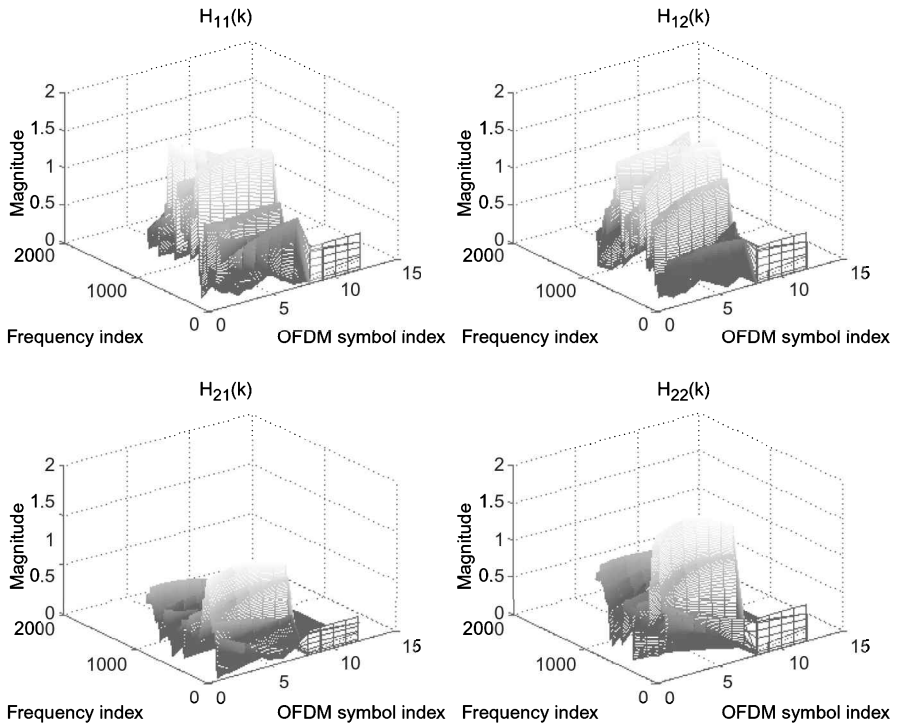


FIGURE 9.105

Example of multipath frequency responses of $N_{TX} = 2$, $N_{RX} = 2$ MIMO channel [141].

The capacity of the MIMO channel can be calculated under various conditions and different assumptions. Depending on whether the receiver has perfect channel knowledge, or whether the channel is flat fading or frequency-selective fading, different expressions for the channel capacity can be obtained. The MIMO channel capacity in Eq. (9.62) can be analyzed under two different assumptions: (1) the transmitter has no channel knowledge; and (2) the transmitter has perfect channel knowledge through feedback from the receiver or reciprocity of the downlink and uplink channels. Let us denote by $\boldsymbol{\Xi}$ the $N_{TX} \times N_{TX}$ covariance matrix of the channel input vector $\mathbf{x}$, and further assume that the channel is unknown to the transmitter, then it can be shown that the MIMO channel capacity can be written as [20]:

$$C_{MIMO} = \log_2(\det[\mathbf{I} + \mathbf{H}\boldsymbol{\Xi}\mathbf{H}^H]) \tag{9.63}$$

where $\text{trace}(\boldsymbol{\Xi}) \leq \gamma$ ensures that the total signal power does not exceed a certain limit. It can be shown that for equal transmit power and uncorrelated sources $\boldsymbol{\Xi} = (\gamma/N_{TX})\mathbf{I}$, Eq. (9.63) becomes identical to Eq. (9.62). This is true when the channel matrix is unknown to the transmitter and the input signal is Gaussian-distributed, maximizing the mutual information. If the receiver measures and

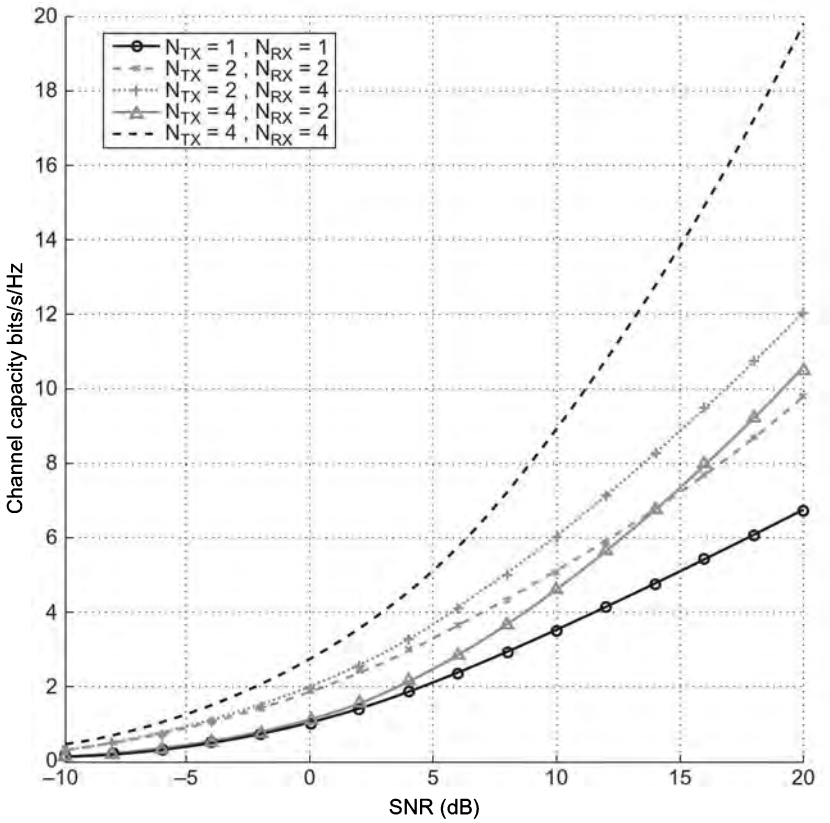


FIGURE 9.106

Comparison of the information theoretic capacity of some MIMO channels [141].

sends channel quality feedback or CSI to the transmitter, the covariance matrix $\boldsymbol{\Xi}$ is not proportional to the identity matrix, rather it is constructed from a water-filling algorithm. If one compares the capacity achieved assuming equal transmit-power and unknown channel with that of perfect channel estimation through feedback, then the capacity gain due to use of feedback is obtained. For the independent identically distributed Rayleigh fading scenario, the linear capacity growth discussed earlier will be observed. It is shown that Eq. (9.62) can be written as follows:

$$C_{\text{MIMO}} = \sum_{i=1}^{\min(N_{\text{TX}},N_{\text{RX}})} \log_2\left(1 + \frac{\gamma\lambda_i^2}{N_{\text{TX}}}\right) \tag{9.64}$$

where $\lambda_i, i = 1, 2, \ldots,$ and $\min(N_{\text{TX}}, N_{\text{RX}})$ are the non-zero eigenvalues of $\mathbf{HH}^{\text{H}}$. We can decompose the MIMO channel into $K \leq \min(N_{\text{TX}}, N_{\text{RX}})$ equivalent

parallel SISO channels using the singular value decomposition (SVD) theorem[21] [28,29]. Let $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$ describe the input–output relationship of the MIMO channel where $\mathbf{y}$ is the output vector with N_{RX} components, $\mathbf{x}$ is the input vector with N_{TX} components, $\mathbf{n}$ is the additive noise vector with N_{RX} components, and $\mathbf{H}$ is the $N_{RX} \times N_{TX}$ channel matrix. Using the SVD theorem, we can show $\mathbf{H} = \mathbf{U\Sigma V}^H$. Let $\widehat{\mathbf{x}} = \mathbf{V}^H\mathbf{x}, \widehat{\mathbf{y}} = \mathbf{U}^H\mathbf{y}$, and $\widehat{\mathbf{n}} = \mathbf{U}^H\mathbf{n}$ denote the unitary transformation of the channel input and output and noise vectors, then it can be shown that $\widehat{\mathbf{y}} = \mathbf{\Sigma}\,\widehat{\mathbf{x}} + \widehat{\mathbf{n}}$. Since $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Sigma} = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{\min(N_{RX},N_{TX})}, 0, 0, \ldots, 0)$, it is clear that the capacity of this model is the same as the capacity of the model $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$. However, $\mathbf{\Sigma}$ is a diagonal matrix with K non-zero elements on the main diagonal, thus $\widehat{y}_1 = \lambda_1\widehat{x}_1 + \widehat{n}_1, \ldots, \widehat{y}_K = \lambda_K\widehat{x}_K + \widehat{n}_K, \widehat{y}_{K+1} = \widehat{n}_{K+1}$. The latter equations are conceptually equivalent to K parallel SISO eigen-channels, each with signal power of $\lambda_i^2, i = 1, 2, \ldots, \min(N_{TX}, N_{RX})$. Hence, the MIMO channel capacity can be rewritten in terms of the eigenvalues of the input signal covariance matrix Ξ.

When the channel knowledge is available at the transmitter and receiver, then $\mathbf{H}$ is known and we can optimize the capacity over Ξ subject to the power constraint $\text{trace}(\Xi) \leq \gamma$. It is shown in the literature that the optimal Ξ in this case exists and is known as the water-filling solution. The channel capacity in this case is given by [20,26]:

$$\begin{aligned} C &= \sum_{k=1}^{K} \log_2(\eta\lambda_i^2)^+ \\ \gamma &= \sum_{k=1}^{K} (\eta - \lambda_i^{-2})^+ \end{aligned} \tag{9.65}$$

where $(x)^+ = x \forall x \geq 0, (x)^+ = 0 \forall x < 0$ and η is a non-linear function of eigenvalues of the channel input covariance matrix. The effect of various channel conditions on the channel capacity has been extensively studied in the literature. For example, increasing the LoS signal strength at fixed SNR reduces capacity in Rician channels [20]. This can be explained in terms of the channel matrix rank

[21]The concept of decomposition of an $N \times N$ Hermitian matrix in terms of quadratic product of $N \times N$ unitary matrix composed of eigenvectors and $N \times N$ diagonal matrix of eigenvalues can be generalized to $M \times N$ complex-valued matrices of rank K. If $\mathbf{A}$ is an $M \times N$ where $(M > N)$ complex-valued matrix of rank K, then $\mathbf{A} = \mathbf{U\Sigma V}^H$ denotes the singular value decomposition of $\mathbf{A}$. The $M \times M$ unitary matrix $\mathbf{U}$ is composed of the eigenvectors of $\mathbf{AA}^H$, i.e., $\mathbf{U} = (\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_m})$, where $\mathbf{AA}^H\mathbf{u}_i = \sigma_i^2\mathbf{u}_i$. The $N \times N$ unitary matrix $\mathbf{V}$ is composed of eigenvectors of $\mathbf{A}^H\mathbf{A}$, i.e., $\mathbf{V} = (\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n})$, where $\mathbf{A}^H\mathbf{Av}_i = \sigma_i^2\mathbf{v}_i$. The elements of the $M \times N$ matrix $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$, $\mathbf{\Lambda} = \mathbf{diag}(\sigma_1, \sigma_2, \ldots, \sigma_K)$ are the square roots of the eigenvalues of matrix $\mathbf{A}^H\mathbf{A}$ which are referred to as the singular values of matrix $\mathbf{A}$. Therefore, matrix $\mathbf{A}$ may be written as $\mathbf{A} = \sum_{i=1}^{K} \sigma_i\mathbf{u}_i\mathbf{v}_i^H$. The number of non-zero singular values of matrix $\mathbf{A}$ is the rank of $\mathbf{A}$. The singular values of matrix $\mathbf{A}$ are positive real numbers which satisfy $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_K > 0$ [29].

or through various eigenvalue properties. The issue of correlated fading is of considerable importance for implementations where the antennas are required to be closely spaced. The optimal water-filling allocation strategy is obtained when the power allocated to each spatial sub-channel is non-negative. Figure 9.107 illustrates an example where the water-filling algorithm is applied to an OFDM system assuming the number of sub-channels = 16, total power = −20 dBm, $N_0 = -80$ dBm, and bandwidth = 1 MHz. It can be observed that depending on the inverse C/N of each sub-channel, the amount of power allocated to each sub-channel differs, provided that the sum of sub-channel powers does not exceed the total available power. The power allocated to each sub-channel corresponds to parameters λ_i^2.

9.13.2 Instantaneous capacity, ergodic capacity, and outage capacity of MIMO systems

In the design of wireless communication systems, the main objective is to exploit the transmission schemes whose performance can approach the channel capacity

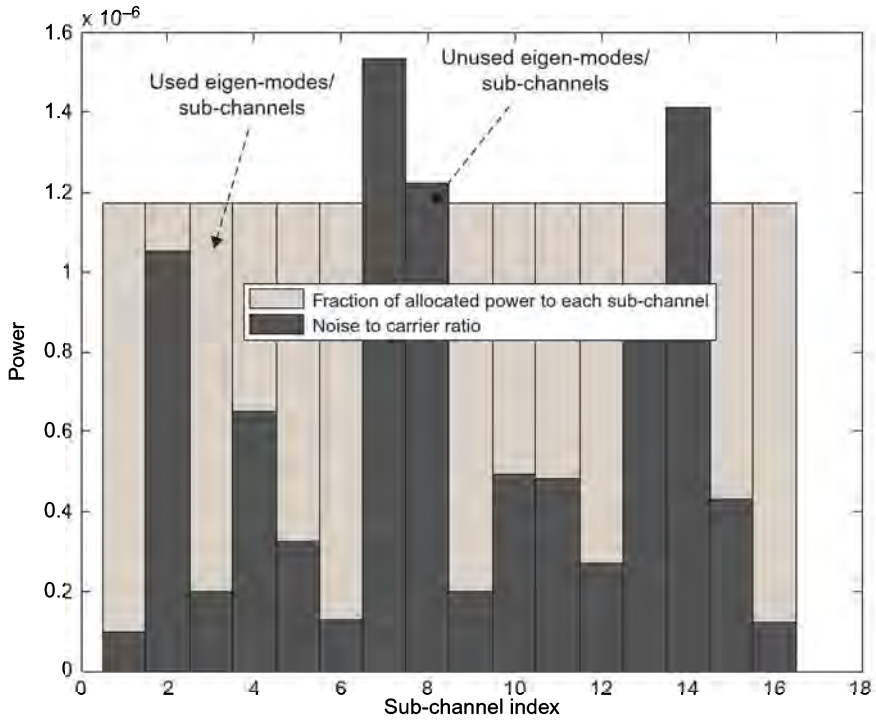


FIGURE 9.107

Example illustration of a water-filling algorithm in an OFDM system [141].

as much as possible. Therefore, it is important to understand the underlying concepts and various information theoretical definitions of channel capacity and what can be pragmatically achieved under realistic channel conditions and transceiver implementations.

Let us begin our concise study with the most generic definition of channel capacity. We denote the input and output of a memory-less SISO wireless channel with the random variables X and Y, respectively, the channel capacity is defined as [39]:

$$C = \max_{p(x)} I(X;Y) \tag{9.66}$$

where $I(X;Y)$ represents the mutual information between X and Y. Shannon's theorem [33] gives an operational meaning to the definition of the instantaneous capacity as the maximum number of bits we can transmit reliably over the channel with vanishing probability of error. The mutual information in Eq. (9.66) is maximized with respect to all possible transmit signal statistical distributions $p(x)$. Mutual information is a measure of the amount of information that one random variable contains about another variable. The mutual information between X and Y can also be written as $I(X;Y) = H(Y) - H(Y|X)$, where $H(Y|X)$ represents the conditional entropy between the random variables X and Y. The entropy of a random variable can be described as the measure of uncertainty in the random variable or the amount of information required on the average to describe the random variable. Thus, the mutual information representation of channel capacity can be described as the reduction in the uncertainty of one random variable due to the knowledge of the other. Note that the mutual information between X and Y depends on the properties of the channel through a channel matrix $\mathbf{H}$, and the properties of X through the probability distribution of X [166].

Throughout this section, it is assumed that the channel matrix $\mathbf{H}$ is random and that the receiver has perfect channel knowledge. It is also assumed that the channel is memory-less, i.e., for each use of the channel an independent realization of $\mathbf{H}$ is drawn. This means that the capacity can be computed as the maximum of the mutual information as defined earlier. The results are also valid when $\mathbf{H}$ is generated by an ergodic process, because as long as the receiver observes the $\mathbf{H}$ process, only the first order statistics are needed to determine the channel capacity.

The ergodic (mean) capacity of a random channel with $N_{RX} = N_{TX} = 1$ and an average transmit power constraint P_T can be expressed as:

$$C_{\text{ergodic}} = \mathscr{E}_{\mathbf{H}}\{\max_{p(x):P \le P_T} I(X;Y)\} \tag{9.67}$$

where P is the average power of a single codeword, transmitted over the channel, and $\mathscr{E}_{\mathbf{H}}\{.\}$ denotes the expectation over all channel realizations. Compared to the generic definition, the capacity of the channel is now defined as the maximum of the mutual information between the input and the output over all statistical

distributions on the input that satisfy the power constraint. In general, the capacity of a random MIMO channel with power constraint P_T can be expressed as [20]:

$$C_{\text{ergodic}} = \mathscr{E}_{\mathbf{H}}\{\max_{p(\mathbf{x}):\text{trace}(\mathbf{\Phi}) \le P_T} I(\mathbf{x};\mathbf{y})\} \tag{9.68}$$

where $\mathbf{\Phi} = \mathscr{E}[\mathbf{x}\mathbf{x}^H]$ is the covariance matrix of the transmit signal vector $\mathbf{x}$. The total transmit power is limited to P_T irrespective of the number of transmit antennas. For a fading channel, the channel matrix $\mathbf{H}$ is a stochastic process, and hence the associated channel capacity $C(\mathbf{H})$ is also a random variable. In this case, the ergodic channel capacity is defined as the average of instantaneous channel capacity over the distribution of $\mathbf{H}$. The ergodic channel capacity of the MIMO transmission scheme is given by:

$$C_{\text{ergodic}} = \mathscr{E}\left\{\max_{\text{trace}(\mathbf{\Phi}) = N_{TX}} \log_2\left(\det\left[\mathbf{I} + \frac{\gamma}{N_{TX}}\mathbf{H}\mathbf{\Phi}\mathbf{H}^H\right]\right)\right\} \tag{9.69}$$

where $\mathbf{\Phi}$ denotes the $N_{TX} \times N_{TX}$ covariance matrix of the channel input vector $\mathbf{x}$. According to information theoretic concepts, this capacity cannot be achieved unless channel coding is employed across an infinite number of independently fading blocks. Let us focus on the case of perfect CSI at the receiver side and no CSI at the transmitter side. Obviously, this implies that the maximization of the latter equation is now more restricted than in the previous case. Nevertheless, it has been shown in the literature that the optimal signal covariance matrix has to be chosen according to $\mathbf{\Phi} = \mathbf{I}$. This means that the antennas should transmit uncorrelated streams with the same average power. With this result, the ergodic MIMO channel capacity reduces to [20,24]:

$$C_{\text{ergodic}} = \mathscr{E}\left\{\log_2\left(\det\left[\mathbf{I} + \frac{\gamma}{N_{TX}}\mathbf{H}\mathbf{H}^H\right]\right)\right\} \tag{9.70}$$

Clearly, this is not the Shannon capacity in a true sense, since as mentioned before and with perfect channel knowledge at the transmitter one can choose a signal covariance matrix that outperforms $\mathbf{\Phi} = \mathbf{I}$ case. Nevertheless, we refer to the expression in Eq. (9.70) as the ergodic channel capacity with CSI at the receiver and no CSI at the transmitter.

The capacity under channel ergodicity was defined as the average of the maximal value of the mutual information between the transmitted and the received signals, where the maximization was carried out with respect to all possible transmit signal statistical distributions. Another measure of channel capacity that is frequently used is outage capacity. With outage capacity, the channel capacity is associated to an outage probability. Capacity is treated as a random variable which depends on the channel instantaneous response and remains constant during the transmission of a finite-length coded block of information. If the channel capacity falls below the outage capacity, there is no possibility that the transmitted block of information can be decoded with no errors, no matter which coding scheme is employed. The probability that the capacity is less than the outage

capacity denoted by C_{outage} is ρ. This can be expressed in mathematical terms by $p(C < C_{\text{outage}}) = \rho$. In this case, the latter expression represents an upper bound since there is a finite probability ρ that the channel capacity is less than the outage capacity. It can also be written as a lower bound, representing the case where there is a finite probability $(1 - \rho)$ that the channel capacity is higher than C_{outage}, which means $p(C > C_{\text{outage}}) = 1 - \rho$. In other words, since the MIMO instantaneous channel capacity is a random variable, it is meaningful to consider its statistical distribution. A useful measure of its statistical behavior is the outage capacity. Outage analysis quantifies the level of performance (in this case capacity) that is guaranteed with a certain level of reliability. The ρ% outage capacity $C_{\text{outage}}(\rho)$ is defined as the information rate that is guaranteed for $(100 - \rho)$% of the channel realizations ($p(C < C_{\text{outage}}) = \rho$). The outage capacity is often a more relevant measure than the ergodic channel capacity, because it describes in some way the quality of the channel. This is due to the fact that the outage capacity measures how the instantaneous rate supported by the channel is distributed, in terms of probability. Thus, if the rate supported by the channel is spread over a wide range, the outage capacity for a fixed probability level may become small, whereas the ergodic channel capacity may be high [166].

9.13.3 Spatial multiplexing and diversity

The main objective of space–time coding schemes is to exploit the inherent spatial (multipath) diversity in MIMO channels through properly designed space–time codes. In this section, two transmit diversity schemes, namely the Alamouti scheme and CDD, are described where both realize full spatial diversity of the channel without requiring channel knowledge at the transmitter. In the Alamouti scheme, we consider a MIMO channel with two transmit antennas and any number of receive antennas. The Alamouti scheme can be described as follows. Two different complex-valued data symbols s_1 and s_2 are transmitted simultaneously from antenna ports 1 and 2, respectively, during the first symbol period. In the second symbol period, symbols $-s_2^*$ and s_1^* are sent from antenna ports 1 and 2, respectively. It must be noted that the rate equals one in the Alamouti scheme since two independent data symbols are transmitted over two symbol periods. Let us consider the following case with two transmit antennas and one receive antenna, thereby we have two channels H_1 and H_2. The received signal at the first and second symbol period can be expressed as follows [58]:

$$\begin{aligned} y_1 &= H_1 s_1 + H_2 s_2 + n_1 \\ y_2 &= -H_1 s_2^* + H_2 s_1^* + n_2 \end{aligned} \tag{9.71}$$

Let $\mathbf{r} = (y_1, y_2^*)^{\text{H}}, \mathbf{s} = (s_1, s_2)^{\text{H}}$, $\mathbf{n} = (n_1, n_2)^{\text{H}}$, and $\mathbf{H} = \begin{bmatrix} H_1 & H_2 \\ H_2^* & -H_1^* \end{bmatrix}$, the relationship between the output and the input of the channel can be written as $\mathbf{r} = \mathbf{Hs} + \mathbf{n}$ where multiplying both sides of the latter equation by $\mathbf{H}^{\text{H}}$ yields

$\tilde{\mathbf{r}} = \mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{s} + \tilde{\mathbf{n}}$. Since $\mathbf{H}$ is orthogonal, i.e., $\mathbf{H}^{\mathrm{H}}\mathbf{H} = \alpha\mathbf{I}$ where $\alpha = |H_1|^2 + |H_2|^2$; dividing both sides of the equation by α yields [58]:

$$\begin{aligned}\frac{\tilde{r}_1}{|H_1|^2 + |H_2|^2} &= s_1 + \frac{H_1^* n_1}{|H_1|^2 + |H_2|^2} + \frac{H_2 n_2}{|H_1|^2 + |H_2|^2} \\ \frac{\tilde{r}_2}{|H_1|^2 + |H_2|^2} &= s_2 + \frac{H_2^* n_1}{|H_1|^2 + |H_2|^2} + \frac{-H_1 n_2}{|H_1|^2 + |H_2|^2}\end{aligned} \tag{9.72}$$

It is assumed that the channels are independent and identically distributed with frequency-flat fading and remain constant over (at least) two consecutive symbol periods. Appropriate processing at the receiver, as shown in Eqs. (9.71) and (9.72), collapses the vector channel into a scalar channel for either of the transmitted data symbols, where $\tilde{\mathbf{r}}$ is the processed received signal corresponding to transmitted symbol $\mathbf{s}$, and $\tilde{\mathbf{n}}$ is the processed noise. Even though the CSI knowledge is not available at the transmitter, the Alamouti scheme achieves $2N_{\mathrm{RX}}$ order diversity. We note that array gain is realized only at the receiver since the transmitter does not have the CSI. The Alamouti scheme may be extended to channels with more than two transmit antennas through orthogonal space–time block coding. The link-level performance of the Alamouti scheme is compared to the SISO and SIMO cases in Figure 9.108 assuming BPSK-modulated input symbols and Rayleigh fading channel.

Delay diversity is an alternative method in which the same symbol is transmitted at different time instances such that each symbol experiences different fading. The delay diversity converts spatial diversity to frequency diversity by transmitting the data signal from the first antenna port and a delayed replica of the signal from the second antenna port. Assuming two transmit antennas and one receive antenna and that the delay induced by the second antenna equals one symbol period, the effective channel seen by the data signal is a frequency-selective fading SISO channel with impulse response $h(k) = h_1\delta(k) + h_2\delta(k-1)$, where h_1 and h_2 are complex-valued flat-fading channels. We note that the effective channel resembles a two-path SISO channel with independent fading paths and equal average path energy. An ML detector will realize full second-order diversity at the receiver. If the same signal is transmitted from two or more transmit antennas, a beamforming effect is created where in some directions the signal is attenuated, while in other directions it may be amplified. The CDD scheme mitigates this problem by transmitting cyclically shifted signals from the transmit antennas. Let us consider an OFDM symbol with constellation d_k at the kth frequency index, the time-domain signal corresponding to this symbol which is transmitted from antenna port 1 can be expressed as follows:

$$s_1(t) = \mathrm{Re}\left\{ \mathrm{e}^{\mathrm{j}\omega_c t} \sum_{k=-(N_{\mathrm{FFT}}-1)/2}^{(N_{\mathrm{FFT}}-1)/2} d_k \, \mathrm{e}^{\mathrm{j}2\pi k(t-t_g)/T_u} \right\} \quad mT_u \le t \le (m+1)T_u \tag{9.73}$$

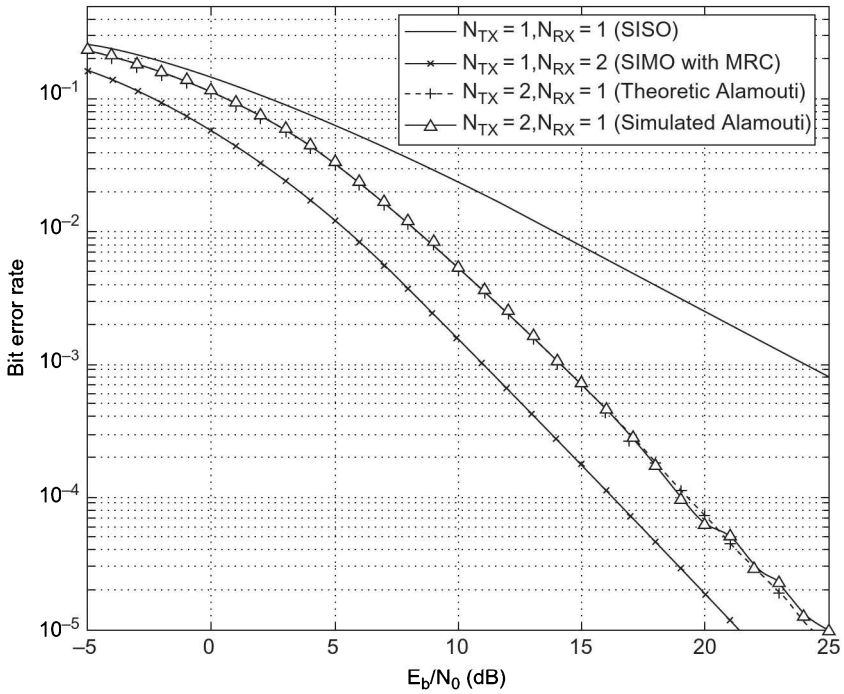


FIGURE 9.108

Performance of the Alamouti scheme [141].

Using the CDD scheme, the transmitted signal from the second antenna port is given by:

$$s_2(t) = \mathrm{Re}\left\{ e^{j\omega_c t} \sum_{k=-(N_{\mathrm{FFT}}-1)/2}^{(N_{\mathrm{FFT}}-1)/2} d_k \, e^{j2\pi k(t-t_g-T_{CDD})/T_u} \right\} \quad mT_u \le t \le (m+1)T_u \quad (9.74)$$

This is equivalent to a cyclic rotation of the signal in the time-domain. The implementation of the CDD scheme is simplified, if the cyclic shift T_{CDD} is an integer multiple of the sampling time. At the receiver, the effect of CDD at the kth index is of receiving through the channel $H_1(k) + \exp(j2\pi k T_{\mathrm{CDD}}/T_u)H_2(k)$. The advantage over the case of $H_1(k) + H_2(k)$, which is the result of transmission of the same signal over the two channels, is the prevention of the beamforming effect, i.e., directed transmission in one direction and undesired emission in the other directions.

The main objective of spatial multiplexing, as opposed to space–time diversity coding, is to maximize transmission data rate. Therefore, $\min(N_{\mathrm{TX}}, N_{\mathrm{RX}})$ independent data symbols are transmitted per symbol period such that the spatial rate is $\min(N_{\mathrm{TX}}, N_{\mathrm{RX}})$. There are several encoding options that can be used in

conjunction with spatial multiplexing to achieve the increased data rate. Single codeword and multi-codeword transmission are the prominent encoding options that are typically used in wireless communication systems. In a multi-codeword scheme, the bit stream to be transmitted is de-multiplexed into separate data streams where each stream is processed through the transmission chain, i.e., independent temporal encoding, symbol mapping, and interleaving, and is then transmitted from the corresponding antennas. The antenna to stream mapping remains static over time. The spatial rate is $\min(N_{TX}, N_{RX})$ and the overall signaling rate is $n_{\text{modulation-order}} r_b \min(N_{TX}, N_{RX})$, where r_b denotes the channel coder input bit rate and $n_{\text{modulation-order}}$ denotes the QAM modulation order. The multi-codeword scheme can at most achieve N_{RX} order diversity, since any given information bit is transmitted from only one transmit antenna and received by N_{RX} receive antennas. While this is a source of sub-optimality of the multi-codeword architecture, it does simplify receiver design. The coding gain achieved by the multi-codeword scheme depends on the coding gain of the temporal code. Furthermore, a maximum array gain of N_{RX} can be realized using the multi-codeword scheme.

In the single-codeword architecture, the bit stream undergoes temporal encoding, symbol mapping, and interleaving after which it is de-multiplexed into $\min(N_{TX}, N_{RX})$ streams transmitted from the individual antennas. This form of encoding can achieve full $N_{TX} \times N_{RX}$ order diversity, provided that the temporal code is designed properly, since each information bit can be spread across all the transmit antennas. Nevertheless, the single-codeword scheme requires joint decoding of the sub-streams, which increases receiver implementation complexity compared to the multi-codeword scheme where the individual data streams can be decoded separately. The spatial rate of the single-codeword is $\min(N_{TX}, N_{RX})$, and the overall signaling rate is $n_{\text{modulation-order}} r_b \min(N_{TX}, N_{RX})$. The coding gain achieved by the single-codeword scheme will depend on the temporal channel coding characteristic and a maximum array gain of N_{RX} can be achieved [24].

Space-frequency block coding is the frequency-domain version of the space–time block codes (STBC). This family of codes is designed to achieve diversity gain by transmitting orthogonal streams which yields optimal SNR with a linear receiver. Since, in the OFDM systems, the number of OFDM symbols in a sub-frame may often be an odd number, while STBC (e.g., Alamouti code) operates on pairs of adjacent symbols in the time-domain, the application of STBC would not be straightforward. For SFBC transmission, the adjacent sub-carriers on each OFDM symbol are paired, encoded, and transmitted from two base station antenna ports where the encoding is defined as follows:

$$\mathbf{W} = \begin{bmatrix} x_k & x_{k+1} \\ -x_{k+1}^* & x_k^* \end{bmatrix} \tag{9.75}$$

where $(x_k, -x_{k+1}^*)$ and (x_{k+1}, x_k^*) are transmitted from antenna ports 1 and 2, respectively. Note that the spatial rate of the SFBC codes in this case is one.

9.13.4 MIMO receivers

An orthogonal space–time block coding scheme transforms a vector-detection problem into a number of scalar-detection problems. Similar concepts can be applied to frequency-selective fading MIMO channels, inspiring receiver design techniques such as zero-forcing (ZF), MMSE estimation, and (optimal) ML sequence detection [20]. The use of transmit diversity techniques such as CDD or frequency-offset diversity transforms a MISO channel into a number of SISO channels, and allows the application of SISO receiver architectures. For a general space–time trellis code, a vector Viterbi decoder can be employed. The space–time trellis coding in general provides improved performance over orthogonal space–time block coding at the expense of receiver complexity. The problem encountered by a receiver for spatial multiplexing is the presence of multi-stream interference, since the signals sent from different transmit antennas interfere with each other. Note that in spatial multiplexing, different data streams are transmitted in the same channel and occupy the same resources in time and frequency. For the sake of simplicity, we restrict our attention to the case where $N_{RX} \geq N_{TX}$.

9.13.4.1 Maximum likelihood receiver

The optimal detection for spatially multiplexed signals is to use maximum likelihood criterion. The ML receiver performs vector decoding and is optimal in the sense of minimizing the error probability. Assuming equally likely, temporally uncoded vector symbols, the ML receiver generates an estimate of the transmitted signal vector as follows:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \left\| \mathbf{y} - \sqrt{\frac{E_s}{N_{TX}}}\mathbf{H}\mathbf{s} \right\|^2 \tag{9.76}$$

where the minimization is performed over all possible transmit vector symbols $\mathbf{s}$. Let Z denote the alphabet size of the scalar constellation transmitted from each antenna, a complete implementation requires an exhaustive search over a total of $Z^{N_{TX}}$ vector symbols or hypotheses, making the decoding complexity of this receiver exponentially grow with the number of transmit antennas. Recent development of fast ML detection algorithms reduces the computational complexity of the ML decoder. The ML receiver realizes N_{RX} diversity order for the multi-codeword scheme and $N_{RX} \times N_{TX}$ diversity order for the single-codeword scheme [20,25].

9.13.4.2 Successive interference cancelation receiver

An alternative non-linear approach to detection of spatially multiplexed signals is successive interference cancelation (SIC) assuming that the spatially multiplexed signals are separately coded (i.e., multi-codeword transmission). As shown in Figure 9.109, in an SIC receiver, one of the spatially multiplexed signals is initially demodulated and decoded. Upon successful decoding, the data is

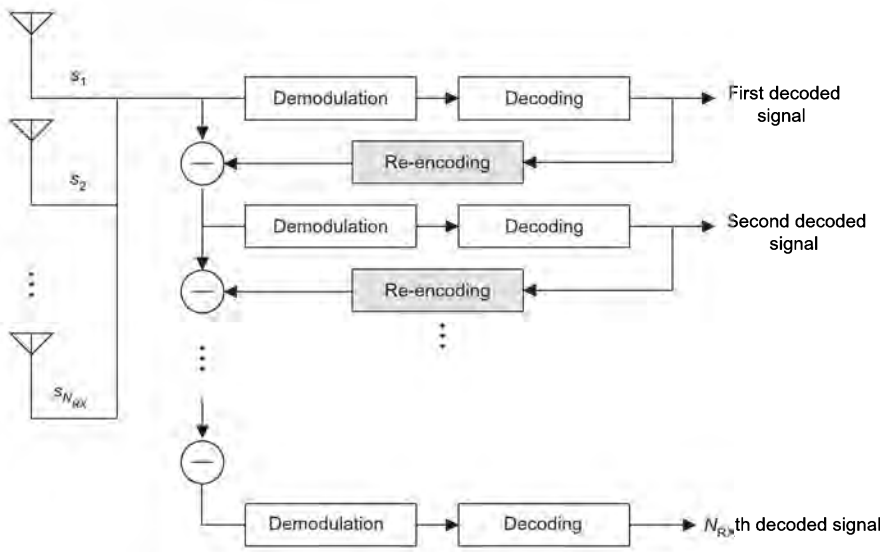


FIGURE 9.109

Structure of a successive interference cancelation MIMO receiver.

re-encoded and subtracted from the received signals. The second spatially multiplexed signal can then be demodulated and decoded in the absence of interference from the first detected signal. The subtraction would ideally improve the signal-to-interference ratio. The procedure is iterated until all spatially multiplexed signals are demodulated and decoded. It is noted that the first signals to be decoded using this approach are subject to higher interference levels compared to those that are decoded later in the process. Therefore, the ith signal to be decoded must be more robustly coded than the $(i+1)$th signal. This can be achieved in multi-codeword transmission by applying different modulation and coding schemes to different spatially multiplexed signals. The latter scheme is often referred to as per-antenna rate control (PARC) [20,25,42].

9.13.4.3 Zero-forcing receiver

The decoding complexity of the ML receiver can be reduced using a linear filter to separate the transmitted data streams and then independently decode each stream. The ZF and MMSE linear detectors are examples of such receivers. The ZF weighting matrix (filter) that decomposes the received signal into its component streams is given by the following equation [20,42]:

$$\mathbf{G}_{\mathrm{ZF}} = \sqrt{\frac{N_{\mathrm{TX}}}{E_{\mathrm{s}}}}\mathbf{H}^{\dagger} = \sqrt{\frac{N_{\mathrm{TX}}}{E_{\mathrm{s}}}}(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{H}} \tag{9.77}$$

where $\mathbf{G}_{ZF}$ is an $N_{TX} \times N_{RX}$ matrix that inverts the channel matrix and $\mathbf{H}^{\dagger}$ denotes the Moore–Penrose inverse[22] of the channel matrix $\mathbf{H}$. The output of the ZF receiver is obtained as follows:

$$\mathbf{y} = \mathbf{s} + \sqrt{\frac{N_{TX}}{E_s}}\mathbf{H}^{\dagger}\mathbf{n} \tag{9.78}$$

which shows that the ZF receiver decomposes the channel matrix into parallel scalar channels with additive spatially colored noise. Each scalar channel is then decoded independently irrespective of noise correlation across the processed streams. The ZF receiver transforms the joint decoding problem into single stream decoding problems, thereby significantly reducing receiver complexity. This complexity reduction is achieved at the expense of noise enhancement, which in general results in a significant performance degradation (compared to the ML decoder). The diversity order achieved by each of the individual data streams equals $N_{RX} - N_{TX} + 1$. Figure 9.110 illustrates the link-level performance of a 2×2 MIMO system with a ZF receiver. It is shown that the performance is close to that of a SISO system, since the diversity order is one.

9.13.4.4 Minimum mean square error receiver

The MMSE receiver balances multi-stream interference mitigation and noise enhancement effect, and minimizes the total error. The MMSE weighting matrix (filter) is given by the following expression [20,42]:

$$\mathbf{G}_{MMSE} = \arg\min_{\mathbf{G}} \mathscr{E}\left\{ ||\mathbf{G}\mathbf{y} - \mathbf{s}||^2 \right\} = \sqrt{\frac{N_{TX}}{E_s}}\left(\mathbf{H}^{H}\mathbf{H} + \frac{N_{TX}N_0}{E_s}\mathbf{I} \right)^{-1} \mathbf{H}^{H} \tag{9.79}$$

where $\mathbf{I}$ is an $N_{RX} \times N_{RX}$ identity matrix. In low signal-to-interference plus noise ratios (i.e., SINR < 0 dB), the performance of the MMSE receiver approaches that of a matched-filter receiver, outperforming the ZF detector which tends to enhance noise. In high signal-to-interference plus noise ratios, however, the performance of the MMSE receiver approaches that of a ZF receiver and therefore realizes $N_{RX} - N_{TX} + 1$ order diversity for each data stream.

---

[22] Given a matrix $\mathbf{A}$ and vector $\mathbf{b}$ and equation $\mathbf{A}\mathbf{x} = \mathbf{b}$, we find the vector $\mathbf{x}$ that minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$. Using the orthogonality principle, the error is taken perpendicular to estimation vector such that $\mathbf{x}^{H}\mathbf{A}^{H}(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0$ which implies that $\mathbf{A}^{H}\mathbf{A}\mathbf{x} = \mathbf{A}^{H}\mathbf{b}$ (since $\mathbf{x} \neq 0$), or $\mathbf{x} = (\mathbf{A}^{H}\mathbf{A})^{-1}\mathbf{A}^{H}\mathbf{b}$. One should note that $\mathbf{A}\mathbf{x} = \mathbf{A}(\mathbf{A}^{H}\mathbf{A})^{-1}\mathbf{A}^{H}\mathbf{b}$. The matrix $\mathbf{A}^{\dagger} = (\mathbf{A}^{H}\mathbf{A})^{-1}\mathbf{A}^{H}$ is defined as the Moore–Penrose inverse of the matrix $\mathbf{A}$. If $\mathbf{A}^{H}\mathbf{A}$ is not invertible, the SVD algorithm is used to calculate the matrix pseudo-inverse. In that case, we may write $\mathbf{A} = \mathbf{V}_K\mathbf{\Sigma}_K\mathbf{U}_K^{H}$ and seek $\mathbf{x}$ such that $||\mathbf{V}_K\mathbf{\Sigma}_K\mathbf{U}_K^{H}\mathbf{x} - \mathbf{b}||$ is minimized. Using the orthogonality principle, we have $\mathbf{x}^{H}\mathbf{U}_K\mathbf{\Sigma}_K\mathbf{V}_K^{H}(\mathbf{V}_K\mathbf{\Sigma}_K\mathbf{U}_K^{H}\mathbf{x} - \mathbf{b})$ or (assuming $\mathbf{x} \neq 0$) $\mathbf{U}_K\mathbf{\Sigma}_K^2\mathbf{U}_K^{H}\mathbf{x} - \mathbf{U}_K\mathbf{\Sigma}_K\mathbf{V}_K^{H}\mathbf{b} = 0$ such that $\mathbf{\Sigma}_K^2\mathbf{U}_K^{H}\mathbf{x} - \mathbf{\Sigma}_K\mathbf{V}_K^{H}\mathbf{b} = 0$, multiplying both sides by $\mathbf{\Sigma}_K^{-2}$, yields $\mathbf{U}_K^{H}\mathbf{x} - \mathbf{\Sigma}_K^{-1}\mathbf{V}_K^{H}\mathbf{b} = 0$. The latter is true when $\mathbf{x} = \mathbf{\Sigma}_{i=1}^{K}\left(\frac{1}{\sigma_i}\mathbf{v}_i^{H}\mathbf{b}\right)\mathbf{u}_i$; therefore $\mathbf{x} = \mathbf{U}_K\mathbf{\Sigma}_K^{-1}\mathbf{V}_K^{H}\mathbf{b}$ [29].

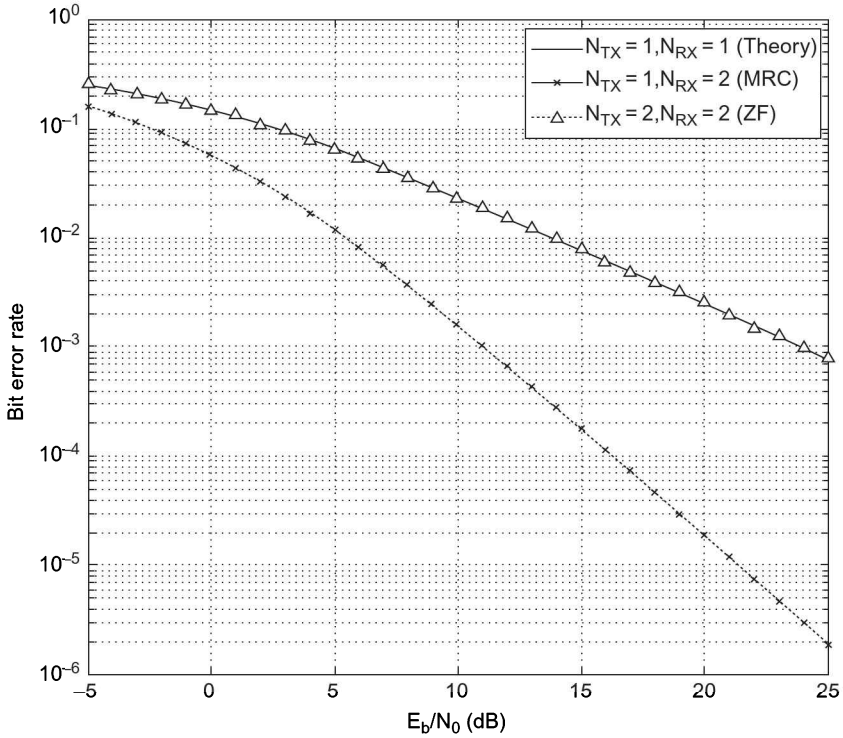


FIGURE 9.110

Link-level performance of 2 × 2 MIMO system with ZF equalizer (Rayleigh channel and BPSK modulation).

9.13.4.5 Maximal ratio combining receiver

In maximal ratio combining (MRC), the output is a weighted sum of all receiver branches, thus the coefficients $\alpha_i, i = 1, 2, \ldots,$ and N_{RX} are all non-zero. The signals are co-phased and $\alpha_i = \beta_i \, e^{-j\theta_i}, i = 1, 2, \ldots, N_{RX}$ where θ_i is the phase of the received signal on the ith branch (Figure 9.111). The magnitude of the combined output can be written as $y = \sum_{i=1}^{N_{RX}} \beta_i r_i$. Let us assume that the noise power spectral density is $N_0/2$ in each branch, which yields the combined noise at the output $n = \sum_{i=1}^{N_{RX}} \beta_i^2 N_0/2$. The SNR at the output of the combiner is given as:

$$\gamma = \frac{1}{N_0} \frac{\left(\sum_{i=1}^{N_{RX}} \beta_i r_i\right)^2}{\sum_{i=1}^{N_{RX}} \beta_i^2} \tag{9.80}$$

The objective is to find $\alpha_i, i = 1, 2, \ldots, N_{RX}$ which maximize the output SNR. It can be understood that branches with higher SNR are weighted more than those

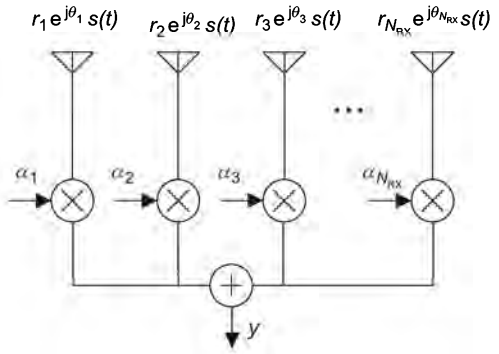


FIGURE 9.111

A generic linear combining receiver.

with lower SNR; therefore, the coefficients $\alpha_i, i = 1, 2, \ldots,$ and N_{RX} are proportional to the branch SNRs. The coefficients are found by taking partial derivatives of the output SNR expression and solving the resulting equations for the optimal weights which yields $\alpha_i = r_i^2/N_0, i = 1, 2, \ldots, N_{\mathrm{RX}}$ and the output SNR will be the sum of the SNRs of the receiving branches. Therefore, the output SNR and the corresponding array gain increase linearly with the number of the diversity branches (i.e., the number of receive antennas).

9.13.4.6 MMSE-interference rejection combining receiver

The interference rejection combining (IRC) receiver is an effective method for improving the cell-edge user throughput, since it attempts to suppress the inter-cell interference. The IRC receiver is typically based on the minimum mean square error criteria, which requires channel estimation and covariance matrix estimation including the inter-cell interference with high accuracy. The gain from the IRC receiver is tied to the estimation of the interference signal, i.e., the covariance matrix, in terms of the downlink user throughput performance in a multi-cell environment. We define the $N_{\mathrm{RX}} \times 1$ received signal vector $\mathbf{y}(k, l)$ corresponding to the kth sub-carrier and the lth OFDM symbol as follows: $\mathbf{y}(k, l) = \sum_{i=0}^{N_{cell}-1} \mathbf{H}_i(k, l)\mathbf{W}_{\mathrm{TX}}^i(k, l)\mathbf{s}_i(k, l) + \mathbf{n}(k, l)$, where $\mathbf{H}_i(k, l)$ represents the $N_{\mathrm{RX}} \times N_{\mathrm{TX}}$ channel matrix between the ith cell and the UE, $\mathbf{W}_{\mathrm{TX}}^i(k, l)$ represents the $N_{\mathrm{TX}} \times N_l$ precoding matrix of the ith cell, $\mathbf{s}_i(k, l)$ denotes the $N_l \times 1$ signal vector of the ith cell, and $\mathbf{n}(k, l)$ is the $N_{\mathrm{RX}} \times 1$ noise vector with variance σ^2. In this context, $N_{\mathrm{TX}}, N_{\mathrm{RX}}, N_l$, and N_{cell} are the number of transmit antennas of each cell, the number of receive antennas at the UE, the number of transmission layers for the UE, i.e., transmission ranks and total number of cells in the model, respectively. The zeroth cell is assumed to be the serving cell for the UE.

The received signal vector at the UE denoted by $\hat{\mathbf{s}}_0(k,l)$ is detected using a $N_l \times N_{\mathrm{RX}}$ weighting matrix $\mathbf{W}^0_{\mathrm{RX}}(k,l)$ expressed as $\hat{\mathbf{s}}_0(k,l) = \mathbf{W}^0_{\mathrm{RX}}(k,l)\mathbf{y}(k,l)$. Based on MMSE criterion, the weighting matrix $\mathbf{W}^0_{\mathrm{RX}}(k,l)$ is obtained by $\mathbf{W}^0_{\mathrm{RX}}(k,l) = P_0\overline{\mathbf{G}}^H_0(k,l)\overline{\mathbf{R}}^{-1}_{yy}$ and $\overline{\mathbf{G}}_0(k,l) = \overline{\mathbf{H}}_0(k,l)\mathbf{W}^0_{\mathrm{TX}}$ where P_0, $\overline{\mathbf{G}}_0$, $\overline{\mathbf{H}}_0$, and $\overline{\mathbf{R}}_{yy}$ denote the transmission power of the serving cell, the composite channel from the serving cell, the $N_{\mathrm{RX}} \times N_{\mathrm{TX}}$ channel matrix of the serving cell, and the covariance matrix of the received signal, respectively [180].

The baseline MMSE receiver is defined as a receiver that only suppresses the inter-layer interference, i.e., separates multiple data streams that achieve the maximum SINR for each data stream within a cell. The inter-cell interference is assumed to be AWGN at each receive antenna, making the covariance matrix a diagonal matrix. In the special case of rank-1 transmission, the MMSE receiver is equivalent to the MRC receiver. When a composite channel from the serving cell is obtained, the covariance matrix without the inter-cell interference can be represented by the composite channel from the serving cell, and the interference power from other cells and noise power. Therefore, the respective MMSE weighting matrices with ideal and realistic channel estimation are obtained as follows [180]:

$$\begin{aligned}
\mathbf{W}^{\mathrm{Ideal}}_{\mathrm{MMSE}}(k,l) &= P_0\mathbf{G}^H_0(k,l)[P_0\mathbf{G}_0(k,l)\mathbf{G}^H_0(k,l) + \boldsymbol{\Omega} + \sigma^2_N\mathbf{I}]^{-1} \\
\boldsymbol{\Omega} &= diag(\sigma^2_{I_0}, \sigma^2_{I_1}, \ldots, \sigma^2_{I_{N_{\mathrm{RX}}}}) \\
\mathbf{W}^{\mathrm{Realistic}}_{\mathrm{MMSE}}(k,l) &= P_0\hat{\mathbf{G}}^H_0(k,l)(P_0\hat{\mathbf{G}}_0(k,l)\hat{\mathbf{G}}^H_0(k,l) + \boldsymbol{\Omega} + \sigma^2_N\mathbf{I})^{-1}
\end{aligned} \tag{9.81}$$

where $\mathbf{G}_0(k,l)$, $\hat{\mathbf{G}}_0(k,l)$, $\sigma^2_{I_j}$, and σ^2_{N} denote the ideal composite channel matrix, the estimated composite channel matrix based on the DM-RSs, which are precoded in the same manner as the data signals, the interference power from other cells at the jth receive antenna, and the noise power, respectively [180].

The MMSE-IRC receiver can suppress not only the inter-layer interference, but also the inter-cell interference when the degrees of freedom at the receiver are high, i.e., the number of receive antennas is higher than that of the desired transmission rank. If the composite channels from N_{cell} cells are perfectly known at the receiver, the ideal IRC weighting matrix is obtained as follows:

$$\mathbf{W}^{\mathrm{Ideal}}_{\mathrm{IRC}}(k,l) = P_0\mathbf{G}^H_0(k,l)\left(\sum_{i=0}^{N_{\mathrm{cell}}-1} P_i\mathbf{G}_i(k,l)\mathbf{G}^H_i(k,l) + \sigma^2_{\mathrm{N}}\mathbf{I}\right)^{-1} \tag{9.82}$$

However, it is assumed that the composite channel matrices from other $N_{\mathrm{cell}} - 1$ cells cannot be estimated at the receiver; therefore, the covariance matrix including the inter-cell interference is obtained by statistical averaging of the receiver signal, yielding:

$$\mathbf{R}_{yy} = \mathscr{E}[\mathbf{y}(k,l)\mathbf{y}^{\mathrm{H}}(k,l)] \tag{9.83}$$

The resulting IRC weighting matrices in the case of ideal and realistic channel estimations can be obtained as follows:

$$\begin{aligned}\mathbf{W}_{\text{IRC}}^{\text{Ideal}}(k,l) &= P_0\mathbf{G}_0^H(k,l)(\mathscr{E}[\mathbf{y}(k,l)\mathbf{y}^{\text{H}}(k,l)])^{-1}\\ \mathbf{W}_{\text{IRC}}^{\text{Realistic}}(k,l) &= P_0\hat{\mathbf{G}}_0^H(k,l)(\mathscr{E}[\mathbf{y}(k,l)\mathbf{y}^{\text{H}}(k,l)])^{-1}\end{aligned} \tag{9.84}$$

The accuracy of the IRC weighting depends on the accuracy of the channel estimation and covariance matrix. Therefore, in order to obtain an accurate covariance matrix, the covariance matrix must be averaged using the received samples which have gone through the same precoding matrix at the transmitter side and the same channel matrix with high correlation. More specifically, in order to ensure the use of the same precoding matrix, one can assume that the covariance matrix is averaged over one resource block using all samples except for the control signaling, the CRSs, and the CSI-RSs. Note that LTE Rel-10 allows the use of different precoding matrices for different resource blocks. Furthermore, it must be noted that some signals, such as control signaling, CRSs, and CSI-RSs, are transmitted without precoding [180].

The time-domain averaging is given more priority over frequency-domain averaging within a resource block because the channel variation within a subframe is smaller than that within a resource block in the frequency-domain. Two averaging methods, i.e., time-domain only, and time- and frequency-domain averaging are performed where the resulting covariance matrices are calculated as follows [180]:

$$\begin{aligned}\mathbf{R}_{yy}^{\text{time}} &= \frac{1}{N_{\text{time}}}\sum_{l}\mathbf{y}(k,l)\mathbf{y}^{\text{H}}(k,l)\\ \mathbf{R}_{yy}^{\text{time-frequency}} &= \frac{1}{N_{\text{time-frequency}}}\sum_{k}\sum_{l}\mathbf{y}(k,l)\mathbf{y}^{\text{H}}(k,l)\end{aligned} \tag{9.85}$$

where N_{time} and $N_{\text{time-frequency}}$ denote the average number of samples for the time-domain and time- and frequency-domain averaging, respectively [180].

9.13.4.7 Sphere decoding

In its search for the ML solution, the sphere decoder (SD) evaluates all transmit signal vectors $\mathbf{x}$ fulfilling the following criterion: $||\mathbf{y}-\mathbf{H}\mathbf{x}||^2 < R^2$, where R is the search radius of a hyper-sphere. It is obvious that the selection of R is a critical issue largely influencing the complexity of any SD algorithm. Choosing R too large would lead to a sphere containing a large number of hypotheses (also referred to as candidates), leading to high detection complexity. On the other hand, choosing R too small would result in an empty sphere and the search having to be restarted with an increased radius, eventually leading to similar problems. The search for candidates fulfilling the above inequality is done by back-substitution algorithms.

Toward this goal, the Cholesky[23] or QR[24] decompositions of the channel matrix $\mathbf{H}$ may be equivalently used. Note that using a Cholesky decomposition may be advantageous in systems with more transmit than receive antennas, since the size of the upper triangular matrix $\mathbf{R}$ is limited by $\min(N_{RX}, N_{TX})$, which for $N_{RX} < N_{TX}$ leads to an under-determined problem for the QR implementation of the SD [181].

In the following, we focus on the symmetric case, i.e., $N_{RX} = N_{TX}$, and use a QR decomposition of $\mathbf{H}$, for practicality of implementation. With $\mathbf{H} = \mathbf{QR}$, where $\mathbf{R}$ is upper-triangular and $\mathbf{Q}$ is unitary, the above inequality may be rewritten as follows:

$$||\mathbf{y}-\mathbf{Hx}||^2 < R^2 \Rightarrow ||\mathbf{Q}^H\mathbf{y}-\mathbf{Rx}||^2 < R^2 \Rightarrow ||\mathbf{y}'-\mathbf{Rx}||^2 < R^2 \tag{9.86}$$

The above expression can be written as follows:

$$\sum_{i=l}^{N_{TX}} \left| y_i' - \sum_{j=i}^{N_{TX}} r_{ij}x_j \right|^2 < R^2 \quad l = 1, 2, \ldots, N_{TX} \tag{9.87}$$

The detection process starts with the last layer $l = N_{TX}$, for which the latter expression reduces to $|y'_{N_{TX}} - r_{N_{TX}N_{TX}}x_{N_{TX}}|^2 < R^2$ and continues until the first layer is detected. This process is quite similar to SIC techniques, i.e., the signals from previously detected layers are subtracted from the received signal before detection within the current layer is performed. However, in SD, detection at the ith layer essentially takes on the following form:

$$|c_{ki} - x_i|^2 < \frac{R_{ki}^2}{r_{ii}^2} \tag{9.88}$$

where c_{ki} is the search center and R_{ki} is the remainder search radius for the currently considered (incomplete) candidate c. Both SIC and SD methods depend on the estimates for previously detected layers, and hence vary from one candidate to another. The above inequality implies that not only a single, but several constellation points may be selected. The SD receiver hence performs its search in a tree like structure, which motivates the search for appropriate enumeration

[23]Given a symmetric positive-definite matrix $\mathbf{H}$, the Cholesky decomposition is an upper triangular matrix $\mathbf{U}$ with strictly positive diagonal entries such that $\mathbf{H} = \mathbf{U}^H\mathbf{U}$ [28,29]. Alternatively, if $\mathbf{H}$ has real entries and is symmetric (or more generally, has complex-valued entries and is Hermitian) and positive-definite, then $\mathbf{H}$ can be decomposed as $\mathbf{H} = \mathbf{LL}^H$, where $\mathbf{L}$ is a lower triangular matrix with strictly positive diagonal entries and $\mathbf{L}^H$denotes the conjugate transpose of $\mathbf{L}$. The Cholesky decomposition is unique, i.e., given a Hermitian, positive-definite matrix $\mathbf{H}$, there is only one lower triangular matrix $\mathbf{L}$ with strictly positive diagonal entries such that $\mathbf{H} = \mathbf{LL}^H$. In the special case that $\mathbf{H}$ is a symmetric positive-definite matrix with real entries, then $\mathbf{L}$ has real entries as well [28,29].

[24]Given a matrix $\mathbf{H}$, its QR decomposition is a matrix decomposition of the form $\mathbf{H} = \mathbf{QR}$ where $\mathbf{R}$ is an upper triangular matrix and $\mathbf{Q}$ is a unitary matrix, i.e., one satisfying $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$ where $\mathbf{Q}^T$ is the transpose of $\mathbf{Q}$ and $\mathbf{I}$ is the identity matrix [28,29].

strategies. Since we eventually consider coded transmission where extracting good quality soft outputs require finding a predefined minimum number of candidates $N_{c_{\min}}$, we take a different approach toward this problem. Our interest lies in ensuring a certain minimum number of entries N_c at the bottom of the search tree. With the entries of $\mathbf{H}$ and the noise $\mathbf{n}$ at the receiver being random variables, N_c is obviously a random variable, as well. If we allow for all hypotheses fulfilling the original inequality to be tested, we obtain an upper bound on the expected complexity. Note that whenever an empty sphere is declared, the search must be restarted with an increased radius. Since the overall detection complexity is the sum of all attempts to find the ML point, it is unfavorable to choose a small initial search radius and then increase it stepwise until $N_{c_{\min}}$ candidates are found.

For our symmetric $N_{RX} = N_{TX}$ MIMO system, the initial value for R reduces to $R^2 = 2\sigma^2 K N_{RX}$. Since in our system we normalize $2\sigma^2 = 1$ and then scale the transmitter signal constellation, this is further reduced to $R^2 = K N_{RX}$ where the critical point lies in the selection of K. It was found empirically that it would be beneficial to use a small initial search radius at low SNR that is subsequently increased with increasing SNR. The results of evaluations provided in the literature suggest that using $K = \frac{1}{60} L R_c \sqrt{\frac{E_b}{N_0}}$, where L is the number of bits per symbol and R_c is the code rate, provides a good choice that minimizes detection complexity in a wide SNR range. The middle factor ensures that the scaling of the transmitter signal set is taken into account and the last term derives from the notion that in the high SNR region, finding too many candidates becomes less probable and that the number of events with an empty sphere should be minimized. The radius can be increased by a factor of 1.5, whenever an empty sphere is declared [181].

9.13.4.8 CSI feedback

The reciprocity principle in wireless communication indicates that the channel from an antenna A to another antenna B is identical to the transpose of the channel from antenna B to antenna A, provided that the two channels are measured at the same time, the same frequency, and the same antenna locations. This principle further suggests that the transmitter can obtain information of the downlink (A to B) channel from the uplink (B to A) channel measurements, which the receiver at A can conduct. This information can involve the instantaneous channel or other channel parameters, including the channel statistics. In real full-duplex communications, however, the downlink and uplink channels cannot use identical frequency, time, and spatial instances. The reciprocity principle may still hold approximately, if the difference in any of these dimensions is relatively small compared to the channel variation across the referenced dimension.

Let us consider a scenario in which the base station measures the uplink channel during reception and uses this measurement for the CSIT of the next transmission in the downlink. In voice communications, the downlink and uplink channels to and from all the users operate in consecutive time slots. Therefore, uplink

channel measurements can be made regularly using reference signals. These measurements periodically refresh the CSIT. In data communications, the downlink and uplink channels may not operate continuously; hence, specially scheduled uplink transmissions for channel measurements, known as channel sounding, are used. A subset of the users, for whom CSIT is required, is scheduled to send sounding signals. The sounding signals are orthogonal among simultaneously scheduled users, using orthogonal sub-carriers in OFDM or orthogonal codes in CDMA systems. Sounding is an effective channel measurement technique for systems with multiple transmit antennas at the base station.

One complication in using reciprocity methods is that the principle only applies to the radio channel between the antennas, while in practice the channel is measured and used at the baseband processor. Different transmit and receive RF hardware chains; therefore, become part of the downlink and uplink channels. Since these chains have different frequency transfer characteristics, reciprocity requires transmit–receive chain calibration to equalize the effects of the two chains. Calibration is expensive and has made open-loop methods less attractive in practice.

Another method of obtaining CSIT is using feedback from the UE. The channel information is measured at the receiver at B during the downlink (A to B) transmission and then sent to the transmitter at A on the uplink. In practice, the downlink transmission from a base station includes reference signals, received by all active users. These users, when instructed by the serving base station, can measure their respective channels. The users then periodically or aperiodically send their measured channel information in the uplink channel to the base station to be used in calculation of their CSIT.

The feedback transmission can either be scheduled separately or piggybacked on ongoing transmissions. In data communications, CSIT may be needed for only a subset of users, who are then scheduled to transmit their channel information. Feedback is not limited by the reciprocity requirements. However, it imposes additional system overhead by consuming radio resources. Techniques to reduce the amount of feedback have been a subject of intense study, e.g., designing vector codebooks, quantizing channel information, or selecting only the important information. Furthermore, feedback information is susceptible to channel variation due to the delay in the feedback loop. The effectiveness of feedback depends on this delay and the channel Doppler spread. For a fast time-varying channel in mobile communications, feedback techniques are usually effective up to a certain mobile speed, depending on the carrier frequency, the transmission frame length, and the round trip time. The effects of feedback delay and error have been analyzed for various precoding techniques in LTE/LTE-Advanced, revealing potentially severe performance degradation. Therefore, the optimal use of feedback must account for the information quality.

The trade-off among transmission rate, error rate, and SINR for the scenarios where the transmitter has no channel knowledge and the receiver has perfect knowledge of CSI has been studied in the literature. If the MIMO channel is

assumed to be block fading, the length of the transmitted codewords is less than or equal to the channel block length, and if the channel was perfectly known to the transmitter, we could choose a signaling rate equal to or less than channel capacity and achieve error-free transmission. The coding scheme to achieve capacity consists of performing modal decomposition which decouples the MIMO channel into parallel SISO channels and then using ideal SISO channel coding. In practice, turbo codes can achieve near MIMO channel capacity performance. If the channel is unknown to the transmitter, modal decomposition is not possible. Since the channel is selected randomly according to a given fading distribution, there will always be a non-zero probability that a given transmission rate is not supported by the channel. We assume that the transmitted codeword is decoded successfully if the rate is at or below the mutual information, assuming a spatially white transmit covariance matrix, associated with the given channel realization. A decoding error is declared if the rate exceeds the mutual information. Therefore, if the transmitter does not have any knowledge of the channel, the PER would be equal to the outage probability associated with the transmission rate. We define the diversity order for a given transmission rate as [78]:

$$d(R) = -\lim_{\gamma \to \infty} \frac{\log[P_e(R,\gamma)]}{\log(\gamma)} \tag{9.89}$$

where $P_e(R,\gamma)$ is the PER corresponding to transmission rate R and SNR γ. Thus, the diversity order is the magnitude of the slope of the PER plotted as a function of the SNR on the logarithmic scale [78].

The differential feedback is a different approach to the enhancement of feedback accuracy. The differential feedback exploits the correlation between precoding matrices adjacent in time or frequencies. It only feeds back the difference between the current and the previous beamforming matrices. If the channel variation between two feedbacks is small, quantization codewords are concentrated in a small region on Grassmannian manifold[25] and not uniformly distributed over the entire beamforming space. For low-mobility users, the differential feedback may improve the feedback accuracy without increasing the number of feedback bits. The feedback starts initially and restarts periodically by sending a single feedback that fully describes the precoder. Differential feedbacks follow the start and restart feedbacks. The start and restart feedbacks employ a codebook that is

[25]A Grassmann manifold is a certain collection of vector subspaces of a vector space. In particular, $g_{n,k}$ is the Grassmann manifold of k-dimensional subspaces of the vector space $\mathbb{R}^n$. It has a natural manifold structure as an orbit space of the Stiefel manifold $v_{n,k}$ of orthonormal k-frames in G^n. One of the main features of Grassmann manifolds is that they are classifying spaces for vector bundles. A manifold is a topological space that is locally Euclidean (i.e., around every point, there is a neighborhood that is topologically the same as the open unit ball in $\mathbb{R}^n$). The Grassmannian $Gr(n,k)$ is the set of k-dimensional subspaces in an n-dimensional vector space. For example, the set of lines $Gr(n,k)$ is projective space. The real Grassmannian (as well as the complex Grassmannian) are examples of manifolds [31,32].

self-contained, e.g., the base codebook. Figure 9.112 illustrates the operation of differential feedback across time [83].

An example of differential feedback is a transformation-based differential codebook. Let t, $\mathbf{D}(t)$, and $\mathbf{V}(t)$ denote the feedback index, the corresponding feedback matrix, and the corresponding precoder, respectively. The sequential index is reset to 0 at $T_{\max}+1$. The index for the start and restart feedbacks are 0. Let $\mathbf{U}$ be a vector or a matrix and $\mathbf{Q}_{\mathbf{U}}$ be a rotation matrix determined by $\mathbf{U}$. The precoders of the subsequent differential feedbacks are denoted as $\mathbf{D}(t)$ for $t = 1, 2, \ldots, T_{\max}$ and the corresponding precoders are given as follows:

$$\begin{aligned} \mathbf{V}(t) &= \mathbf{Q}_{\mathbf{V}(t-1)}\mathbf{D}(t) \\ \mathbf{D}(t) &= \mathbf{Q}^{H}_{\mathbf{V}(t-1)}\mathbf{V}(t) \end{aligned} \qquad t = 0, 1, 2, \ldots, T_{\max} \tag{9.90}$$

where the rotation matrix $\mathbf{Q}_{\mathbf{V}(t-1)}$ is an $N_{\mathrm{TX}} \times N_{\mathrm{TX}}$ unitary matrix derived from the previous precoder matrix $\mathbf{V}(t-1)$, and N_{TX} denotes the number of transmit antennas. The dimension of the feedback matrix $\mathbf{D}(t)$ is $N_{\mathrm{TX}} \times N_l$, where N_l denotes the number of spatial streams. The feedback matrix $\mathbf{D}(t)$ can be viewed as a description of time/frequency correlation properties which does not change over the entire frequency band and a long period of time, and $\mathbf{Q}_{\mathbf{V}(t-1)}$ can be viewed as a description of narrowband and short-term channel properties which correspond to a specific sub-band and subframe. Due to correlation between beamforming matrices across time and frequency, feedback overhead can be reduced, if the correlation property can be efficiently exploited. Let us take the time-domain correlation as an example to explain the differential feedback scheme. The hypothetical operations at the UE and at the eNB can be described as follows [18]:

1. Derivation of the differential codeword matrix at the UE, $\mathbf{D} = \hat{\mathbf{Q}}^{\mathrm{H}}_{\mathbf{V}(t-1)}\mathbf{V}(t) = [\hat{\mathbf{V}}(t-1), \hat{\mathbf{V}}^{\perp}(t-1)]^{\mathrm{H}}\mathbf{V}(t)$.
2. Quantization of the differential codeword matrix at the UE, $\hat{\mathbf{D}} = \arg\max_{\mathbf{D}_i \in \Omega} ||\mathbf{D}^{\mathrm{H}}\mathbf{D}_i||_F$ where $||\mathbf{D}^{\mathrm{H}}\mathbf{D}_i||_F$ denotes the Chordal distance between

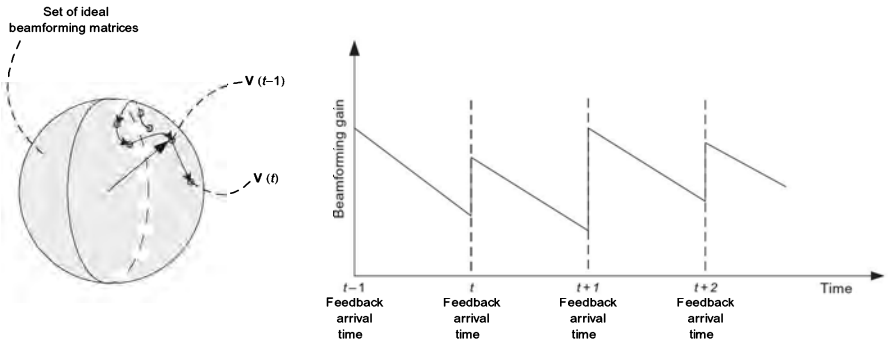


FIGURE 9.112

Example illustration of differential feedback and effect on beamforming gain.

the two codeword matrices $\mathbf{D}$ and $\mathbf{D}_i$. Therefore, the optimal codebook is selected based on the maximum distance between $\mathbf{D}$ and the codeword matrices contained in the codebook Ω.

3. Beamforming matrix reconstruction at the eNB, $\hat{\mathbf{V}}(t) = \hat{\mathbf{Q}}_{\mathbf{V}}(t-1)\hat{\mathbf{D}} = [\hat{\mathbf{V}}(t-1), \hat{\mathbf{V}}^{\perp}(t-1)]\hat{\mathbf{D}}$.
4. Beamforming at the eNB, $\mathbf{y} = \mathbf{H}\hat{\mathbf{V}}(t)\mathbf{s} + \mathbf{n}$.

In the above equations, $\mathbf{V}(t)$ and $\hat{\mathbf{V}}(t)$ denote the ideal and quantized $N_{\mathrm{TX}} \times N_l$ beamforming matrices at time instant t, respectively, the columns of $\hat{\mathbf{V}}^{\perp}(t-1)$ have unit norm and are orthogonal to $\hat{\mathbf{V}}(t-1)$ where $\hat{\mathbf{V}}^{\perp}(t-1)$ can be calculated from $\hat{\mathbf{V}}(t-1)$ by a Householder transformation,[26] $\mathbf{D}_i$ is the codeword matrix of the differential codebook denoted by Ω. The quantization criterion maximizes the received signal power. Other criteria such as channel capacity or mean squared error can also be used. If $N_l = 1$, $\mathbf{V}(t-1)$ would be a vector (with unit norm) and the rotation matrix is given by $\mathbf{Q}_{\mathbf{V}(t-1)} = \mathbf{I} - 2\varphi\varphi^{\mathrm{H}}; ||\varphi|| > 0$ wherein $\varphi = \exp(-\mathrm{j}\theta)\mathbf{V}(t-1) - \varepsilon$ and θ is the phase of the first entry of $\mathbf{V}(t-1)$ and $\varepsilon = (1, 0, 0, \ldots, 0)^{\mathrm{T}}$.

9.13.5 Precoding and beamforming

Multiple antennas at the transmitter and receiver can be used to achieve array and diversity gain instead of capacity gain. In this case, the same symbol weighted by a complex-valued scale factor is sent from each transmit antenna so that the input covariance matrix has unit rank. This scheme is referred to as beamforming. It must be noted that there are two conceptually and practically different classes of beamforming: (1) direction-of-arrival beamforming (i.e., adjustment of transmit or receive antenna directivity); and (2) eigen-beamforming (i.e., a mathematical approach to maximize signal power at the receive antenna based on certain criterion). In this section, we only consider eigen-beamforming schemes.

A classic eigen-beamforming scheme usually performs linear, single-layer, complex-valued weighting on the transmitted symbols such that the same signal is transmitted from each transmit antenna using appropriate weighting factors. In this scheme, the objective is to maximize the signal power at the receiver output. When the receiver has multiple antennas, the single-layer beamforming cannot

[26] In linear algebra, Householder transformation is defined as a linear transformation that describes a reflection about a plane or hyper-plane containing the origin. The Householder transformations are widely used in linear algebra to perform QR decompositions. This is a linear transformation given by the Householder matrix $\mathbf{Q} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^{\mathrm{H}}$ where $\mathbf{I}$ denotes the identity matrix and $\mathbf{u}$ is given as a column unit vector. The Householder matrix has the following properties: (1) it is hermitian, i.e., $\mathbf{Q} = \mathbf{Q}^{\mathrm{H}}$; and (2) it is unitary, i.e., $\mathbf{Q}^{-1} = \mathbf{Q}^{\mathrm{H}}$ [31,32].

simultaneously maximize the signal power at every receive antennas, hence, precoding is used for multi-layer beamforming in order to maximize the throughput of a multi-antenna system. Precoding is a generalized beamforming scheme to support multi-layer transmission in a MIMO system. Using precoding, multiple streams are transmitted from the transmit antennas with independent and appropriate weighting per each antenna such that the throughput is maximized at the receiver output.

Let us begin our concise study of eigen-beamforming and MIMO precoding using a simplified model. In a single-user MIMO (SU-MIMO) system, identity matrix precoding (for open-loop) and SVD precoding (for closed-loop) can be used to achieve link-level MIMO channel capacity. In addition, random unitary precoding can achieve the open-loop MIMO channel capacity with no signaling overhead in the uplink. The SVD precoding, on the other hand, has been shown to achieve the MIMO channel capacity when CSI is signaled to the transmitter. In a precoded SU-MIMO system with N_{TX} transmit antennas and N_{RX} receive antennas, the input–output relationship can be described as $\mathbf{y} = \mathbf{HWs} + \mathbf{n}$ where $\mathbf{s} = (s_1, s_2 \ldots, s_{\text{M}})^{\text{T}}$ is an $M \times 1$ vector of normalized complex-valued modulated symbols, $\mathbf{y} = (y_1, y_2, \ldots, y_{N_{\text{RX}}})^{\text{T}}$ and $\mathbf{n} = (n_1, n_2, \ldots, n_{N_{\text{RX}}})^{\text{T}}$ are the $N_{\text{RX}} \times 1$ vectors of received signal and noise, respectively, $\mathbf{H}$ is the $N_{\text{RX}} \times N_{\text{TX}}$ complex-valued channel matrix, and $\mathbf{W}$ is the $N_{\text{TX}} \times M$ linear precoding matrix. The superscript "T" denotes the transpose operator.

In the receiver, a hard-decoded symbol vector $\hat{\mathbf{s}}$ is obtained by decoding the received vector $\mathbf{y}$ using a vector decoder, assuming perfect knowledge of the channel and the precoding matrices. We assume that the entries of $\mathbf{H}$ are independent and identically distributed according to $\mathbb{Z}(0, 1)$, and the entries of noise vector $\mathbf{n}$ are independent and identically distributed according to $\mathbb{Z}(0, N_0)$. The input vector $\mathbf{s}$ is assumed to be normalized, thus $\text{E}[\mathbf{ss}^{\text{H}}] = \mathbf{I}$ where $\mathbf{I}$ is an identity matrix. Let us further assume that precoding matrix $\mathbf{W}$ is unitary, thus $\mathbf{WW}^{\text{H}} = \mathbf{I}$. The receiver selects a precoding matrix $\mathbf{W}_i, i = 1, 2, \ldots, N_{\text{codebook}}$ from a finite set of quantized precoding matrices $\Omega = \{\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{N_{\text{codebook}}}\}$ and sends the index of the chosen precoding matrix back to the transmitter over a low-delay feedback channel. There are two important issues concerning the above precoding scheme: (1) optimal selection criterion for choosing a precoding matrix from set $\mathbf{\Omega}$; and (2) design of codebook $\mathbf{\Omega}$. The matrix $\mathbf{W}_i, i = 1, 2, \ldots, N_{\text{codebook}}$ can be selected from $\mathbf{\Omega}$ by using either of the following optimization criteria [82]: (1) minimizing the trace of the mean squared error (MMSE-trace selection); (2) minimizing the determinant of the mean squared error (MMSE-determinant selection); (3) maximizing the minimum singular value of $\mathbf{HW}$ (singular value selection); (4) maximizing the instantaneous capacity (capacity selection); or (5) maximizing the minimum received symbol vector distance (minimum distance selection). The above selection criteria may be evaluated at the receiver using a full search over all matrices in $\mathbf{\Omega}$. Using distortion functions based on the selection criteria, it can be shown that the codebook $\mathbf{\Omega}$ is designed using Grassmannian subspace packing [82]. If MMSE-trace, singular value, or minimum distance selection is used, the codebook

is designed such that $\varepsilon = \min_{\mathbf{W_i} \neq \mathbf{W_j}} ||\mathbf{W}_i\mathbf{W}_i^{\mathrm{H}} - \mathbf{W}_j\mathbf{W}_j^{\mathrm{H}}||_2$[27] is maximized. If the MMSE-determinant or capacity selection optimization method is used, the codebook is designed such that $\varepsilon = \min_{\mathbf{W_i} \neq \mathbf{W_j}} \arccos|\det(\mathbf{W}_i^{\mathrm{H}}\mathbf{W}_j)|$ is maximized [82].

The MIMO codebooks are designed based on a trade-off between performance and complexity. The following are some desirable properties of the codebooks:

1. Low-complexity codebooks can be designed by choosing the elements of each constituent matrix or vector from a small binary set, e.g., a four alphabet $(\pm 1, \pm j)$ set, which eliminates the need for matrix or vector multiplication. In addition, the nested property of the codebooks can further reduce the complexity of CQI calculation when rank adaptation is performed.
2. The base station may perform rank overriding which results in significant CQI mismatch, if the codebook structure cannot adapt to it. A nested property with respect to rank overriding can be exploited to mitigate the mismatch effects.
3. Power amplifier balance is taken into consideration when designing codebooks with constant modulus property, which may eliminate unnecessary increase in PAPR.
4. Good performance for a wide range of propagation scenarios, e.g., uncorrelated, correlated, and dual-polarized antennas, is expected from the codebook design algorithms. A DFT-based codebook is optimal for linear array with small antenna spacing since the vectors match with the structure of the transmit array response. Additionally, with an optimal selection of the matrices and the entries of the codebook (rotated block diagonal structure), significant gains can be obtained in dual-polarized scenarios.
5. Low feedback and signaling overhead are desirable from an operation and performance perspective.
6. Low memory requirement is another design consideration for the MIMO codebooks.

An alternative approach to codebook-based precoding is to use differential codebook precoding, in which the transmitter computes the precoding matrix by using a differential codebook precoding matrix specified by the receiver and the previous precoding matrix used in the transmitter, thus $\mathbf{W}_k = \mathbf{C}_{m_{\mathrm{opt}}}\mathbf{W}_{k-1}$ where each differential codebook precoding matrix $\mathbf{C}_m, m = 1, 2, \ldots, N'_{\mathrm{codebook}}$ is an $N_{\mathrm{TX}} \times N_{\mathrm{TX}}$ matrix. The receiver selects the optimum differential precoding matrix which maximizes the instantaneous capacity as follows [83]:

$$\mathrm{Optimum-codebook-index}: m_{\mathrm{opt}} = \arg \max_{\mathrm{m}} \log_2 \det\left(\mathbf{I} + \frac{\gamma}{N_{\mathrm{TX}}}\mathbf{H}_k\mathbf{W}\mathbf{C}'_m{}^{\mathrm{H}}\mathbf{H}_k^{\mathrm{H}}\right) \tag{9.91}$$

[27]The Euclidean norm of square matrix $\mathbf{A}$ is defined as $||\mathbf{A}||_2 = \sup_{||\mathbf{x}||=1}||\mathbf{A}\mathbf{x}||_2$. The spectral norm of matrix $\mathbf{A}$ is the largest singular value of $\mathbf{A}$ or the square root of the largest eigenvalue of the positive-semi-definite matrix $\mathbf{A}^{\mathrm{H}}\mathbf{A}$; i.e., $||\mathbf{A}||_2 = \sqrt{\lambda_{\max}(\mathbf{A}^{\mathrm{H}}\mathbf{A})}$ where $\mathbf{A}^{\mathrm{H}}$ denotes the conjugate transpose of $\mathbf{A}$ [44].

where $\{\mathbf{C}'_1,\mathbf{C}'_2,\ldots,\mathbf{C}'_{N'_{\text{codebook}}}\} = \{\mathbf{C}_1\mathbf{W}_{k-1},\mathbf{C}_2\mathbf{W}_{k-1},\ldots,\mathbf{C}_{N'_{\text{codebook}}}\mathbf{W}_{k-1}\}$ and γ is the signal-to-noise ratio. The initial precoding matrix can be set as the first M columns of the $N_{\text{TX}} \times N_{\text{TX}}$ identity matrix. One advantage of the differential precoding scheme is that the candidate precoding matrices $\{\mathbf{C}'_1,\mathbf{C}'_2,\ldots,\mathbf{C}'_{N'_{\text{codebook}}}\}$ are refined in each precoding update instance, which virtually increases the codebook size to realize finer quantization of the channel [83]. Since the differential precoding schemes compute the precoding matrix by using the previous precoding matrix, different criteria should be used to generate precoding matrices. In a slow fading environment where precoding is more effective, the channel variation during the precoding matrix update interval is relatively small. In this case, the ideal codebook matrix which is multiplied by the previous precoding matrix should be close to the identity matrix. Thus, the codebook design strategy is to optimize the codebook such that the codebook includes quasi-diagonal matrices. More specifically, codebook matrices are computed from two unitary matrices as $\mathbf{C}_{\text{m}} = \mathbf{G}^{\text{H}}(t)\mathbf{G}(t + \Delta t)$, where $\mathbf{G}(t) \in \mathbb{C}^{N_{\text{TX}} \times N_{\text{TX}}}$ is a random unitary matrix. Note that since both $\mathbf{G}(t)$ and $\mathbf{G}(t + \Delta t)$ are unitary matrices, each codebook matrix $\mathbf{C}_{\text{m}} \in \mathbb{C}^{N_{\text{TX}} \times N_{\text{TX}}}$ also has unitary properties, and small perturbation Δt produces quasi-diagonal matrices whose diagonal entries are more dominant than other entries. A set of codebooks $\{\mathbf{C}_1,\mathbf{C}_2,\ldots,\mathbf{C}_{N'_{\text{codebook}}}\}$ are generated by using various Δt values where the codebook is optimized by capacity maximization criterion.

In SVD precoding, assuming the perfect knowledge of the CSI, the channel matrix is decomposed as $\mathbf{H} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\text{H}}$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices (i.e., $\mathbf{U}^{\text{H}}\mathbf{U} = \mathbf{I}, \mathbf{V}^{\text{H}}\mathbf{V} = \mathbf{I}$) and $\boldsymbol{\Sigma}$ is a diagonal matrix containing eigenvalues of the channel matrix. If the relationship between the output and input of the channel is described as $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$, then by replacing the channel matrix with its SVD form and multiplying both sides by $\mathbf{U}^{\text{H}}$, one can write $\mathbf{U}^{\text{H}}\mathbf{y} = \mathbf{U}^{\text{H}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\text{H}}\mathbf{s} + \mathbf{U}^{\text{H}}\mathbf{n}$. Further simplification of the latter equation would yield $\tilde{\mathbf{y}} = \boldsymbol{\Sigma}\mathbf{x} + \tilde{\mathbf{n}}$, if the input signal is precoded with $\mathbf{x} = \mathbf{V}^{\text{H}}\mathbf{s}$. Using SVD precoding, the channel matrix is diagonalized and the spatial interference is removed without any matrix inversion or non-linear processing. Since $\mathbf{U}$ is unitary, $\mathbf{U}^{\text{H}}\mathbf{n}$ still has the same variance as $\mathbf{n}$, thus, the SVD precoding does not result in noise enhancement. Due to substantial feedback and complexity of singular value decomposition of the channel matrix, the SVD precoding is not considered a viable precoding scheme. This procedure is illustrated in Figure 9.113.

In general, the precoding algorithms for multi-user MIMO (MU-MIMO) can be classified into linear and non-linear classes. Linear precoding schemes can achieve reasonable performance with lower complexity compared to non-linear precoding techniques. Linear precoding includes unitary and zero-forcing precoding schemes. Non-linear precoding can achieve near optimal capacity at the expense of increased complexity. Non-linear precoding is designed based on the concept of dirty paper coding (DPC), where any known interference at the transmitter can be canceled if the optimal precoding scheme is applied to the transmit signal.

Unitary precoding includes unitary and semi-unitary precoding, both of which are a simple extension of SVD precoding in SU-MIMO with the addition of the SDMA-based user scheduling technique. The SDMA-based opportunistic user

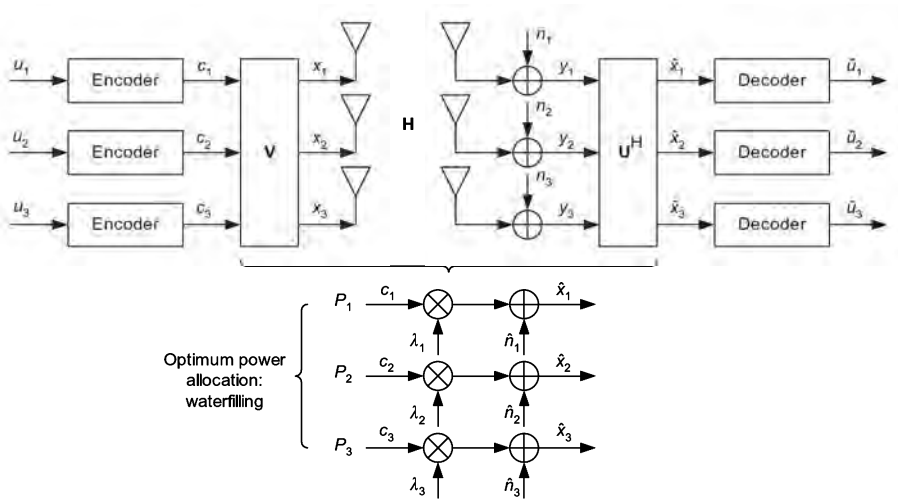


FIGURE 9.113

Illustration of the SVD concept [46].

scheduling technique groups near orthogonal users to avoid intra-group interferences at the cost of additional signaling overhead, which results in higher performance relative to the SU-MIMO. For example, it can increase diversity order to approximately the number of transmit antennas even with simple linear decoding at the receiver.

The zero-forcing precoding consists of zero-forcing and regularized zero-forcing precoding. If the transmitter has perfect knowledge of the downlink CSI, the ZF-based precoding can achieve near channel capacity performance when the number of users is large. With limited CSI at the transmitter, ZF-precoding requires increased feedback to achieve the full multiplexing gain. Hence, inaccurate CSI at the transmitter may result in significant loss of performance due to residual interference among transmit streams.

DPC is a coding technique that cancels known interference without power penalty provided that the transmitter has perfect knowledge of the interfering signals regardless of whether the receiver has knowledge of CSI. This category includes Costa precoding [89], Tomlinson-Harashima precoding [80], and the vector perturbation technique [80]. Vector perturbation uses modulo operation at the transmitter to perturb the transmitted signal vector to avoid the transmit power enhancement incurred by zero-forcing methods. The optimal perturbation method is found by solving a minimum distance problem, and thus can be implemented using sphere decoding or full search-based algorithms [73–80].

Now let us consider MU-MIMO systems and briefly study how precoding is applied in those scenarios. In the downlink direction of a precoded MU-MIMO system (alternatively known as broadcast channel in the literature) with N_{TX} transmit antennas at the base station and one receive antenna at the kth mobile station, the input–output relationship can be written as $y_k = \mathbf{h}_k^H \mathbf{x} + n_k, k = 1, 2, \ldots, N_{user}$, where

$\mathbf{x} = \sum_{i=1}^{N_{\text{user}}} s_i \mathbf{w}_i$ is the $N_{\text{TX}} \times 1$ vector of weighted transmitted symbols s_i, y_k and n_k are the received signal and noise, respectively, $\mathbf{h}_k$ is the kth $N_{\text{TX}} \times 1$ channel vector, where matrix $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_{\text{user}}})^{\text{T}}$ is the $N_{\text{user}} \times N_{\text{TX}}$ complex-valued downlink channel matrix, and $\mathbf{w}_k$ is the kth $N_{\text{TX}} \times 1$ normalized linear precoding vector.

The mathematical relationship for the input and output of a precoded MU-MIMO system in the uplink (alternatively known as multiple access channel in the literature) with N_{RX} receive antennas at the base station and one transmit antenna at each user's terminal can be written as $\mathbf{y} = \sum_{k=1}^{N_{\text{user}}} s_k v_k \mathbf{h}_k + \mathbf{n}$ where $s_k v_k$ is the weighted complex-valued modulated symbol from the kth user, $\mathbf{y} = (y_1, y_2, \ldots, y_{N_{\text{RX}}})^{\text{T}}$ and $\mathbf{n} = (n_1, n_2, \ldots, n_{N_{\text{RX}}})^{\text{T}}$ are the $N_{\text{RX}} \times 1$ vectors of received signal and noise, respectively, $\mathbf{h}_k$ is the kth $N_{\text{RX}} \times 1$ channel vector, where matrix $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_{\text{user}}})^{\text{T}}$ is the $N_{\text{RX}} \times N_{\text{user}}$ complex-valued uplink channel matrix. As mentioned earlier, perfect knowledge of the CSI is necessary at the transmitter in order to achieve the capacity of a MU-MIMO channel. However, in practical systems, the receiver only provides partial CSI through uplink feedback channels to the transmitter, i.e., the MU-MIMO precoding with limited feedback.

The received signal in the downlink of a MU-MIMO system with limited feedback precoding is mathematically expressed as: $y_k = \mathbf{h}_k^{\text{H}} \sum_{i=1}^{N_{\text{user}}} s_i \hat{\mathbf{w}}_i + n_k, k = 1, 2, \ldots, N_{\text{user}}$. The transmit vector for limited feedback precoding is modeled as $\hat{\mathbf{w}}_i = \mathbf{w}_i + \boldsymbol{\varepsilon}_i$ where $\boldsymbol{\varepsilon}_i$ is the error vector generated as a result of the limited feedback and vector quantization, and the expression for the received signal can be rewritten as:

$$y_k = \mathbf{h}_k^{\text{H}} \sum_{i=1}^{N_{\text{user}}} s_i \mathbf{w}_i + \mathbf{h}_k^{H} \sum_{i=1}^{N_{\text{user}}} s_i \boldsymbol{\varepsilon}_i + n_k, \quad k = 1, 2, \ldots, N_{\text{user}} \tag{9.92}$$

where $\mathbf{h}_k^{\text{H}} \sum_{i=1}^{N_{\text{user}}} s_i \boldsymbol{\varepsilon}_i$ is the residual interference due to the limited-feedback precoding. To reduce the residual interference term, one should use more accurate CSI feedback which results in the use of more uplink resources for the feedback. It is shown in the literature that the number of feedback bits per user N_{feedback} must be increased linearly with the SNR γ_{dB} (in decibels) at the rate of $N_{\text{feedback}} = (N_{\text{TX}} - 1)\log_2 \gamma = \gamma_{\text{dB}}(N_{\text{TX}} - 1)/3$ in order to achieve the full multiplexing gain of N_{TX}. In addition, the scaling of N_{feedback} guarantees that the throughput loss relative to zero-forcing precoding with perfect CSI knowledge at the transmitter is upper-bounded by N_{TX} bps/Hz, which corresponds to approximately 3 dB power offset [78]. The throughput of a feedback-based zero-forcing system is bounded, if the SNR approaches infinity and the number of feedback bits per user is fixed. Reducing the number of feedback bits according to $N_{\text{feedback}} = \alpha \log_2 \gamma$ for any $\alpha < N_{\text{TX}} - 1$ results in a strictly inferior multiplexing gain of $N_{\text{TX}}[\alpha/(N_{\text{TX}} - 1)]$ where N_{TX} is the number of transmit antennas and γ is the SNR of the downlink channel. In order to calculate the amount of feedback required to maintain certain

throughput, the difference between the feedback rates of zero-forcing precoding with perfect feedback $R_{PF-ZF}(\gamma)$ and with limited feedback $R_{LF-ZF}(\gamma)$ is required to satisfy the following constraint $\Delta R(\gamma) = R_{PF-ZF}(\gamma) - R_{LF-ZF}(\gamma) \leq \log_2 b$. In order to maintain a rate offset less than $\log_2 b$ (per user) between zero-forcing with perfect CSI and with finite-rate feedback (i.e., $\Delta R(\gamma) \leq \log_2 b, \forall \gamma$), it is sufficient to scale the number of feedback bits per user according to $N_{feedback} = \gamma_{dB}(N_{TX} - 1)/3 - (N_{TX} - 1)\log_2(b - 1)$. The rate offset of $\log_2 b$ (per user) is translated into a power offset, which is a more useful metric from the design perspective. Since a multiplexing gain of N_{TX} is achieved with zero-forcing, the zero-forcing curve has a slope of N_{TX} bps/Hz/3 dB at asymptotically high SNR. Therefore, a rate offset of $\log_2 b$ bps/Hz per user corresponds to a power offset of $3\log_2 b$ decibels. To feedback $N_{feedback}$ bits through the uplink channel, the throughput of the uplink feedback channel should be larger than or equal to $N_{feedback}$, i.e., $w_{FB} \log_2(1 + \gamma_{FB}) \geq N_{feedback}$ where γ_{FB} denotes the SNR of the feedback channel. Thus, the required feedback resource to satisfy the constraint $\Delta R(\gamma) \leq \log_2 b$ can be shown to be given as follows $w_{FB} \geq [\gamma_{dB}(N_{TX} - 1)/3 - (N_{TX} - 1)\log_2(b - 1)]/\log_2(1 + \gamma_{FB})$, i.e., the required feedback resource is a function of both downlink and uplink channel conditions [78].

We defined precoding as adaptive or non-adaptive weighting of the spatial streams prior to transmission from each antenna port (in a multi-antenna configuration) using a precoding matrix for the purpose of improving the reception or separation of the spatial streams at the receiver. Both feed-back and feed-forward precoder matrix selection schemes can be used in order to select the optimal weights. Feed-back precoding matrix selection techniques do not rely on channel reciprocity, rather they use feedback channels provided that the feedback latency is less than the channel coherence time. In feed-forward approaches, the necessary CSI can be theoretically obtained through direct feedback, where the CSI is explicitly signaled to the transmitter by the receiver, or sounding reference signals. The direct channel feedback methods preclude the channel reciprocity requirement, whereas channel sounding methods rely on channel reciprocity. Therefore, explicit control signaling is required for the PMI-based (feedback) schemes. However, in reciprocity-based schemes, the sounding signals in the uplink and precoded pilots in the downlink are used to assist the transmitter and receiver to appropriately select the precoding matrix. The reciprocity-based schemes have the additional advantage of not being constrained to a finite set of codebooks. Beamforming relies on long-term statistics of the radio channel and unlike reciprocity-based techniques, does not require short-term correlation between the uplink and downlink in order to properly function.

9.13.6 SU-MIMO and MU-MIMO

Single-user MIMO techniques are point-to-point schemes that improve channel capacity and reliability through the use of space–time/space–frequency codes (transmit/receive diversity) in conjunction with spatial multiplexing schemes. In a

SU-MIMO transmission, the advantage of MIMO processing is obtained from the coordination of processing among all the transmitters or receivers. In the multi-user channel, on the other hand, it is usually assumed that there is no coordination among the users. As a result of the lack of coordination among users, uplink and downlink MU-MIMO channels are different. In the uplink scenario, users transmit to the base station over the same channel. The challenge for the base station is to separate the signals transmitted by the users, using array processing or multi-user detection methods. Since the users are not able to coordinate with each other, there is not much that can be done to optimize the transmitted signals with respect to each other. If some channel feedback is allowed from the transmitter back to the users, some coordination may be possible, but it may require that each user knows all the other users' channels rather than only its own. Otherwise, the challenge in the uplink is mainly in the processing done by the base station to separate the users. In the downlink channel, where the base station simultaneously transmits to a group of users over the same channel, there is some inter-user interference for each user which is generated by the signals transmitted to other users. Using multi-user detection techniques, it may be possible for a given user to overcome the multiple access interference, but such techniques are often extremely complicated for use at the receivers. Ideally, one would like to mitigate the interference at the transmitter by carefully designing the transmitted signal. If CSI is available at the transmitter, it is aware of what interference is being created for user i by the signal it is transmitting to user j, and vice versa. The inter-user interference can be mitigated by beamforming or the use of dirty paper codes.

In general, SU-MIMO and MU-MIMO schemes are compared as follows [60–64]:

- SU-MIMO is a point-to-point link with predictable link capacity, whereas a MU-MIMO channel is a broadcast channel in the downlink direction and a multiple access channel in the uplink direction whose link-level data rates are characterized in terms of capacity regions.
- Multi-layer SU-MIMO schemes offer layer/stream diversity in the sense that if one stream has a poor SNR, the system will not necessarily experience an outage, whereas in the same situation, a MU-MIMO system will be in outage. This is due to the fact that in MU-MIMO schemes, users have typically an equal target data rate and symbol error rate on their respective links, while in SU-MIMO systems, only the sum rate of the overall link is considered, since all streams are delivered to same user.
- The MU-MIMO schemes suffer from a near-far problem due to significant differences between the path losses experienced by each user, resulting in a large deviation in the SINR of the corresponding user links. This would benefit the users with better channel conditions, while there is no near-far problem in SU-MIMO systems. The near-far problem in MU-MIMO systems may be alleviated via appropriate grouping of the users with similar channel conditions.

Ex.1005.00674

- The use of cooperative collocated transmit antennas in SU-MIMO schemes can facilitate the encoding at the transmitter and decoding at the receiver. In contrast, the users in a MU-MIMO scheme can cooperate in encoding at the base station in the downlink and decoding in the uplink; however, the users cannot cooperate in decoding in the downlink or encoding in the uplink directions.
- The capacity of the downlink and uplink are theoretically identical in the downlink and uplink for the SU-MIMO systems (given the same transmit power and perfect channel knowledge in the transmitter and the receiver); however, the capacity of the MU-MIMO broadcast channel and multiple access channel are not identical.
- The capacity of the SU-MIMO schemes is less impacted by lack of CSI at the transmitter, whereas the capacity of the MU-MIMO broadcast channel significantly suffers from lack of CSI at the transmitter.
- SU-MIMO suffers from limited exploitation of multi-user diversity. The number of spatial dimensions is limited by the number of antennas at the UE. There is a potential that spatial dimensions are wasted, if the UEs have fewer antennas compared to the base station.
- MU-MIMO more efficiently exploits the multi-user diversity since all spatial dimensions which are supported by the base station can be exploited. It will achieve capacity gain, if UEs have fewer antennas relative to the base station. Stronger spatial dimensions are exploited, particularly in the case of a low rank channel. The utilized spatial dimensions may be weak in the case of a low rank channel due to spatial correlation.

The advantages/disadvantages of SU-MIMO and MU-MIMO schemes are summarized in Table 9.27.

An important metric for measuring the performance of any communication channel is the information theoretic capacity. In a SU-MIMO channel, the capacity is the maximum amount of information that can be transmitted as a function of available bandwidth given a constraint on transmitted power. In SU-MIMO channels, it is common to assume that the total power distributed among all

Table 9.27 Comparison of SU-MIMO and MU-MIMO Schemes [46]

	SU-MIMO	MU-MIMO
Advantages	High user throughput High peak data rates	High system capacity Full exploitation of multi-user diversity
Disadvantages	Multiple transmit antennas at the base station are not fully exploited Multi-user diversity is not fully exploited	Degradation of peak data rates due to inter-user interference (ZF may not work perfectly due to imperfect CSIT)

transmit antennas is limited. For the MU-MIMO channel, the problem is somewhat more complex. Given the constraint on the total transmit power, it is possible to allocate varying fractions of that power to different users in the network, thus for any value of total power, different information rates are obtained. The result is a capacity region shown in Figure 9.114 for a two-user MU-MIMO channel. The maximum capacity for user 1 is achieved when 100% of the power is allocated to user 1, and for user 2, the maximum capacity is also obtained when it is allocated the full power. For every possible power distribution, there is an achievable information rate, which results in the capacity regions depicted in the figure. Two regions are shown in Figure 9.114, the larger one for the case where both users have roughly the same maximum capacity (similar channel conditions), and the other region for a case where one of the users has a much better channel condition than the other user. For N_{user} users, the capacity region is characterized by an N_{user}-dimensional hyper-region.

Let us use a simple MU-MIMO system model to demonstrate how the sum rate of the system is calculated. The simplified model of MU-MIMO broadcast channel is shown in Figure 9.115. Based on this model, the transmit vector $\mathbf{x}$ can be expressed as the weighted sum (precoded) of the input data symbols $s_k|_{k=1,2,\ldots,N_{\text{user}}}$ as follows: $\mathbf{x} = \sum_k \mathbf{p}_k s_k = \mathbf{Ps}$ where $N_{\text{user}} \times 1$ dimensional vector $\mathbf{s} = (s_1, s_2, \ldots, s_{N_{\text{user}}})^{\text{T}}$ denotes the data symbols from N_{user} users, and $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{N_{\text{user}}})$ is the precoding matrix comprising N_{user} precoding vectors. It is assumed that the finite transmit power at the transmitter can be calculated as $P_{\text{TX}} = E\{\mathbf{x}^{\text{H}}\mathbf{x}\}$. The kth complex-valued output of the system can be written as

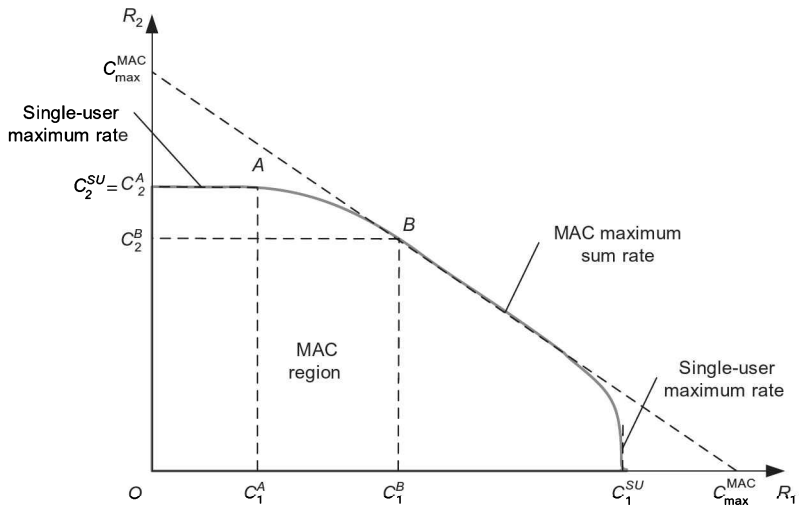


FIGURE 9.114

Capacity region of MU-MIMO multiple access channel (MAC) with two users compared to SU-MIMO [46].

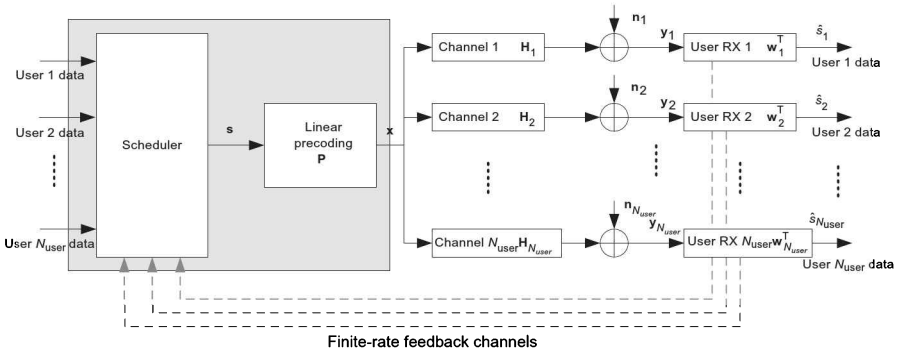


FIGURE 9.115

MU-MIMO broadcast channel model [46]

$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{n}_k \in \mathbb{C}^N$ where N denotes the dimension of vector $\mathbf{y}$. The kth branch user data can be detected using a linear MMSE receiver as $\hat{s}_k = \mathbf{w}_k^{\mathrm{T}}\mathbf{x}_k \in \mathbb{C}$ in which the MMSE weighting matrix is given by [46]:

$$\mathbf{w}_k = \left(\mathbf{H}_k^{*}\mathbf{P}^{*}\mathbf{P}^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}} + \frac{N_{user}}{P_{\mathrm{TX}}}\mathbf{I}_N\right)^{-1}\mathbf{H}_k^{*}\mathbf{p}_k^{*} \tag{9.93}$$

It can be further shown that the SINR at the kth output is given by:

$$\gamma_k = \frac{|\mathbf{w}_k^{\mathrm{T}}\mathbf{H}_k\mathbf{p}_k|^2}{||\mathbf{w}_k||_2^2\frac{N_{user}}{P_{\mathrm{TX}}} + \sum\limits_{i(i\neq k)}|\mathbf{w}_k^{\mathrm{T}}\mathbf{H}_k\mathbf{p}_i|^2} \tag{9.94}$$

The sum rate of the system is given as $R_{\mathrm{sum}} = \sum_k \log_2(1+\gamma_k)$ [46].

Figure 9.116 illustrates a generic MU-MIMO scenario where N_{user} users, each having N_{TX_i} transmit antennas, transmit simultaneously to the eNB in the uplink. The maximum achievable throughput of the system is determined by the point on the N_{user}-dimensional hyper-region that maximizes the sum of all of the users' information rates, and is referred to as the sum capacity of the channel (Figure 9.117). In the case of a near-far problem where one user has a more strongly attenuated channel than other users, or some users are closer to the eNB than others, obtaining the sum capacity would cause the user with the worst channel to achieve a lower rate.

As shown in Figure 9.116, in the uplink of a MU-MIMO system, the received signal at the eNB can be written as $\mathbf{y} = \sum_{k=1}^{N_{\mathrm{user}}}\mathbf{H}_k^{\mathrm{H}}\mathbf{x}_k + \mathbf{n}$ where $\mathbf{x}_k$ is the $N_{\mathrm{TX}_k} \times 1$ transmitted signal vector of the kth UE with N_{TX_k} transmit antennas, $\mathbf{H}_k \in \mathbb{C}^{N_{\mathrm{TX}_k} \times N_{\mathrm{RX}}}$ denotes the flat-fading channel matrix from the kth user to the eNB, and $\mathbf{n} = (n_1, n_2, \ldots, n_{N_{\mathrm{RX}}}), n_k \sim \mathbb{N}(0,1)$ is an independent and identically

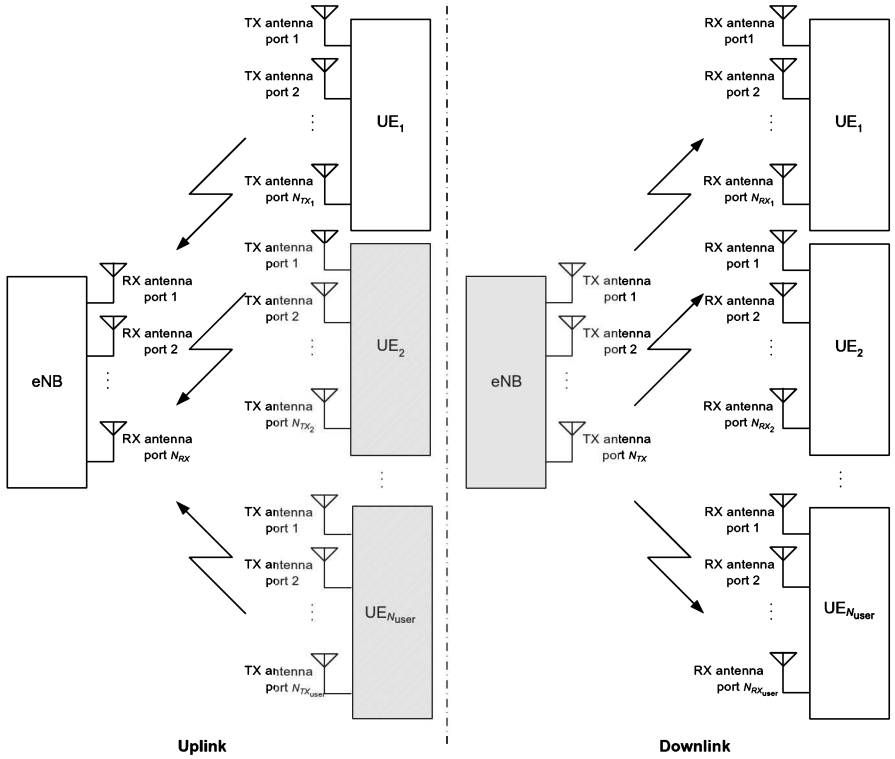


FIGURE 9.116

A generic multi-user MIMO transmission model [18].

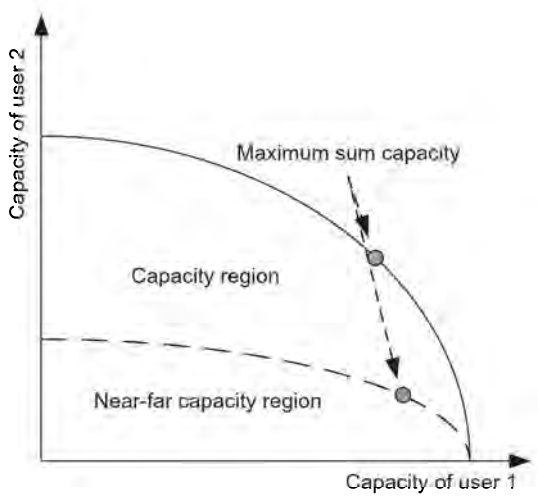


FIGURE 9.117

An example illustration of capacity region.

distributed additive white Gaussian noise vector at the eNB. We assume that the receiver k has perfect and instantaneous knowledge of the channel matrix $\mathbf{H}_k$. Note that the eNB is equipped with N_{TX} transmit and N_{RX} receive antennas.

In the downlink as illustrated in Figure 9.116, the received signal at the kth receiver can be written as $\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{n}_k \forall k = 1, 2, \ldots, N_{\text{user}}$ where $\mathbf{H}_k \in \mathbb{C}^{N_{\text{RX}_k} \times N_{\text{TX}}}$ is the downlink channel and $\mathbf{n}_k \in \mathbb{C}^{N_{\text{RX}_k} \times 1}$ is the complex-valued additive Gaussian noise vector at the kth receiver. We assume that each receiver also has perfect and instantaneous knowledge of its own channel matrix $\mathbf{H}_k$. The transmitted signal $\mathbf{x}$ is a function of multiple users' information data, i.e., $\mathbf{x} = \sum_{k=1}^{N_{\text{user}}} \mathbf{x}_k$ where $\mathbf{x}_k$ is the signal carrying the kth user's message with covariance matrix $\mathbf{\Omega}_k = E\{\mathbf{x}_k\mathbf{x}_k^{\text{H}}\}$. The power allocated to the kth user is given by $\rho_k = \text{trace}(\mathbf{\Omega}_k)$. Under a sum power constraint at the eNB, the power allocation needs to maintain $\sum_{k=1}^{N_{\text{user}}} \rho_k \leq P_{\text{total}}$. Assuming a unit variance for the noise, it is now known that the capacity region for a given matrix channel realization can be written as [80]:

$$C_{\text{DL}} = \bigcup_{(\rho_1, \rho_2, \ldots, \rho_{N_{\text{user}}} | \sum \rho_k \leq P_{\text{total}})} (R_1, R_2, \ldots, R_{N_{\text{user}}}) \in \mathbb{R}^{+N_{\text{user}}}, R_l \leq \log_2 \frac{\det\left[\mathbf{I} + \mathbf{H}_i\left(\sum_{j \geq i} \mathbf{\Omega}_j\right)\mathbf{H}_i^{\text{H}}\right]}{\det\left[\mathbf{I} + \mathbf{H}_i\left(\sum_{j > i} \mathbf{\Omega}_j\right)\mathbf{H}_i^{\text{H}}\right]} \Bigg\} \tag{9.95}$$

where $\mathbb{R}^{+N_{\text{user}}}$ is the N_{user}-dimensional set of positive real numbers. The above equation may be optimized over each possible user ordering. Although difficult to realize in practice, the computation of the capacity region can be simplified using the assumption that the downlink capacity region can be calculated through the union of regions of the dual multiple access channel with all uplink power allocation vectors meeting the sum power constraint. The fundamental effect of the use of multiple antennas at either the eNB or the user terminals in increasing the channel capacity is best comprehended by examining how the sum capacity, i.e., the point obtained by the maximum $\sum_{k=1}^{N_{\text{user}}} R_k$ in the capacity region, scales with the number of active users.

The capacity region in the uplink of a MU-MIMO system can be calculated for two different cases: (1) joint decoding, i.e., the signals from a group of users are decoded in a cooperative manner; and (2) independent decoding where the signals from users are independently decoded in parallel.

The capacity of the MU-MIMO channels can be calculated using DPC techniques. The method was inspired by interference cancelation schemes. It is shown that the capacity of a channel where the transmitter has knowledge of interference is the same as if there were no interference. The DPC concept can be better

comprehended by comparing the scheme to writing on a dirty paper with a properly selected ink color so that the scripts have high contrast and are clearly readable. It can be shown that the capacity of the MU-MIMO multiple-access channel in the above cases is given as follows [75]:

$$R_i \leq \log_2\left[\frac{\det(\mathbf{R}_\mathbf{y})}{\det(\mathbf{R}_\mathbf{y} - E_{s_i}\mathbf{h}_i\mathbf{h}_i^{\mathrm{H}})}\right], \quad i = 1, 2, \ldots, N_{\text{user}} \quad \text{(independent decoding)}$$

$$\sum_{k\in\Re} R_k \leq \log_2 \det\left(\mathbf{I}_{N_{\text{RX}}} + \frac{1}{N_0}\mathbf{H}_\Re\mathbf{R}_{\mathbf{s}_\Re}\mathbf{H}_\Re^{\mathrm{H}}\right) \quad \text{(joint decoding)} \tag{9.96}$$

In the above equations N_{RX} denotes the number of receive antennas at the base station, N_{user} is the number of users each with only one transmit antenna, $\Re$ is a subset of users whose signals are jointly decoded, $\mathbf{H}_\Re$ is the $N_{\text{RX}} \times N_\Re$ complex-valued channel matrix where $N_\Re$ is the cardinality of the set $\Re$, $\mathbf{R}_{\mathbf{s}_\Re}$ denotes the covariance matrix of the signals transmitted by the users contained in the set $\Re$, $\mathbf{I}_{N_{\text{RX}}}$ is an $N_{\text{RX}} \times N_{\text{RX}}$ identity matrix, R_k is the rate that can be reliably achieved by user k, s_i is the complex-valued symbol transmitted by user $i, i = 1, 2, \ldots, N_{\text{user}}$, E_{s_i} is the average energy of the signal from the ith user which differs from one user to another, $\mathbf{R}_\mathbf{y}$ is the covariance matrix of the received signal at the base station, and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_{\text{user}}}]$ is the normalized $N_{\text{RX}} \times N_{\text{user}}$ complex-valued channel matrix composed of individual user complex-valued channel vectors $\mathbf{h}_i$.

An efficient UE pairing scheme is required at the eNB to choose the correct pair of UEs for MU-MIMO transmission in MU-MIMO systems. This pairing scheme is required to maintain minimal interference among scheduled UEs in MU-MIMO transmission. A proper pairing scheme can be designed by maximizing the Chordal distance[28] between the feedback precoding matrices of the UEs. The Chordal distance between two matrices is given in [31,32] and represented by $d_{\text{c}}(\mathbf{p}_i, \mathbf{p}_j) = (1/\sqrt{2})||\mathbf{p}_i\mathbf{p}_i^{\mathrm{H}} - \mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}||_F$ where $||.||_{\text{F}}$ denotes the Frobenius norm of the matrix. The Chordal distance generalizes the distance between two points on the unit sphere through an isometric embedding from complex Grassmann manifold $Gr(N_{\text{TX}}, N_l)$ to the unit sphere. Assuming an infinite number of UEs served by the current eNB, the kth UE with reported precoding matrix $\mathbf{p}_k$ will be paired with the mth UE, where the mth UE reports precoding matrix $\mathbf{p}_m$ and the Chordal distance between precoding matrices is maximized. With the maximized Chordal

[28]The asymptotic performance of a coding scheme is dominated by the shortest distance between any pair of codewords. The relevant distance measure between two codewords (vectors) $\mathbf{x}_1$ and $\mathbf{x}_2$ of an orthogonal code (of size M) for a non-coherent MIMO system is the Chordal distance defined as $d^2(\mathbf{x}_1, \mathbf{x}_2) = M - \left\|\mathbf{x}_1\mathbf{x}_2^{\mathrm{H}}\right\|_{\text{F}}^2$ [31,32].

distance criterion, $\mathbf{p}_m$ stays in the anti-polar position of $\mathbf{p}_k$, hence $||\mathbf{H}_k\mathbf{p}_m||^2$ is minimized, yielding the minimized inter-UE interference. Consequently, the UE pairing scheme for MU-MIMO transmission in practical systems is designed to find the best match between two UEs (e.g., the *m*th UE and the *k*th UE) based on the reported precoding matrices and the following criterion $\mathbf{p}_m = \arg \max_{\forall \mathbf{p}_i \in \mathbf{P}_{UE}} d_c(\mathbf{p}_i, \mathbf{p}_k)$ with $\mathbf{P}_{UE}$ representing the set containing all reported precoding matrices at a certain eNB.

If the conventional AWGN channel were modified to include an additive interference term that is known at the transmitter, and the received signal were expressed as the sum of the transmitted signal + interference + noise, one could set the transmitted signal equal to the desired data minus the interference term; however, such an approach requires increased power and perfect knowledge of interference. It is shown in the literature that the capacity in such a scenario is the same as if the interference were not present; no additional power is required to cancel the interference than is used in a nominal additive white Gaussian noise channel. Therefore, writing on dirty paper, from an information theoretic perspective, is equivalent to writing on clean paper when one knows in advance where the dirt is located. This concept has been used to characterize the sum-capacity and capacity region of the multi-antenna multi-user channels. A DPC technique for the downlink of a MU-MIMO system uses the QR decomposition[29] of the channel matrix as the product of a lower triangular matrix $\mathbf{R}$ with a unitary matrix $\mathbf{Q}$ such that $\mathbf{H} = \mathbf{QR}$. The signal to be transmitted is precoded with the $\mathbf{Q}^H$ matrix, resulting in the effective channel $\mathbf{R}$. The first user of this equivalent system sees no interference from other users, i.e., the signal may be chosen irrespective of other users. The second user sees interference only from the first user; this interference is known and thus may be canceled using DPC. The other users are dealt with in a similar manner. Alternative approaches apply dirty paper techniques directly, rather than for individual users. An important difference between the MU-MIMO channel and the interference channels for which dirty paper techniques are applied is that the interference depends on the signal being designed [80].

Block diagonalization (BD) is a linear precoding technique for the downlink of MU-MIMO systems. It decomposes a MU-MIMO downlink channel into multiple parallel orthogonal SU-MIMO channels. The signal of each user is pre-processed at the transmitter using a modulation matrix that lies in the null space of all other users' channel matrices; thereby the multi-user interference in the system is effectively zero. The block diagonalization scheme is restricted

[29]In linear algebra, the QR decomposition of a matrix is defined as decomposition of a matrix into an orthogonal matrix and an upper triangular matrix. The QR decomposition is often used to solve the linear least squares problems. The QR decomposition of a real square matrix $\mathbf{A}$ is $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q}$ is an orthogonal matrix, i.e., $\mathbf{Q}^H\mathbf{Q} = \mathbf{I}$ and $\mathbf{R}$ is an upper triangular matrix. This can be generalized to a complex-valued square matrix $\mathbf{A}$ and a unitary matrix $\mathbf{Q}$. If $\mathbf{A}$ is non-singular, then this factorization is unique if one requires that the diagonal elements of $\mathbf{R}$ are positive-valued [29].

to channels where the number of transmit antennas N_{TX} is greater or equal to the total number of receive antennas in the network N_{RX}. Let us define the precoder matrices as $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_{N_{user}}] \in \mathbb{C}^{N_{RX} \times N_l}$ where $\mathbf{F}_i \in \mathbb{C}^{N_{RX} \times N_{l_i}}$ is the ith user precoder matrix. We assume that $N_l \leq N_{RX}$ is the total number of the transmitted data streams, while $N_{l_i} \leq N_{RX_i}$ is the number of data streams transmitted to the ith user. We can find the optimal precoding matrix $\mathbf{F}$ such that multi-user interference is zero by choosing a precoding matrix $\mathbf{F}_i$ that lies in the null space of the other users' channel matrices. Thereby, an MU-MIMO downlink channel is decomposed into multiple parallel independent SU-MIMO channels [80].

Minimum mean-squared-error beamforming improves the system performance by allowing a certain amount of interference especially for users equipped with a single antenna. However, it suffers from performance loss when it attempts to mitigate the interference between two closely spaced antennas, a situation that usually occurs when the user terminal is equipped with more than one receive antenna. Successive MMSE (SMMSE), which is an improved version of the MMSE precoding algorithm, was developed to mitigate this problem by successively calculating the columns of the precoding matrix for each of the receive antennas separately. Linear equalization suffers from noise enhancement, and hence has poor power efficiency in some cases. The same drawback is experienced by linear precoding, which alleviates noise by boosting the transmit power [76].

The Tomlinson–Harashima precoding (THP) algorithm is a non-linear precoding technique originally developed for multipath SISO channels. The THP algorithm is performed by moving the feedback part of the decision-feedback equalizer to the transmitter. It has also been applied for the pre-equalization of multi-user interference in MU-MIMO systems; hence, no error propagation occurs and the precoding can be performed for the interference-free channel. The MMSE precoding in combination with THP is another approach to balance the multi-user interference in order to reduce the performance loss that occurs with zero interference techniques while THP is used to reduce the multi-user interference and to improve the diversity [46].

Space-division multiple access (SDMA) is a special case of MU-MIMO where several users share the same time–frequency physical resource block due to their spatial separation. SDMA uses antenna beam direction or angle (using directional antennas) as another dimension in signal space which can be channelized and assigned to different users (Figure 9.118). Orthogonal channels can be assigned only if the angular separation between users exceeds the angular resolution of the directional antenna. In practice, SDMA is often implemented using sectorized antenna arrays. The 360° angular space is divided into a number of sectors and in each sector a highly directional antenna array is used to minimize the inter-user interference. The users within each sector can be served using multiple access schemes such as OFDMA and CDMA. The SDMA system adapts to the mobility of the users by beam-steering or assigning a user to a new channel when the user changes its geographical location [25].

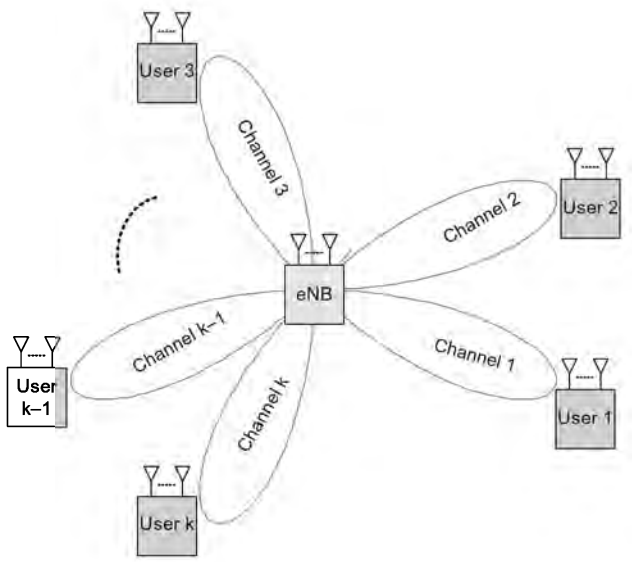


FIGURE 9.118

Illustration of the SDMA concept.

9.13.7 Collaborative MIMO and collaborative spatial multiplexing

Collaborative spatial multiplexing is a virtual MIMO technique where users transmit over the same time–frequency resource unit in the uplink. This type of spatial multiplexing improves the sector throughput without requiring multiple transmit antennas at the mobile station. The received signals from the simultaneous uplink transmission of the users are processed in the base station using maximum likelihood detection techniques. The base station's scheduler groups the users with similar channel conditions and assigns their uplink transmissions to the same time–frequency region. The operation of a collaborative spatial multiplexing scheme is conceptually similar to the regular spatial multiplexing scheme using transmit antennas that are located at different locations (user terminals are geographically apart). The throughput of collaborative spatial multiplexing theoretically increases by the number of users which are scheduled to simultaneously transmit over the same physical resource units.

The collaborative MIMO (Co-MIMO) is an extension of conventional single eNB-based MIMO techniques. It allows multiple base stations to serve one or multiple mobile stations simultaneously over the same radio resource through coordination among the participating base stations. The Co-MIMO is characterized by the following three features:

1. In Co-MIMO, each UE is jointly served by multiple base stations, which is different from conventional MIMO techniques where each UE is only served by a single eNB. Through inter-eNB coordination, the inter-cell interference among these coordinated base stations can be significantly reduced when using low-frequency reuse factors.

2. In Co-MIMO, each eNB can also simultaneously serve multiple mobile stations using the same radio resource in order to increase the system throughput. The co-channel interference among these mobile stations can be minimized through use of SDMA techniques such as beamforming or MU-MIMO schemes.
3. In Co-MIMO, SDMA processing is implemented independently within individual coordinated base stations based on the CSI between the eNB and the mobile stations that are served by the eNB. Therefore, the Co-MIMO limits the amount of information exchange between base stations, which is different from other eNB coordination approaches requiring considerable information exchange.

A central scheduler, that is a logical functional module in the network, is responsible for coordinating a group of base stations, collecting information from the participating base stations, and determining the serving relationship between base stations and mobile stations. The overall complexity of the network operation increases with increasing number of coordinated base stations in Co-MIMO scheme. The Co-MIMO scheme has the advantage of inter-cell interference reduction and spectral efficiency improvement. The inter-cell interference mitigation is realized by turning interference from neighboring cells into useful signals and separating signals for different users via MU-MIMO or SDMA techniques. The spectral efficiency improvement is achieved via interference mitigation and the number of users supported over the same time–frequency resource region. Moreover, the Co-MIMO provides a simple approach to implementation of eNB coordination and a reasonable trade-off between the performance gain and the stringent requirement on the core network traffic and implementation complexity. In practice, the user data delivery to multiple coordinated base stations in a Co-MIMO scheme can be realized in the same manner as the multicast over backhaul is done in the multicast and broadcast services. There is no additional computational complexity in base stations supporting Co-MIMO compared to the base stations which support MU-MIMO techniques [138].

In order to demonstrate the advantages of Co-MIMO, an example system-level simulation was conducted where the collaborative zone is defined over three neighboring sectors belonging to different eNBs as shown in Figure 9.119. Given that the inter-cell interference is more severe in the cell-edge, the evaluation of Co-MIMO performance is performed in the cell-edge area. The cell-edge area is characterized by the cell-edge length in the simulation, as shown in Figure 9.119. We assume a 19-cell collaborative-zone configuration in the simulation as depicted in Figure 9.119. The users are randomly dropped with a uniform distribution in the cell-edge area, one user per sector. The system performance with different sizes of cell-edge area is evaluated by setting the cell-edge length to 150, 100, and 50 m. In each collaborative region, three coordinated base stations communicate with three mobile stations located in their cell-edge areas allocating the same time and frequency resource. Each UE in the cell-edge area will be served by two coordinated base stations, and each eNB serves two mobile stations

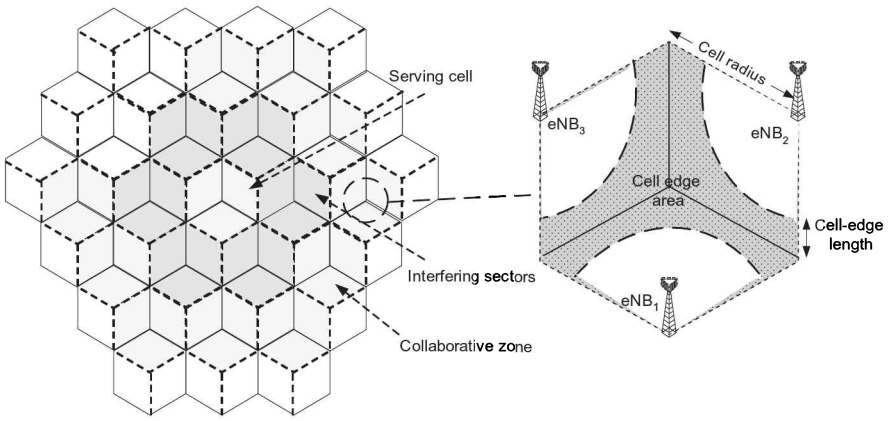


FIGURE 9.119

Example of cell configuration and cell-edge area in a collaborative zone [138].

simultaneously. The precoding matrices for the two served mobile stations are calculated independently within each single sector of a coordinated eNB. Each coordinated eNB transmits with full power and equal power allocation to the mobile stations. A single narrowband sub-carrier is assumed in the system without loss of generality. A conventional SU-MIMO scheme is used as the benchmark for comparison where three mobile stations in the cell-edge area are served independently by different base stations without any coordination in frequency reuse one deployment. As shown in Figure 9.120, even though there remains a certain amount of residual inter-cell interference, the interference power from neighboring cells within the collaborative zone is considerably reduced. If the interference from cells outside of the collaborative zone is taken into consideration, the reduction in overall interference level is less noticeable, and can be further improved by extending the collaborative zone to contain more coordinated base stations. The cumulative distribution function of the cell-edge users within the cell-edge area as a function of cell-edge length for SU-MIMO and Co-MIMO are compared and shown in Figure 9.120. It is shown that the interference power increases with increasing cell-edge length (see Figure 9.119). There is a considerable reduction in the interference power from cells within the Co-MIMO zone when using Co-MIMO techniques compared to the conventional SU-MIMO schemes. The performance gain decreases when comparing the overall interference [138].

9.13.8 Non-linear precoding and dirty-paper coding

In this section, we discuss some advanced non-linear multi-antenna transmission techniques that are not supported in current releases of LTE/LTE-Advanced, but may be considered as viable improvements in future releases of LTE standards. While the following techniques are theoretically proven to achieve higher

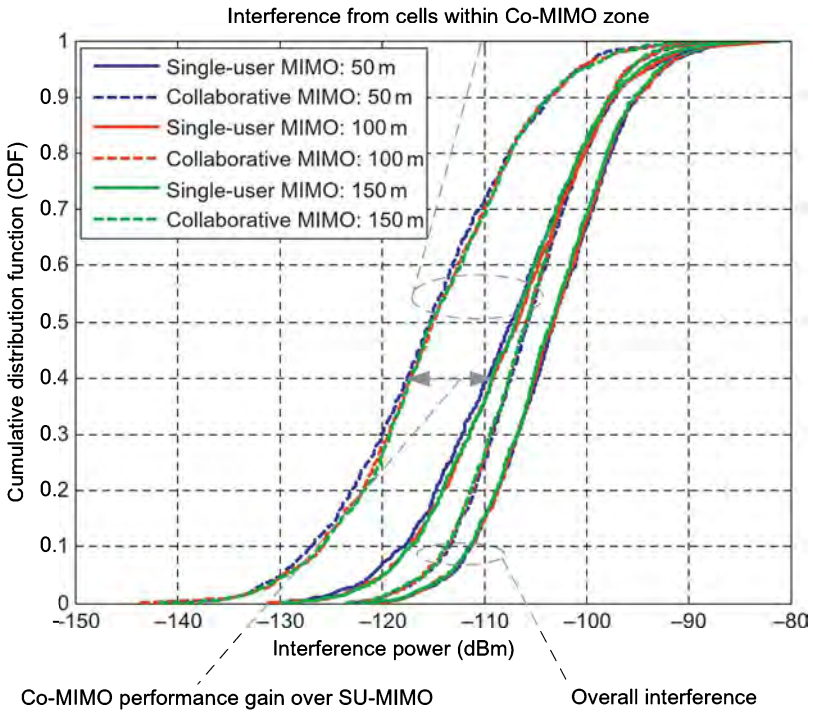


FIGURE 9.120

CDF of interference power of the cell-edge users as a function of cell-edge length [138].

capacities and user throughputs, their implementation complexities have prevented their application in real-life scenarios.

When the channel is frequency-selective, the precoder can also exploit the frequency selectivity and become frequency dependent. For single-carrier systems, non-linear precoding techniques using spatial extensions of the Tomlinson–Harashima precoder can be employed [46]. For multi-carrier systems such as OFDM, frequency-flat precoding techniques can be applied on a per sub-carrier basis. To reduce feedback overhead in OFDM, the CSIT feedback is sampled and interpolated in the frequency-domain. Exploiting the OFDM structure and sub-carrier correlation results in precoders with frequency-dependent eigen-beam directions and frequency-beam-dependent power allocation. In multi-user scenarios, partial CSIT is also highly relevant, since the channel time variation makes it impractical to attain perfect CSIT for all users. The studies have shown that the loss of degrees of freedom due to no CSIT reduces the capacity region of an isotropic vector broadcast channel to that of a scalar one. Imperfect CSIT also severely reduces the growth of the sum-rate broadcast capacity at high SNRs. Some techniques such as opportunistic scheduling, which require only an SNR feedback, can

achieve an optimal throughput growth rate in broadcast channels with a large number of users. With finite-rate feedback, however, the feedback rate needs to be increased with the SNR to achieve the full multiplexing gain. Figure 9.121 illustrates two hypothetical linear and non-linear precoding schemes in a MU-MIMO scenario where $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{\text{user}}})^{\text{T}}$ and the precoding matrix $\mathbf{V}$ is given as follows [46]:

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{N_{\text{user}}})^{\text{T}} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1N_{\text{user}}} \\ v_{21} & v_{22} & \cdots & v_{2N_{\text{user}}} \\ \vdots & \vdots & \ddots & \vdots \\ v_{N_{\text{TX}}1} & v_{N_{\text{TX}}2} & \cdots & v_{N_{\text{TX}}N_{\text{user}}} \end{bmatrix} \tag{9.97}$$

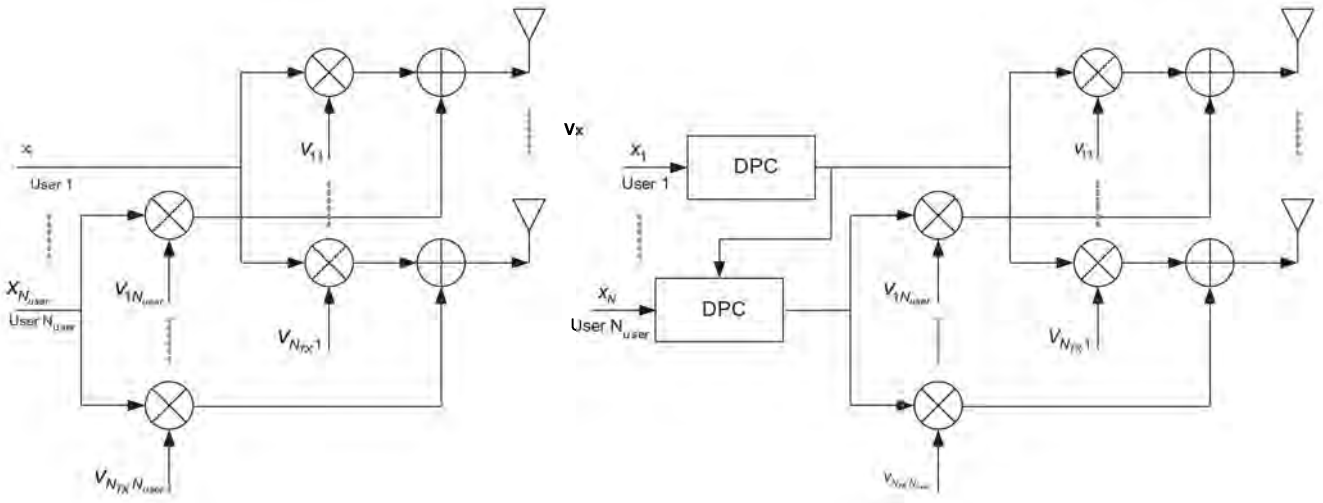
In the non-linear scheme, it is assumed that the interference is known to the transmitter and is mitigated via DPC techniques, as discussed later in this section.

Let us consider a scenario where we have a downlink channel with one base station equipped with N_{TX} transmit antennas and N_{user} users each with one antenna as the simplest form of MU-MIMO Gaussian broadcast channel. A time sample of this channel is characterized by $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$ where $\mathbf{x}$ is the signal vector transmitted in parallel by N_{TX} transmit antennas, $\mathbf{y} = (y_0, y_1, \ldots, y_{N_{\text{user}}-1})^{\text{T}}$ is the vector of signals individually received by N_{user} users, and $\mathbf{n} \sim \mathbb{Z}(0,\mathbf{I})$ is an independent and identically distributed Gaussian noise vector. The channel matrix $\mathbf{H} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{N_{\text{user}}-1})^{\text{H}}$ contains the channel coefficients from the N_{TX} transmit antennas to the N_{user} users, where the row vector $\mathbf{h}_i$ is the channel as seen by the ith user. The input signal power is constrained such that $\text{trace}[\mathscr{E}(\mathbf{xx}^{\text{H}})] \leq P$ where P denotes the total downlink transmitted power or the energy per channel use. Let us further assume that perfect CSI is available at the transmitter and ideal coherent receivers are utilized at the user terminals.

Under the above assumptions, a linear precoding scheme can be defined as a scheme where the transmit signal $\mathbf{x}$ is given by $\mathbf{x} = \boldsymbol{\Theta}\mathbf{u}$ where $\boldsymbol{\Theta} \in \mathbb{C}^{N_{\text{TX}} \times N_{\text{user}}}$ and the vector $\mathbf{u} = (u_1, u_2, \ldots, u_{N_{\text{user}}-1})^{\text{T}}$ contains the user coded symbols. The symbols u_i are independently generated by channel encoders for the users. Without loss of generality, we assume that $E(\mathbf{uu}^{\text{H}}) = \mathbf{I}$ (i.e., normalized signal power), the power constraint implies that $\text{trace}(\boldsymbol{\Theta}\boldsymbol{\Theta}^{\text{H}}) \leq P$. The kth user SINR under linear precoding is given as follows:

$$\text{SINR}_k = \frac{\varphi_{kk} p_k}{1 + \sum_{i \neq k} \varphi_{ki} p_i} \tag{9.98}$$

where $\boldsymbol{\Phi} = [\varphi_{ij}]|_{i=0,1,\ldots,N_{\text{user}}-1; j=0,1,\ldots,N_{\text{TX}}-1}$ denotes a matrix whose entries are defined as $\varphi_{ij} = |\mathbf{h}_i \mathbf{w}_j^{\text{H}}|^2$, $\mathbf{w}_j = \boldsymbol{\theta}_j / \sqrt{p_j}$ is the precoding vector, $\boldsymbol{\theta}_j$ denotes the jth column of $\boldsymbol{\Theta}$, and $p_j = |\boldsymbol{\theta}_j|^2$. It can be shown that assuming Gaussian codes and minimum distance decoding at the receivers, the user rates $R_k = \log_2(1 + \text{SINR}_k)$ are achievable. The transmit matrix $\boldsymbol{\Theta}$ can be designed in several ways. The



Linear precoding structure **Non-linear precoding structure**

FIGURE 9.121

Illustration of linear and non-linear precoding concept [46].

throughput optimization criterion $\max_{\Theta} \sum_k R_k$ is a non-convex optimization problem that yields a maximal sum rate in a MU-MIMO scenario. The constrained maximization with respect to Θ can be replaced by an unconstrained maximization with respect to $\hat{\Theta}$, where the transmit signal is given by the following expression:

$$\mathbf{x} = \sqrt{\frac{P}{\text{trace}(\hat{\Theta}\hat{\Theta}^H)}}\hat{\Theta}\mathbf{u} \tag{9.99}$$

The input constraint is automatically satisfied with equality for any arbitrary $\hat{\Theta}$ [182].

A non-linear precoder, on the other hand, produces the transmitted signal $\mathbf{x}$ in the form of $\mathbf{x} = \gamma\Theta[\mathbf{u} + \delta - \lambda(\mathbf{u}, \delta)]$ where Θ is a transmit matrix such that $\text{trace}(\Theta\Theta^H) \leq P$, δ is a random dithering signal assumed to be known at the transmitter side, and λ is a data-dependent lattice point, i.e., λ is a function of $\mathbf{u}$ and δ. The normalization coefficient γ is defined as follows:

$$\gamma^2 = \frac{P}{\text{trace}(\Theta^H\Theta\mathscr{E}\{(\mathbf{u}+\delta-\lambda)(\mathbf{u}+\delta-\lambda)^H\})} \tag{9.100}$$

where the expectation operator corresponds to the data vectors $\mathbf{u}$ and δ [182].

The capacity of the additive white Gaussian noise channel with interference known at the transmitter was determined by Costa [89]. In his model, as shown in Figure 9.122, the channel input vector $\mathbf{x} = (x_1, x_2, \ldots, x_K)^T$ consisting of K real-valued components is modified by two additive memory-less Gaussian noise and interference sources. The interference source produces the vector $\mathbf{s} = (s_1, s_2, \ldots, s_K)^T$ with normally distributed components $s_i \sim \mathbb{N}(0, \sigma_I^2)$. The interference vector is observed by the encoder before the transmission process starts. The encoder produces the code vector $\mathbf{x} = f(w,\mathbf{s})$. Here $w \in \{1, 2, \ldots, M\}$ is the index of the message that must be conveyed. All messages are equally likely, which means that $\Pr(W = w) = 1/M \quad \forall w$. All code vectors $\mathbf{x}$ should satisfy a power constraint, i.e., $P_{\mathbf{x}} = \mathscr{E}[||\mathbf{x}||^2] \leq P_T$ where P_T denotes the total transmit power. The channel noise vector $\mathbf{z} = (z_1, z_2, \ldots, z_K)^T$ has normally distributed components

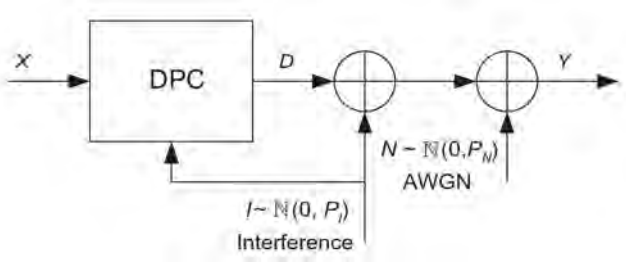


FIGURE 9.122

Costa's dirty paper coding model [89].

$z_i \sim \mathbb{N}(0, \sigma_z^2)$. The channel output vector $\mathbf{y} = (y_1, y_2, \ldots, y_K)^{\mathrm{T}} = \mathbf{x} + \mathbf{s} + \mathbf{z}$ is processed by the decoder and results in the message estimate $\hat{w} = D(\mathbf{y})$. The system has two parameters that determine its performance. For a given total transmit power $\mathrm{P_T}$, and interference and noise variances σ_I^2 and σ_z^2, the information rate $R \sim \log_2 M$ should be as large as possible, while on the other hand the expected error probability $P_e = Pr(\hat{w} \neq w)$ should be as small as possible. The largest rate that is achievable with arbitrarily small expected error probability is called the channel capacity. Costa showed that this capacity is determined only by the transmitter power P_{T} and the variance of the channel noise σ_z^2, i.e., the noise that is unknown to the transmitter. It appears that the channel capacity in this case is determined as $C = 0.5 \log_2(1 + P_{\mathrm{T}}/\sigma_z^2)$ bits/transmission. Since the interference is known to the transmitter a priori, this signaling method is known as writing on dirty paper [89]. Interference which is known to the transmitter does not cause any loss in terms of capacity.

The above concept can be generalized to the case of MU-MIMO when the interference vector is known to the transmitter a priori. Without loss of generality, let us consider the transmission model shown in Figure 9.123, in which a hypothetical base station is precoding and transmitting signals to three mobile stations each equipped with one receive antenna. Let $\mathbf{h}_k = (h_k^1, h_k^2, \ldots, h_k^{N_{\mathrm{TX}}})^{\mathrm{T}}$ denote the channel vector as seen by the kth mobile station. The corresponding precoding vector at the transmitter is represented by $\mathbf{v}_k = (v_1^k, v_2^k, \ldots, v_{N_{\mathrm{TX}}}^k)^{\mathrm{T}}$. The precoding vector $\mathbf{v}_k$ is designed such that it is orthogonal to all $\mathbf{v}_j : j < k$ by subtracting a linear combination of all previous $\mathbf{v}_j$ from $\mathbf{h}_k^*$ (Gram–Schmidt orthogonalization procedure[30]), i.e., $\mathbf{v}_k = \frac{\mathbf{P}_k \mathbf{h}_k^*}{\|\mathbf{P}_k \mathbf{h}_k^*\|}$ where $\mathbf{P}_1 = \mathbf{I}_{N_{TX}}$ and $\mathbf{P}_{k+1} = \mathbf{P}_k(\mathbf{I}_{N_{\mathrm{TX}}} - \mathbf{v}_k \mathbf{v}_k^{\mathrm{H}})$. The objective is to achieve the following $\mathbf{v}_l^{\mathrm{H}} \mathbf{v}_k = 0 \forall l < k, \mathbf{v}_k^{\mathrm{H}} \mathbf{v}_k = 1$, and $\mathbf{h}_k^H \mathbf{v}_j = 0 \ \forall k < j$. The latter criteria are meant to minimize the interference toward a particular mobile station from downlink transmissions to other mobile stations by appropriately selecting the precoding vectors at the transmitter. As a result, the received signal at the kth mobile station is given as follows [46]:

$$y_k = \underbrace{\sum_{i=1}^{k-1} \mathbf{h}_k^{\mathrm{T}} \mathbf{v}_i \mathbf{x}_i}_{\text{eliminated by DPC}} + \underbrace{\mathbf{h}_k^{\mathrm{T}} \mathbf{v}_k \mathbf{x}_k}_{\text{desired signal}} + \underbrace{\sum_{i=k+1}^{N_{user}} \mathbf{h}_k^{\mathrm{T}} \mathbf{v}_i \mathbf{x}_i}_{\text{eliminated by ZF}} + n_k \tag{9.101}$$

[30] In linear algebra, the Gram–Schmidt orthogonalization procedure is defined as a method for orthonormalizing a set of vectors in an inner product space (e.g., the Euclidean space $\mathbb{R}^n$). The Gram–Schmidt procedure takes a finite set of linearly-independent vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \forall k \leq n$ and generates an orthonormal set of vectors. $S' = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ that spans the same k-dimensional subspace of $\mathbb{R}^n$ as the original set. We define the projection operator by $\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}$ where $\langle \mathbf{u}, \mathbf{v} \rangle$ denotes the inner product of the vectors $\mathbf{u}$ and $\mathbf{v}$. This operator projects the vector $\mathbf{v}$ orthogonally onto the line spanned by vector $\mathbf{u}$. The Gram–Schmidt procedure then continues as follows to generate the remaining vectors in the orthonormal set S' [44]: $\mathbf{u}_1 = \mathbf{v}_1$ and $\mathbf{u}_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \mathrm{proj}_{\mathbf{u}_i}(\mathbf{v}_k)$

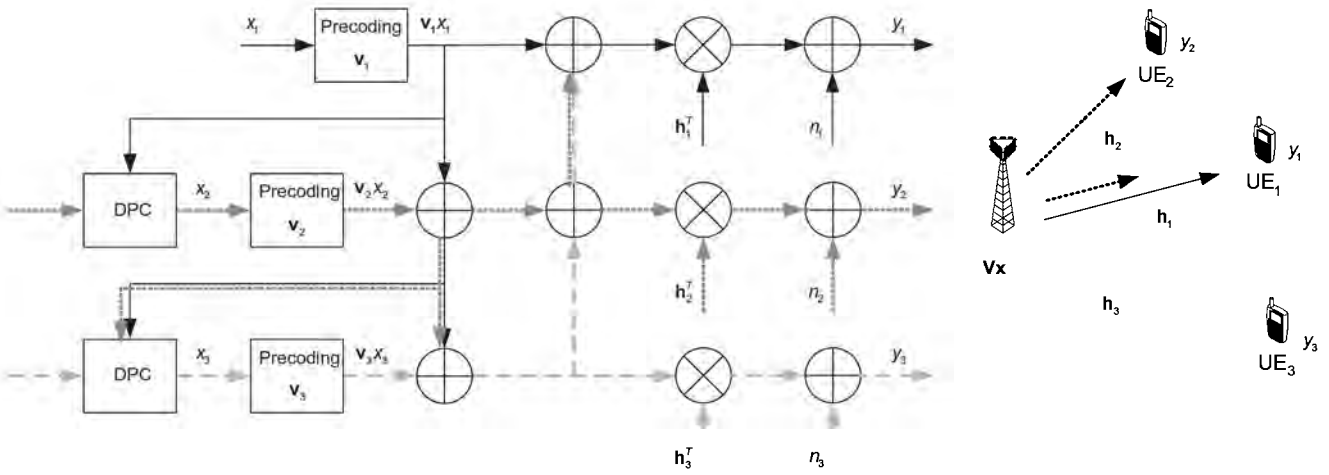


FIGURE 9.123

Sequential encoding with dirty paper coding (DPC) and ZF [46].

The sequential generation of the precoding vectors $\mathbf{v}_k$ through the Gram–Schmidt orthogonalization procedure allows optimized ordering of the users and scheduling of the users even when $N_{\text{user}} > N_{\text{TX}}$ [46]. Note that when $N_{\text{user}} > N_{\text{TX}}$, only N_{TX} users can be scheduled assuming each user has only one receive antenna. The scheduling rule in this case can be summarized as follows:

At each step $k = 1, 2, \ldots, N_{\text{TX}}$:

1. Compute precoding vector $\mathbf{v}_l$ for all users that have not been served yet.
2. For each user l, determine the effective channel $\mathbf{h}_l^{\text{T}}\mathbf{v}_l$ and the resulting sum capacity increment which would be achieved, if user l were scheduled at step k.
3. Allocate the respective resource units to the user who achieves the highest sum capacity increment, considering $C_{\text{sum-capacity}} = \sum_{k=1}^{N_{\text{user}}} C_k$, where C_k is the capacity/rate of the kth user.

9.14 LTE/LTE-Advanced downlink multi-antenna transmission schemes

Having discussed the underlying theory of multi-antenna operation, we now focus on the multi-antenna transmission techniques that have been used in LTE/LTE-Advanced. There are some practical constraints which impact the real-life performance of the theoretical MIMO systems that include limitations in practical cell deployment and antenna configurations, propagation conditions, channel characterization, measurement inaccuracies and latencies, hardware imperfections and implementation margins, and inherent implementation complexity of the multi-antenna schemes in practice. These aspects are often crucial for assessing the performance of a particular multi-antenna transmission scheme in a given propagation channel and configuration, and have been of great importance in the selection of MIMO schemes for LTE/LTE-Advanced.

The advantages of MIMO, i.e., array gain, diversity gain, and multiplexing gain, observed through link-level and system-level simulations assume ideally uncorrelated antennas and full-rank MIMO channel matrices. In this regard, the propagation environment and the antenna design (e.g., the spacing and polarization) play a significant role. If the antenna separation at the eNB and the UE is small relative to the operating wavelength, strong correlation will be observed between the spatial signatures especially in an LoS scenario, limiting the usefulness of spatial multiplexing schemes. However, this can to some extent be circumvented by means of dual-polarized antennas, which provide orthogonality even in LoS environments. Aside from use of such antennas, the condition of spatial signature independence with SU-MIMO can only be satisfied using a rich random multipath propagation environment. In diversity-based schemes, the invertibility of the channel matrix is not required, yet the entries of the channel

matrix should be statistically uncorrelated. Although a greater LoS to non-LoS power ratio will tend to correlate the fading coefficients on the various antennas, this effect will be counteracted by the reduction in fading measured through the LoS component.

Another source of inconsistency between theoretical MIMO gains and practically achieved performance lies in the inability of the receiver and/or the transmitter to acquire reliable knowledge of the propagation channel. At the receiver, channel estimation is typically performed over finite samples of reference signals. In the case of beamforming and MIMO precoding, the transmitter must acquire the CSI from the receiver usually through a limited feedback channel, which causes further degradation to the available CSIT. The potential advantages of MU-MIMO over SU-MIMO include robustness with respect to the propagation environment and the preservation of spatial multiplexing gain, even in the case of UEs with a small number of antennas. However, these advantages rely on the ability of the base station to compute the required beamforming coefficients, which requires accurate CSIT; if no CSIT is available and the fading statistics are identical for all the UEs, the MU-MIMO gains disappear and the SU-MIMO becomes more advantageous. Consequently, one of the challenges in making MU-MIMO practical for cellular applications, and particularly for an FDD system, is developing efficient feedback mechanisms that allow accurate CSI to be estimated at the base station. This requires the use of appropriate codebooks for quantization.

Another issue which arises in the practical implementation of MIMO schemes is the interaction between the physical layer and the scheduling protocol. As noted earlier, in both uplink and downlink cases, the number of UEs which can be scheduled using MU-MIMO is typically limited to the number of the transmit antennas at the base station. Often one may even decide to limit the number of scheduled users to a value strictly less than the number of transmit antennas at the eNB to preserve some degrees of freedom for per-user diversity. As the number of active users typically exceeds the number of scheduled users, a sophisticated selection algorithm must be implemented to identify which set of users will be scheduled for simultaneous transmission over a particular time–frequency resource. This algorithm is not included in the specification, and various approaches are possible where a combination of rate maximization and QoS constraints will typically be considered. It is important to note that the choice of UEs that will maximize the sum rate (the sum over the individual rates for a given subframe) is one that favors UEs exhibiting not only good instantaneous SINR, but also spatial separability among their spatial signatures [14].

A 2×2 antenna configuration for MIMO is assumed as the baseline in the LTE downlink, i.e., two transmit antennas at the base station and two receive antennas at the terminal side [1,2]. Configurations with up to eight transmit or receive antennas are also supported in the specification. Different downlink MIMO modes are supported in LTE/LTE-Advanced, which can be adapted based on channel and traffic conditions, QoS requirements, and UE capability. These include transmit diversity, open-loop spatial multiplexing, codebook-based and

non-codebook-based closed-loop spatial multiplexing, SU-MIMO and MU-MIMO, and precoding. The maximum number of codewords is two irrespective of the number of antennas with fixed mapping between codewords to layers. There is a semi-static switching between SU-MIMO and MU-MIMO per UE.

The objective of dynamic SU-MIMO/MU-MIMO switching is to improve system performance by allowing the network to serve a UE using either SU-MIMO or MU-MIMO. The SU/MU-MIMO switching aspect is important because, depending on the channel and traffic conditions, some UEs may be best served with SU-MIMO, while others benefit from MU-MIMO operation. As channel and traffic conditions may vary from one subframe to another, the dynamic switching aspect is important to optimize system performance. Important factors that influence the suitability of SU-MIMO versus MU-MIMO include the spatial characteristics of the MIMO channel (e.g., potential rank deficiency that could be relaxed by performing MU-MIMO), as well as the UE's typical signal-to-interference-plus-noise ratio conditions. In addition, channel and traffic variations are important factors (e.g., UE mobility or the characteristics of user traffic). MU-MIMO is more suitable when UE mobility is low and traffic conditions are stable as this improves the ability to find a suitable grouping of UEs for MU-MIMO operation.

In spatial multiplexing, up to two codewords can be mapped to different layers (i.e., up to eight layers). One codeword represents the output from one channel coder. The number of spatial layers available for transmission is equal to the rank of the channel matrix.[31] Precoding in the transmitter side is used to support spatial multiplexing. This is achieved by multiplying the signal with a precoding matrix prior to transmission. The optimum precoding matrix is selected from a predefined codebook which is known to both the eNB and UE. The optimum precoding matrix is the one which maximizes the capacity. The UE estimates the channel and selects the optimum precoding matrix. This feedback is provided to the eNB. Depending on the available bandwidth, this information is made available per resource block or group of resource blocks, since the optimum precoding matrix may vary between resource blocks. The network may configure a subset of the codebooks that the UE is able to select from. In the case of UEs with high velocity, the quality of the feedback may deteriorate. Thus, an open-loop spatial

[31]The column rank of matrix $\mathbf{P}$ is the maximum number of linearly independent column vectors of $\mathbf{P}$. The row rank of matrix $\mathbf{P}$ is the maximum number of linearly independent row vectors of $\mathbf{P}$. Equivalently, the column rank of $\mathbf{P}$ is the dimension of the column space of $\mathbf{P}$, whereas the row rank of $\mathbf{P}$ is the dimension of the row space of $\mathbf{P}$. An important property in linear algebra is that for any matrix, the column rank and the row rank are always equal. Therefore, the number of linearly independent rows or columns is simply referred to as the rank of matrix $\mathbf{P}$. It can be easily verified that the rank of $\mathbf{P}$ and $\mathbf{P}^T$ are equal. The rank is also defined as the dimension of the image of the linear transformation $\mathbf{y} = \mathbf{P}\mathbf{x}$. If a linear operator on a vector space with infinite-dimension has finite-dimensional range (i.e., a finite-rank operator), then the rank of the operator is defined as the dimension of the range. The rank of an $M \times N$ matrix is always less than or equal to min (M, N). A matrix is said to be full rank, if its rank is equal to min (M, N); otherwise, the matrix is rank deficient [29].

multiplexing mode is also supported which is based on predefined settings for spatial multiplexing and precoding. In the case of four antenna ports, different precoders are assigned cyclically to the resource elements. The eNB will select the optimum MIMO mode and the precoding configuration. The information is conveyed to the UE as part of the downlink control information on the PDCCH.

LTE Rel-8 supports spatial multiplexing with a maximum of four layers in SU-MIMO. In LTE-Advanced, the maximum number of layers for SU-MIMO was extended to eight in order to increase the peak spectral efficiency. Aside from the increased number of spatial layers, the SU-MIMO in Rel-8 and Rel-10 differ in the use of beamforming or closed-loop precoding. The SU-MIMO in Rel-8 was developed under the codebook-based precoding framework where transmit precoding vectors/matrices at the eNB are restricted within a finite set. This is because the demodulation is based on the cell-specific reference signals that are not precoded. Therefore, the precoding matrix at the eNB has to be explicitly signaled to the UE in the physical downlink control channel to enable MIMO decoding. It is obvious that codebook-based precoding limits the eNB's precoding flexibility, which is particularly important for MU-MIMO that relies on beamforming to mitigate the inter-user interference. The latter inflexibility motivated design of non-codebook-based precoding schemes in LTE-Advanced through the introduction of multi-layer demodulation reference signals. Because the demodulation reference signals are precoded with the same precoding vector/matrix as the user data/control information, the effective composite channel after precoding can be readily measured with demodulation reference signals, obviating the need to explicitly signal the transmit precoder to the UE. Furthermore, the eNB may choose arbitrary precoding vectors/matrices and achieve greatly improved precoding flexibility for both SU-MIMO and MU-MIMO schemes. Although non-codebook-based precoding has a small performance difference relative to codebook-based precoding for SU-MIMO, it is proven to be the key enabling feature for MU-MIMO in LTE-Advanced. The CSI feedback in LTE is based on implicit feedback where the UE reports a set of recommended MIMO transmission metrics (RI/PMI/CQI). The same feedback scheme is applied in LTE-Advanced for both SU-MIMO and MU-MIMO.

The CSI feedback allows downlink transmission to be adapted based on the instantaneous channel conditions. In LTE Rel-8, the CSI measurement is based on cell-specific reference signals, which is also used for data demodulation. In contrast, the CSI measurement in LTE-Advanced is based on a set of newly designed CSI-RS signals, which are low duty-cycle and low-density, and further allow a higher reuse factor compared to cell-specific reference signals. The feedback mechanisms of LTE Rel-8 and LTE Rel-10 are both based on the implicit feedback framework that has been established and tested since early 3GPP releases. The UE measures the downlink channel quality through reference signals and feeds back the CSI in the form of recommended transmission formats, which includes RI, PMI, and CQI metrics. In LTE, the CQI is defined as a set of transport block sizes, each of which translates to a maximum code rate and modulation

order that can be received by the UE at a certain block error rate. As a criterion for testing the CQI report accuracy, when the reported code rate and modulation order is used for actual data transmission, the UE must be able to decode the data with a BLER of less than 0.01.

In order to support MIMO operation, the UE can be configured to report PMI and RI, in addition to CQI reports. In the case of codebook-based spatial multiplexing, precoding at the eNB is applied relative to the phase of the cell-specific reference signals for each antenna port. As a result, if the UE knows the precoding matrices that could be applicable (as defined in the configured codebook), and if it knows the transfer function of the channels from the transmit antenna ports, it can determine which codebook is the most suitable under the current channel conditions and signals this to the eNB. The preferred codebook, whose index constitutes the PMI report, is the precoder that maximizes the aggregate number of data bits which could be received across all layers. The number of resource blocks used to calculate each PMI depends on the feedback mode configured by the eNB as follows (Figure 9.124):

- Wideband PMI: The UE reports a single PMI corresponding to the preferred precoder assuming transmission over the entire system bandwidth. This is applicable for PMI feedback that is configured to be sent periodically on PUCCH, and also for PMI feedback that is sent aperiodically in conjunction with UE-selected sub-band CQI reports on PUSCH.
- Sub-band PMI: The UE reports one PMI for each sub-band over the system bandwidth, where the sub-band size is between 4 and 6 resource blocks depending on the system bandwidth (Table 9.28). This is used in scenarios where PMI feedback is sent aperiodically on PUSCH in addition to a wideband CQI report. If the eNB intends to schedule a wideband data

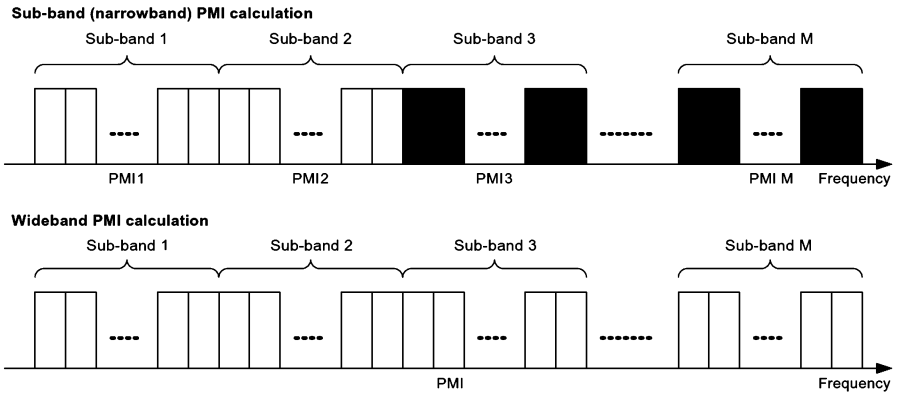


FIGURE 9.124

Illustration of wideband and sub-band PMI concept.

Table 9.28 Sub-band Sizes as a Function of System Bandwidth Applicable to the Best-M Method [5]

System Bandwidth (Number of Resource Blocks)	Sub-band Size *k* (Number of Resource Blocks)	Number of Preferred Sub-bands *M*
6–7	N/A	N/A
8–10	2	1
11–26	2	3
27–63	3	5
64–110	4	6

transmission to a UE in a frequency-selective channel, the precoder used by the eNB can change from one sub-band to another within one transport block.

- UE-selected sub-band PMI (Best-M): The UE selects a set of M preferred sub-bands, each consists of k resource blocks, and reports a single PMI corresponding to the preferred precoder assuming transmission over all of the M selected sub-bands. This is applicable in conjunction with UE-selected sub-band CQI reports on PUSCH. The eNB is not obligated to use the precoder requested by the UE; however, if the eNB selects another precoder, then it has to adjust the transmission format according to the inconsistency with the reported CQI. The eNB may also restrict the set of precoders which the UE may evaluate and report. This is known as codebook subset restriction. It enables the eNB to prevent the UE from reporting precoders which may not be useable, e.g., in some eNB antenna configurations or in correlated fading scenarios. In the case of open-loop spatial multiplexing, codebook subset restriction is equivalent to a restriction on the rank which the UE may report.

9.14.1 Transmit diversity

In LTE/LTE-Advanced, transmit diversity is only defined for two and four transmit antennas and one codeword. In order to maximize diversity gain, the antennas typically need to be uncorrelated, so they need to be sufficiently separated relative to the operating wavelength or have different polarizations. Transmit diversity is useful in a number of scenarios where the SNR is low, or for applications with low delay tolerance. Diversity schemes are also suitable for channels for which no uplink feedback is available (e.g., multimedia broadcast multicast services, physical broadcast channel, and synchronization signals), or when the feedback is not sufficiently accurate such as high-speed scenarios. Transmit diversity is also used as a fallback scheme in LTE/LTE-Advanced when other multi-antenna schemes fail. In LTE/LTE-Advanced, the multi-antenna schemes are independently configured for the physical control channels and the traffic channels, and in the case of traffic channels, it is configured independently per UE. In this

section, we discuss space–frequency block codes and frequency-switched transmit diversity (FSTD) diversity schemes. Transmit diversity may be utilized for transmission of PBCH, PDCCH, and for PDSCH, if the transmit diversity mode is configured for a UE. Another transmit diversity technique which is commonly associated with OFDM is CDD. The CDD is not used in LTE as a diversity scheme, but rather as a precoding scheme for spatial multiplexing on the PDSCH.

LTE Rel-8 supports downlink transmissions on one, two, or four cell-specific antenna ports, each corresponding to one, two, or four cell-specific reference signals, where each reference signal corresponds to one antenna port. An additional antenna port associated with one UE-specific reference signal is available, as well. This antenna port can be used for conventional beamforming, especially in the case of TDD operation. The number of downlink spatial layers was increased to eight through addition of demodulation reference signals and CSI reference signals in LTE-Advanced. Transmit diversity with rank-1 is supported through Alamouti-based space–time codes in the frequency-domain. The encoding operation is performed over space and frequency dimensions so the output block $\mathbf{X}_n$ from the precoder is limited to consecutive data resource elements over a single OFDM symbol. If single-codeword transmission is assumed, the modulated symbols of a single codeword are mapped to all layers. The transmit-diversity schemes for two and four cell-specific antenna ports are also supported. Note that these transmit-diversity schemes can be used for PBCH, PDCCH, and PCFICH. Furthermore, the number of cell-specific antenna ports used to encode the PBCH is the same as the total number of configured cell-specific antenna ports, and those antenna ports are also used for other control channels. Thus, all UEs must support up to four cell-specific antenna ports and the corresponding transmit-diversity schemes. For the case of two antenna ports, the output from the precoder is given as follows:

$$\begin{matrix}\text{Antenna Port } 0\rightarrow \\ \text{Antenna Port } 1\rightarrow\end{matrix} \mathbf{X}_n = \begin{bmatrix} s_{n1} & s_{n2} \\ -s_{n2}^{*} & s_{n1}^{*} \end{bmatrix} \tag{9.102}$$

where the rows correspond to the antenna ports and the columns to consecutive data resource elements over the same OFDM symbol. When four antenna ports are utilized, a combination of SFBC and FSTD is used to provide robustness against the correlation between channels from different transmit antennas and for less-complex UE receiver implementation. The output from the precoder is given as follows:

$$\begin{matrix}\text{Antenna Port } 0\rightarrow \\ \text{Antenna Port } 1\rightarrow \\ \text{Antenna Port } 2\rightarrow \\ \text{Antenna Port } 3\rightarrow\end{matrix} \mathbf{X}_n = \begin{bmatrix} s_{n0} & s_{n1} & 0 & 0 \\ 0 & 0 & s_{n2} & s_{n3} \\ -s_{n1}^{*} & s_{n0}^{*} & 0 & 0 \\ 0 & 0 & -s_{n3}^{*} & s_{n2}^{*} \end{bmatrix} \tag{9.103}$$

The precoder in Eq. (9.103) consists of two SFBC codes which are transmitted on antenna ports 0, 2, and 1, 3, respectively. The reason for distributing a single SFBC code in such an interlaced manner on every other antenna port instead of

consecutive antenna ports is related to the fact that the first two cell-specific antenna ports have a higher reference signal density than the other reference signals; therefore, the channel estimation accuracy may be lower on the third and fourth antenna ports. The above transmit-diversity scheme can be used for all downlink channels other than PHICH. For the latter case, four different ACK/NACK bits are multiplexed using orthogonal codes with a spreading factor of four over a group of four sub-carriers, and the resulting group is repeated three times in the frequency-domain to achieve frequency-diversity gain. To maintain the orthogonality between different codes in each repetition of four sub-carriers, antenna switching is not applied within each repetition. Instead, the set of antennas changes across different repetitions.

9.14.2 Spatial multiplexing

Let us review a few terms before turning our focus on spatial multiplexing schemes in LTE/LTE-Advanced. As mentioned earlier, a spatial layer refers to one of the streams generated by spatial multiplexing, or alternatively a mapping of symbols to the transmit antenna ports. Each layer is identified by a precoding vector of size equal to the number of transmit antenna ports. The rank is defined as the number of transmission layers. A codeword represents an independently encoded data block corresponding to a single transport block delivered to the physical layer by the MAC sub-layer in the transmitter and protected with a CRC. For ranks greater than one, two codewords may be transmitted (as shown in Figure 9.125). Note that the number of codewords is always less than or equal to the number of layers, which in turn is always less than or equal to the number of antenna ports.

In principle, an SU-MIMO spatial multiplexing scheme can either use a single codeword mapped to all the available layers or multiple codewords mapped to one or more spatial layers. The advantage of using one codeword is the obvious reduction in the amount of control signaling required for CQI reporting where only a single value would be needed for all layers, and for HARQ ACK/NACK

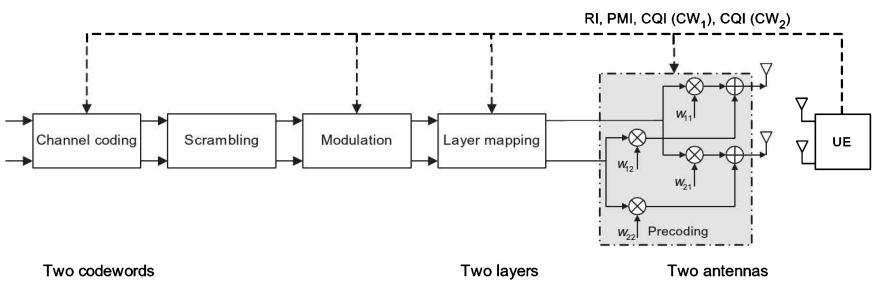


FIGURE 9.125

Example of closed-loop spatial multiplexing with precoding rank = 2 [46].

where only one ACK or NACK would be signaled per subframe per UE. In that case, the MLD receiver is optimal in terms of minimizing the BLER. At the opposite extreme, a separate codeword could be mapped to each of the layers. The advantage of such mapping is that significant gains can be achieved using an SIC receiver at the expense of additional control signaling. An MMSE-SIC receiver is shown to approach the Shannon capacity. Note that an MMSE receiver is more practical for both transmitter structures (i.e., single-codeword and multi-codeword). For LTE/LTE-Advanced, at most two codewords are used even though four layers may be transmitted. The codeword-to-layer mapping is static, since only minimal gains were shown for a dynamic mapping method. Note that in LTE/LTE-Advanced, all resource blocks associated with the same codeword use the same modulation and coding scheme (MCS), even if a codeword is mapped to multiple layers.

The PDSCH transmission modes for open-loop and closed-loop spatial multiplexing use precoding vectors derived from a predefined codebook to form the transmit layers. Each codebook consists of a set of predefined precoding matrices. In the case of closed-loop spatial multiplexing, the UE sends the index of the most desirable entry from a predefined codebook as feedback to the eNB. The preferred precoder is the weighting matrix which would maximize the capacity based on the receiver capabilities. In a single-cell and interference-free environment, the UE will typically indicate the precoder that would result in a transmission with an effective SNR that most closely follow the largest singular value of its estimated channel matrix. An example algorithm by which a UE selects the codebook which is the best fit to the propagation channel at a specific time is as follows [158]:

1. Calculate metric $\zeta(i) = \mathbf{W}_i(\mathbf{H}^{\mathrm{H}}\mathbf{H})\mathbf{W}_i^{\mathrm{H}}$ for every element in the codebook, where $\mathbf{H}$ is the channel matrix and $\mathbf{W}_i$ is the ith precoding matrix/vector.
2. Select the codebook element which minimizes the above metric $\mathrm{PMI} = \arg\min_i(\zeta(i))$.
3. Report the index of the selected codebook element (PMI) to the eNB.

Some important properties of the codebooks utilized in the precoding are as follows [14,15]:

- The LTE/LTE-Advanced precoders only perform phase corrections and no amplitude changes. This is to ensure that the power amplifiers connected to each antenna are uniformly loaded. One exception is the use of the identity matrix as the precoder. Although the identity matrix precoder may completely switch off one antenna on one layer, since each layer is still connected to one antenna at constant power, the effect across the layers is still constant modulus toward the power amplifier.
- The LTE/LTE-Advanced precoders were designed to exploit the nested property. The nested property is a method of organizing the codebooks of different ranks so that the lower rank codebook is formed by a subset of the

higher rank codebook vectors. This property simplifies the CQI calculation across different ranks. It ensures that the precoded transmission for a lower rank is a subset of the precoded transmission for a higher rank, thereby reducing the number of calculations required for the UE to generate the feedback. For example, if a specific index in the codebook corresponds to columns 1, 2, and 3 from the precoder $\mathbf{W}$ in the case of a rank-3 transmission, then the same index in the case of rank-2 transmission consists of either columns 1 and 2, or columns 1 and 3 from $\mathbf{W}$.

- The LTE/LTE-Advanced precoders were designed to minimize complex computations. The codebook associated with dual antenna transmission consists of ± 1 and $\pm j$, which eliminates the need for any complex multiplications. The four transmit antenna codebook contains ± 1 and $\pm j$ along with magnitude scaling to enhance the performance.

Figures 9.126 and 9.127 illustrate effect of precoding on the antenna radiation pattern using precoding matrices $\mathbf{W}_0$ and $\mathbf{W}_1$ (closed-loop spatial multiplexing

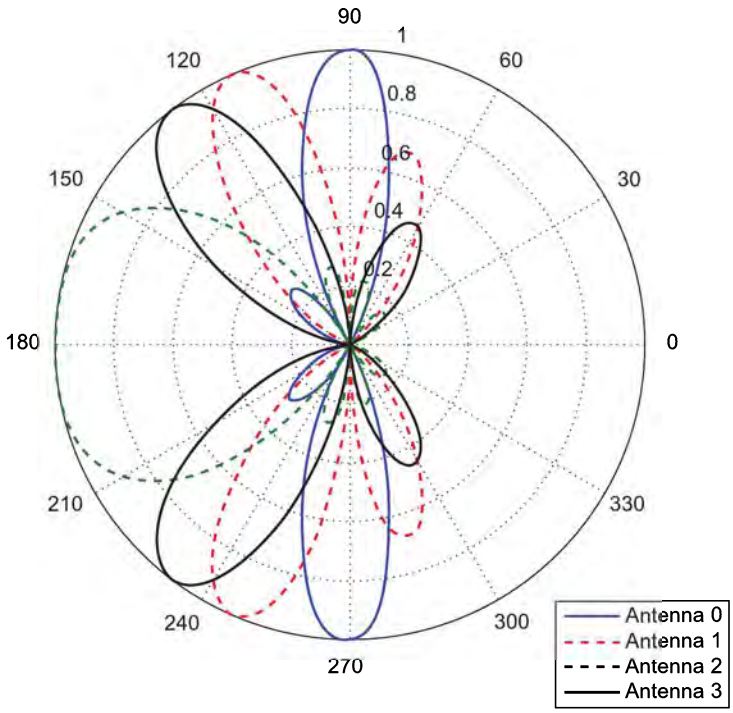


FIGURE 9.126

Effect of precoding on the antenna radiation pattern using precoding matrix $\mathbf{W}_0$ (closed-loop spatial multiplexing with four transmit antennas and rank-4 transmission).

with four transmit antennas and rank-4 transmission), respectively. The precoding matrices are defined as follows:

$$\mathbf{W}_0 = \begin{pmatrix} 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0.5 & -0.5 & -0.5 & 0.5 \end{pmatrix}$$

and

$$\mathbf{W}_1 = \begin{pmatrix} 0.5 & -0.5\text{j} & -0.5 & 0.5\text{j} \\ 0.5\text{j} & 0.5 & 0.5\text{j} & 0.5 \\ -0.5 & -0.5\text{j} & 0.5 & 0.5\text{j} \\ -0.5\text{j} & 0.5 & -0.5\text{j} & 0.5 \end{pmatrix}$$

It can be seen that the eigen-beamforming changes the radiation pattern of the transmit antennas to direct the transmit power in a certain direction consistent with instantaneous propagation channel conditions in order to maximize the signal-to-noise ratio at the UE receiver.

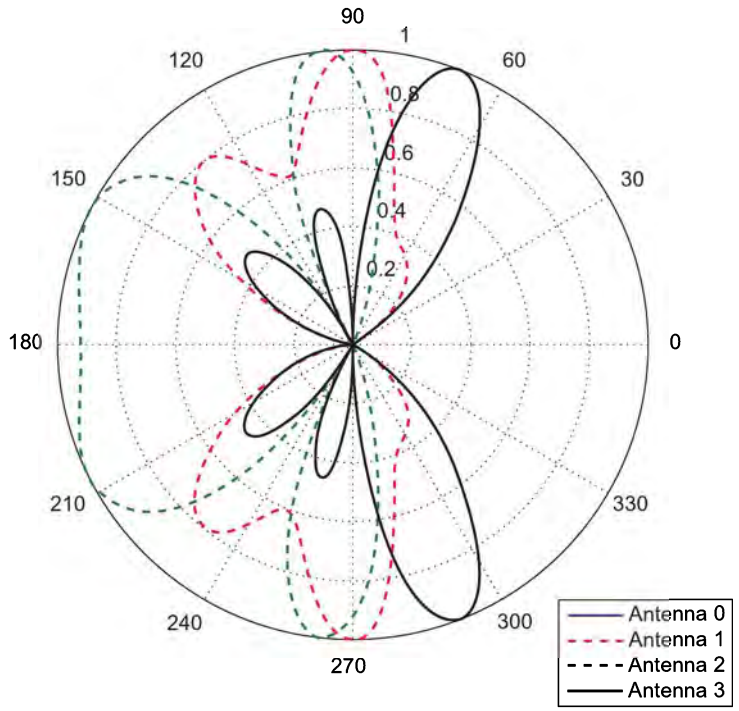


FIGURE 9.127

Effect of precoding on the antenna radiation pattern using precoding matrix $\mathbf{W}_1$ (closed-loop spatial multiplexing with four transmit antennas and rank-4 transmission).

9.14.3 Cyclic delay diversity

In the case of open-loop spatial multiplexing, the UE does not feedback PMI; rather, it sends the rank of the channel. In this mode, if the rank used for PDSCH transmission is greater than one, LTE uses a CDD scheme where the same set of OFDM symbols on the same set of sub-carriers are transmitted from multiple transmit antennas with different delays applied to each antenna. The delay is applied before the cyclic prefix is added, thereby guaranteeing that the delay is cyclic over the FFT size. Adding a time delay is equivalent to applying a phase shift in the frequency-domain. As the same time delay is applied to all sub-carriers, it corresponds to a linear phase shift across the sub-carriers. Each sub-carrier experiences a different beamforming pattern as the non-delayed sub-carrier from one antenna interferes constructively or destructively with the delayed version from another antenna. The diversity effect of CDD originates from the fact that linearly phase-shifted sub-carriers are affected differently in a multipath channel, thus artificially increasing the frequency-selectivity of the channel. Although this approach does not optimally exploit the channel in the way that ideal precoding would be matching the precoding matrix to the eigenvectors of the channel, it does help to ensure that any destructive fading is constrained to individual sub-carriers rather than affecting the entire transport block. This can be particularly beneficial if the channel information at the transmitter is unreliable, e.g., due to the limited feedback or high UE mobility. The fact that the delay is added before the cyclic prefix insertion suggests that any delay value can be used without increasing the overall delay spread of the channel. In contrast, if the delay were added after the addition of the cyclic prefix, then the range of applicable delays would be limited in order to ensure that the delay spread of the delayed symbol is not more than the maximum delay spread of the channel. The equivalence of time-delay and phase-shift implies that the CDD operation can be implemented as a frequency-domain precoder for the affected antennas, where the precoder phase is changed per sub-carrier basis according to a fixed linear function.

In LTE Rel-8, precoding for spatial multiplexing using antenna ports with cell-specific reference signals is only used in combination with layer mapping for spatial multiplexing. Spatial multiplexing supports two or four antenna ports and the set of antenna ports used is $p = 0, 1$ or $p = 0, 1, 2, 3$, respectively with p denoting the antenna port number. The precoding for spatial multiplexing without CDD is defined as $(y_i^0, y_i^1, \ldots, y_i^{N_p-1})^{\mathrm{T}} = \mathbf{W}_i(x_i^0, x_i^1, \ldots, x_i^{N_l-1})^{\mathrm{T}}$ where $\mathbf{W}_i$ is the $N_p \times N_l$ precoding matrix, N_p is the number of antenna ports, N_l denotes the number of layers, $i = 0, 1, \ldots, N_s^p - 1$, and $N_s^p = N_s^l$. For spatial multiplexing, the value of $\mathbf{W}_i$ is selected among the precoder elements in the codebook configured in the eNB and the UE. The eNB can further confine the precoder selection in the UE to a subset of the elements in the codebook using codebook subset restrictions [3].

For large-delay CDD, precoding for spatial multiplexing is defined as $(y_i^0, y_i^1, \ldots, y_i^{N_p-1})^{\mathrm{T}} = \mathbf{W}_i\mathbf{D}_i\mathbf{U}(x_i^0, x_i^1, \ldots, x_i^{N_l-1})^{\mathrm{T}}$ (Figure 9.128), where $\mathbf{D}_i$ and $\mathbf{U}$ are $N_l \times N_l$ matrices supporting CDD as given in Table 9.29 for different numbers

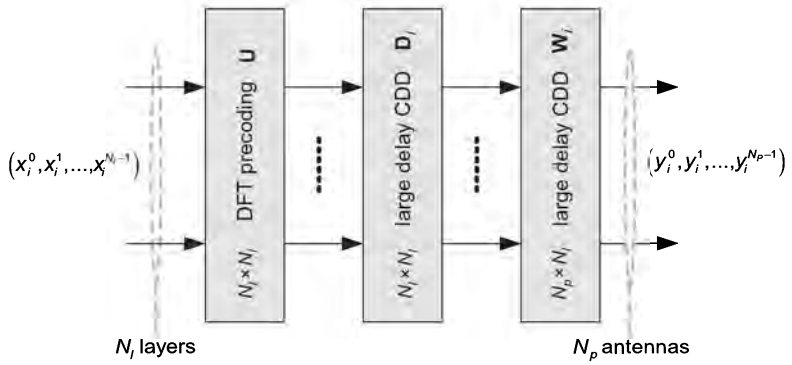


FIGURE 9.128

Open-loop spatial multiplexing with N_p antenna ports and N_l transmission layers [3].

Table 9.29 Large-Delay Cyclic Delay Diversity [3]

Number of Layers N_l	U	D_i
2	$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & e^{-j2\pi/2} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & e^{-j2\pi i/2} \end{bmatrix}$
3	$\frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-j2\pi/3} & e^{-j4\pi/3} \\ 1 & e^{-j4\pi/3} & e^{-j8\pi/3} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-j2\pi i/3} & 0 \\ 0 & 0 & e^{-j4\pi i/3} \end{bmatrix}$
4	$\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & e^{-j2\pi/4} & e^{-j4\pi/4} & e^{-j6\pi/4} \\ 1 & e^{-j4\pi/4} & e^{-j8\pi/4} & e^{-j12\pi/4} \\ 1 & e^{-j6\pi/4} & e^{-j12\pi/4} & e^{-j18\pi/4} \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-j2\pi i/4} & 0 & 0 \\ 0 & 0 & e^{-j4\pi i/4} & 0 \\ 0 & 0 & 0 & e^{-j6\pi i/4} \end{bmatrix}$

of layers. For two antenna ports, the precoder is selected according to $\mathbf{W}_i = \mathbf{C}_1$ where $\mathbf{C}_1$ denotes the precoding matrix corresponding to precoder index zero in Table 9.30. For four antenna ports, the UE may assume that the eNB cyclically assigns different precoders to different vectors $(x_i^0, x_i^1, \ldots, x_i^{N_l-1})^T$ on the physical downlink shared channel. A different precoder is used every N_l vectors, where N_l denotes the number of transmission layers in the case of spatial multiplexing. In particular, the precoder is selected according to $\mathbf{W}_i = \mathbf{C}_k$, where k is the precoder index given by $k = \left(\lfloor i/N_l \rfloor \bmod 4\right) + 1 \in \{1, 2, 3, 4\}$, and $\mathbf{C_1}$, $\mathbf{C_2}$, $\mathbf{C_3}$, $\mathbf{C_4}$ denote precoder matrices. When two antenna ports are configured, the number of codewords is equal to the transmission rank, and codeword n is mapped to layer n.

Table 9.30 Codebook for Transmission on Antenna Ports 0 and 1 [3]

	Number of Layers N_l	
Codebook Index	**1**	**2**
0	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$	$\frac{1}{\sqrt{2}}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$
1	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1 & 1\\1 & -1\end{bmatrix}$
2	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\j\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}1 & 1\\j & -j\end{bmatrix}$
3	$\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-j\end{bmatrix}$	–

For transmission on two antenna ports ($p = 0,1$), the precoding matrix $\mathbf{W}_i$ is selected from Table 9.30 or a subset of it. The open-loop spatial multiplexing may be used when reliable PMI feedback is not available at the eNB, e.g., when the UE moves fast or when the feedback overhead on the uplink is extremely high. The open-loop spatial multiplexing with N_l layers and N_p antenna ports ($N_p \geq N_l$) is illustrated in Figure 9.128. The feedback consists of the RI and the CQI in open-loop spatial multiplexing. In contrast to the closed-loop spatial multiplexing, the eNB only determines the transmission rank, and a fixed set of precoding matrices are applied cyclically across all the scheduled sub-carriers in the frequency-domain. It is noted that in the open-loop spatial multiplexing mode, the transmit diversity scheme is applied when the transmission rank is set to one [3].

9.14.4 Beamforming and MU-MIMO

In LTE/LTE-Advanced, two beamforming techniques are supported for the PDSCH transmission: (1) closed-loop rank-1 precoding; and (2) beamforming with UE-specific reference signals. In closed-loop rank-1 precoding, only a single layer is transmitted, and it can also be seen as a special case of the SU-MIMO spatial multiplexing using codebook-based precoding. In this mode, the UE sends CSI feedback (PMI) to the eNB to indicate suitable precoding to apply for the beamforming. In a beamforming scheme using UE-specific reference signals, the precoding is not restricted to a predefined codebook, thus the UE will rely on DM-RS as opposed to cell-specific reference signals for PDSCH demodulation. Rel-8 of LTE only supported single-layer beamforming and the scheme was extended to support dual-layer beamforming in Rel-9. These transmission modes are used to extend cell coverage by concentrating the eNB power in the directions in which the radio channels provide the strongest paths to the UE. They are typically implemented with an array of closely spaced antenna elements

(or pairs of cross-polarized elements) for creating directional transmissions (Figures 9.129 and 9.130).

The geometric design of the antenna array significantly affects the radiation characteristics. This is discussed here using the example of conventional base

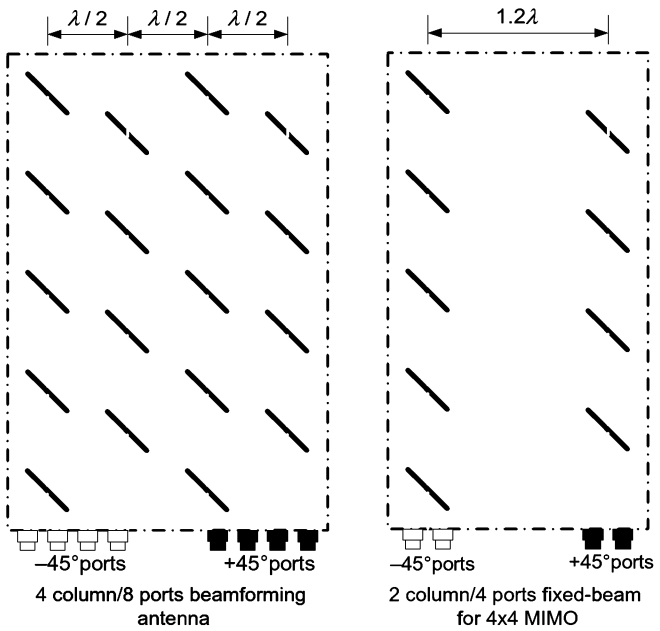


FIGURE 9.129

Practical configuration of beamforming and fixed-beam quad-port cross-polarized physical antenna elements [146].

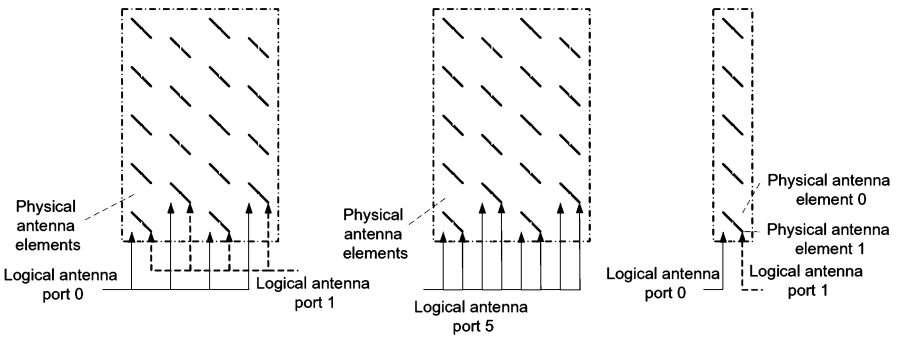


FIGURE 9.130

Example mapping of one/two antenna ports to multiple cross-polarized physical antenna elements [13].

station antennas. At present, conventional passive base station antennas are typically made up of multiple cross-polarized elements. In the *y*-axis, multiple elements are combined in order to satisfy the cell coverage requirements. All elements that have the same polarity radiate the same signal (see Figures 9.127 and 9.128). Particularly relevant for MIMO and beamforming is the arrangement of the cross-polarized elements and the columns in the *x*-axis. In Figure 9.129, the antenna at the left consists of two elements arranged at 90° relative to each other (cross-polarized). Each polarization column represents an antenna element that can transmit a different signal. This makes it possible to transmit four signals with a compact antenna arrangement, such as in 4×4 MIMO or transmit diversity. Analogously, the antenna shown in the figure can radiate four independent signals ($4 \times N$ MIMO), while the antenna at the right can radiate eight independent signals ($2 \times N$ MIMO). The antennas shown in Figure 9.129 can be used for beamforming. Beamforming requires correlated channels, i.e., elements with the same polarization ($+45°$ or $-45°$) must be used. Also, the distance between the columns should be small relative to the wavelength. Beamforming can be carried out with two antenna elements (columns with the same polarization) in the antenna array shown in the left side of Figure 9.129, or with four antenna elements in the layout on the right side. Base station antenna designs are constantly evolving. Active antennas are an important trend that allow seamless integration of beamforming concepts, e.g., by implementing dedicated transceivers for the required number of antenna elements [164].

The mapping of the (logical) antenna ports to physical antenna elements of the eNB may be realized in different ways depending on the multi-antenna scheme that is used. Figure 9.130 shows three example cases where one or two antenna ports are mapped to an array of cross-polarized physical antenna elements. As shown in the figure, the array of physical antennas radiates one signal corresponding to antenna port 5 or two signals corresponding to antenna ports 0 and 1. In the case of dual antenna port transmission, the UE sees two downlink transmissions, one associated with antenna port 0 and another one associated with antenna port 1. The physical antenna mapping and arrangement are transparent to the UE. The composite radiation pattern of the eNB transmit antennas are determined based on the combination of individual antenna radiation pattern, array antenna radiation pattern, and the precoding weights (vectors) that are applied for eigen-beamforming. Two example composite radiation patterns are shown in Figures 9.126 and 9.127.

In the case where UE-specific reference signals are used, due to the fact that no precoding codebook is used (non-codebook-based transmission), an arbitrary number of transmit antennas is theoretically supported; in practical deployments up to eight antennas are typically used (e.g., an array of four cross-polarized pairs). The signals from the correlated antenna elements are phased appropriately so that they add up constructively at the location of the UE. The UE is not aware that it is receiving a precoded directional transmission as opposed to a cell-wide transmission, except that the UE is instructed to use UE-specific reference signals as the phase reference for demodulation. To the UE, the phased array of antennas

appears as one antenna port. In LTE Rel-8, only a single UE-specific antenna port and associated reference signal is defined. As a consequence the beamformed PDSCH can only be transmitted with a single spatial layer. The support of two UE-specific antenna ports was introduced in Rel-9, thus enabling beamforming with two spatial layers with direct one-to-one mapping between the layers and antenna ports.

The UE-specific reference signals in LTE Rel-9 were designed to be orthogonal in order to facilitate separation of the spatial layers at the receiver. The two associated PDSCH layers can then be transmitted to a single user (SU-MIMO) in good channel conditions to increase the individual data rate, or to multiple users (MU-MIMO) to increase the system capacity. The LTE Rel-10 further extended the support for beamforming with an increased number of UE-specific reference signals and spatial layers. When two layers are transmitted over multiple spatial streams, they share the available transmit power of the eNB, thus the increase in data rate comes at the expense of a reduction of coverage. This makes rank adaptation, which is used in the codebook-based spatial multiplexing transmission mode, an important mechanism for the dual-layer beamforming. In this case, the UE is configured to feedback RI in conjunction with the PMI. It should be noted that beamforming in LTE can only be applied to PDSCH and not to the downlink control channels, although the coverage of a traffic channel can be extended through beamforming, the overall cell coverage may still be limited by the control channels. Similar to any other precoding technique, the beamforming schemes in LTE require reliable CSIT, which can ideally be obtained either via UE feedback or based on estimation of the uplink channel condition.

In the single-layer beamforming mode of Rel-8, the UE does not feed back any precoding-related information, and the eNB extracts this information from the uplink, e.g., using DoA estimations.[32] It is worth noting that in this case

[32]If the position of the UE is known, the beamforming weights can be adapted accordingly to optimize transmission for the UE. Therefore, signal classification algorithms such as multiple signal classification (MUSIC) can be used in the base station to determine the DoA for the UE signal and thus determine its location. A uniform linear array (ULA) antenna array is typically used, where the distance d between the individual antennas is the same and $d \leq \lambda/2$. This type of array can be seen as a spatial filtering and sampling in the signal space. Just as the Nyquist criterion applies to sampling a signal over time, the distance here must be $d \leq \lambda/2$ in order to determine the DoA [164].

The MUSIC algorithm estimates the frequency content of a signal or autocorrelation matrix using an eigen-space method. This method assumes that a signal $x(n)$ consists of L complex exponentials in the presence of additive white Gaussian noise. Given an $M \times M$ autocorrelation matrix $\mathbf{R}_{\mathbf{x}}$, if the eigenvalues are sorted in decreasing order, the eigenvectors corresponding to the L largest eigenvalues span the signal subspace. Note that for $M = L + 1$, the result from MUSIC algorithm is identical to that of Pisarenko's method. The general idea is to use averaging to improve the performance of the Pisarenko's estimator. The frequency estimation function for MUSIC is given by $\hat{P}_{\mathrm{MU}}(e^{j\omega}) = \left(\sum_{i=L+1}^{M} \left|\mathbf{e}^{\mathrm{H}}\mathbf{v}_i\right|^2\right)^{-1}$ where $\mathbf{v}_i$'s are the noise eigenvectors and vector $\mathbf{e} = (1, e^{j\omega}, e^{j2\omega}, \ldots, e^{j(M-1)\omega})^{\mathrm{T}}$. The locations of the L largest peaks of the estimation function give the frequency estimates for the L signal components [28].

calibration of the eNB RF paths may be necessary. The dual-layer beamforming mode of LTE Rel-9 provides the possibility for the UE to help the eNB to derive the beamforming precoding weights by sending PMI feedback in the same way as for codebook-based closed-loop spatial multiplexing. For beamforming, this is particularly suitable for FDD deployments where channel reciprocity cannot be exploited as effectively as in the case of TDD. Nevertheless, this PMI feedback can only provide partial CSIT, since it is based on measurements of the cell-specific reference signals which, in beamforming deployments, are usually transmitted from multiple antenna elements with broad radiation patterns; furthermore, the PMI feedback uses the same codebook as for closed-loop spatial multiplexing; therefore it is constrained by quantization granularity. For CQI feedback, the eNB must take into account the fact that in the beamforming modes the channel quality experienced by the UE on the beam-formed PDSCH resources will typically be different than on the cell-specific reference signals based on which the CQI is calculated. The UE-specific reference signals cannot be used for CQI estimation because they are only in existence in the resource blocks assigned to a specific user. It must be noted that beamforming also reduces the interference to neighboring cells. In addition, the beamforming may be coordinated among neighboring eNBs to take advantage of the spatial interference created by beamforming in multiple cells [14,15].

Other algorithms may be used in which the optimum beamforming precoder is determined by channel estimation. In a TDD system, uplink and downlink are on the same frequency, thus the channel characteristics are the same based on the reciprocity principle. That is why a feedback is not needed from the UE when a suitable uplink signal is present that the base station can use to estimate the channel. In the case of LTE TDD, the uplink sounding reference signal can be used.

In LTE Rel-8, a UE configured in MU-MIMO mode assumes that an eNB transmission on the PDSCH would be performed on one layer using one of the rank-1 codebook entries defined for two or four antenna ports. Therefore, the MU-MIMO transmission is limited to a single layer and single codeword per scheduled UE, and the RI and PMI feedback are similar to a SU-MIMO rank-1 transmission. The eNB can simultaneously schedule multiple UEs on the same resource blocks, typically by pairing UEs that report orthogonal PMIs. The eNB can therefore signal to those UEs (using one bit in the DCI format 1D message on PDCCH) that the PDSCH power is reduced by 3 dB relative to the cell-specific reference signals. This assumes that the transmission power of the eNB is equally divided between the two codewords and implies that in practice a maximum of two UEs may be concurrently scheduled in Rel-8 MU-MIMO mode. The CQI reporting for MU-MIMO in Rel-8 does not take into account any interference from transmissions to jointly scheduled users. The CQI values reported are equivalent to those that would be reported for a rank-1 SU-MIMO transmission. In LTE Rel-9, the support for MU-MIMO is enhanced by the introduction of dual-layer beamforming in conjunction with precoded UE-specific reference signals. This eliminates the constraint on the use of the feedback for precoding and

provides more flexibility to utilize other approaches to MU-MIMO user separation, e.g., according to a zero-forcing beamforming criterion. Note that with an increasing number of UEs, zero-forcing beamforming converges to unitary precoding. The use of precoded DM-RSs eliminates the need for explicitly signaling the precoder to the UEs. Instead, the UEs estimate the effective channel, i.e., the combination of the precoding matrix and physical channel. The DM-RS design in Rel-9 allows for up to four UEs to be scheduled for MU-MIMO transmissions [14].

9.15 PDSCH transmission modes

Each PDSCH transmission mode defines a specific set of physical layer procedures and configurations based on which the transport blocks are processed and the complex-valued baseband signals for transmission on the antenna ports are generated. There are 10 transmission modes that have been defined for LTE/LTE-Advanced of which transmission modes 1–7 apply to LTE Rel-8, transmission mode 8 was defined for LTE Rel-9, transmission mode 9 was developed in Rel-10, and transmission mode 10 was introduced in LTE Rel-11. The transmission modes in LTE/LTE-Advanced can be explained as follows [5,14,15]:

- *Transmission mode 1* uses only one transmit antenna and is a SISO or SIMO mode.
- *Transmission mode 2* utilizes transmit diversity and is the default MIMO mode in LTE. It sends the same information via various antennas, whereby each antenna stream uses different space–time coding and different frequency resources. This improves the signal-to-noise ratio and makes transmission more robust. In LTE, transmit diversity is used as a fallback option for some transmission modes, such as when spatial multiplexing cannot be used. Control channels, such as PBCH and PDCCH, are also transmitted using transmit diversity. For two antennas, a frequency-based version of the Alamouti code [58] or SFBC is used, while for four antennas, a combination of SFBC and FSTD is used.
- *Transmission mode 3* supports spatial multiplexing of two to four layers that are multiplexed to two to four antennas, respectively, in order to achieve higher data rates. It requires less UE feedback on the channel condition, i.e., no precoding matrix indicator is included and is used when channel information is missing or when the channel rapidly changes, e.g., for UEs moving at high velocity. In addition to the precoding, the signal is fed to each antenna with a specific delay (CDD), thus artificially creating frequency diversity. For two transmit antennas, a fixed precoding (codebook index 0 is used), while for four antennas, the precoders are cyclically switched.
- *Transmission mode 4* supports spatial multiplexing with up to four layers that are multiplexed to up to four antennas, respectively, in order to achieve higher data rates. To allow channel estimation at the receiver, the base station

transmits cell-specific reference signals, distributed over various resource elements and over various time slots. The UE sends a response regarding the channel condition, which includes information about which precoding is preferred from a predefined codebook. This is accomplished using an index (PMI) to the codebook, which is a table with possible precoding matrices that is known to both sides.

- *Transmission mode 5* is similar to mode 4. It uses codebook-based closed-loop spatial multiplexing; however, one layer is dedicated for each UE.
- *Transmission mode 6* is a special type of closed-loop spatial multiplexing in mode 4. In contrast to mode 4, only one layer is used (rank-1). The UE estimates the channel and sends the index of the most suitable precoding matrix back to the base station. The base station sends the precoded signal via all antenna ports. The precoding in the baseband changes the magnitude and phase of the signals to different antennas and results in beamforming effect. With four transmit antennas, there are 16 different beamforming diagrams. This implicit beamforming effect is distinguished from classical beamforming used in transmission modes 7 and 8, that are aiming at achieving a direct impact on the antenna pattern, e.g., for covering particular areas of a cell.
- *Transmission mode 7* uses UE-specific reference signals. Both data and reference signals are transmitted using the same antenna precoding. Because the UE requires only the UE-specific reference signal for demodulation of the PDSCH, the data transmission for the UE appears to have been received from only one transmit antenna, and the UE does not see the actual number of transmit antennas. Therefore, this transmission mode is also called single antenna port (port 5). The transmission appears to be originated from a single virtual antenna port 5. The LTE standard does not specify any methods for determining the beamforming parameters. Other methods such as beam-switching are also possible. The number of antennas and the antenna architecture are not specified and are implementation specific.
- *Transmission mode 8* specifies dual-layer beamforming which allows the base station to weight two layers individually at the antennas so that beamforming can be combined with spatial multiplexing for one or more UEs. As in mode 7, UE-specific reference signals are used in this mode. Since the same resource elements are used, the reference signals must be coded differently so that the UE can distinguish between them. Because two layers are used, both layers can be assigned to one UE (SU-MIMO), or the two layers can be assigned to two separate UEs (MU-MIMO).
- *Transmission mode 9* was introduced in Rel-10 and uses non-codebook-based precoding which supports up to eight layers using UE-specific reference signals.
- *Transmission mode 10* was introduced in Rel-11 to support CoMP operation and to configure Rel-11 non-zero-power CSI-RS resources and interference measurement resources. This new transmission mode was defined to provide Rel-11 UEs capable of operating on new carrier types, which do not carry CRS, to conduct the CSI measurement and feedback necessary for CoMP

essential functions. More specifically, transmission mode 10 is used for all channel measurements on Rel-11 CSI-RS resources, all interference measurements on Rel-11 interference measurement resources, and PDSCH demodulation based on UE-specific reference signals.

The performance of LTE Rel-8 transmission modes in terms of BLER and throughput as a function of SNR for a number of antenna configurations have been simulated and are shown in Figures 9.131 and 9.132. In these figures, OLSM and CLSM refer to open-loop and closed-loop spatial multiplexing,

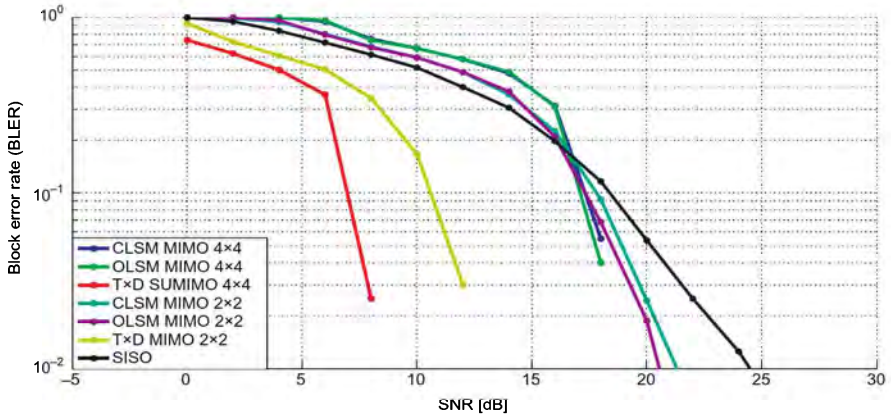


FIGURE 9.131

BLER of LTE transmission modes (modified ITU-VehA 120 km/h,CQI9,800 Subframes) [141,142].

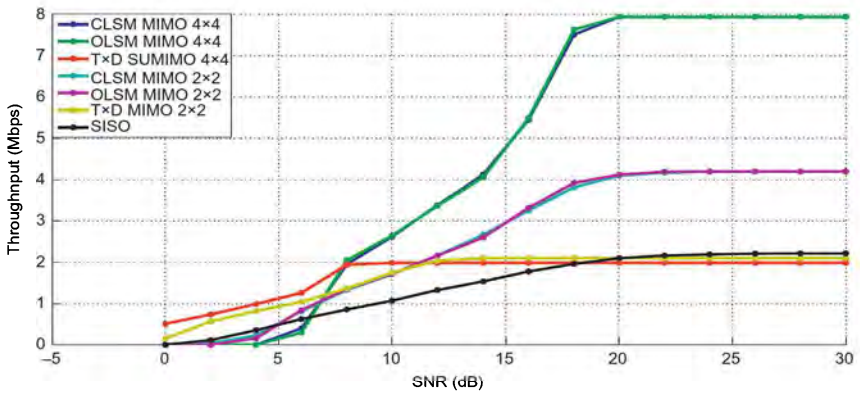


FIGURE 9.132

Throughput of LTE transmission modes (modified ITU-VehA 120 km/h) [141,142].

respectively (transmission modes 3 and 4), and TxD denotes transmit diversity (transmission mode 2) [141,142].

When a UE is configured to decode PDCCH with CRC that has been scrambled by C-RNTI, the UE is required to decode PDCCH and the corresponding PDSCH according to combinations defined in Table 9.31. Note that PDSCH transmission corresponding to these PDCCHs is scrambled by C-RNTI. Similar PDSCH transmission specifications when configured with other UE identifiers such as SPS C-RNTI are given in 3GPP TS 36.213 [5].

The UE can be semi-statically configured via higher layer signaling to receive PDSCH data transmissions signaled via ePDCCH according to one of the transmission modes 1−10. In transmission mode 10, a UE can be configured with scrambling identities $n_{\mathrm{ID}}^{\mathrm{DM\text{-}RS}_i}|_{i=0,1}$ for UE-specific reference signal generation to decode PDSCH according to a detected PDCCH/ePDCCH with CRC scrambled by the C-RNTI with DCI format 2D intended for the UE [5].

A UE in transmission mode 10 can be configured with one or more CSI processes per serving cell by higher layers. Each CSI process is associated with a CSI-RS resource and a CSI-interference measurement (CSI-IM) resource. In transmission mode 10, PUSCH PMI feedback/CQI feedback modes 1-2, 2-2, and 3-1 are supported, if the UE is configured by the eNB to report PMI/RI and the number of CSI-RS ports is greater than one; otherwise, PUSCH PMI feedback/CQI feedback modes $2-0$, and $3-0$ are supported, if the UE is configured without PMI/RI reporting or the number of CSI-RS ports is equal to one. A UE in transmission mode 10 can be configured by higher layers for multiple periodic CSI reports corresponding to one or more CSI processes per serving cell on PUCCH.

In LTE Rel-11 the concept of quasi-collocated antenna ports was introduced in order to support distributed antenna systems. Two antenna ports are quasi-collocated, if the large-scale properties of the channel over which a symbol on one antenna port is conveyed can be inferred from the channel over which a symbol on the other antenna port is conveyed. The large-scale properties include one or more of delay spread, Doppler spread, Doppler shift, average gain, and average delay. The UE does not assume that two antenna ports are quasi-collocated unless specified otherwise.

As a result, when a Rel-11 UE is configured in transmission mode 10 on a given serving cell, it can be configured with up to 4 parameter sets by higher layer signaling in order to decode PDSCH according to a detected PDCCH/ePDCCH with DCI format 2D which is intended for the UE and the given serving cell. The UE then uses the parameter set according to the value of the PDSCH Resource Element Mapping and Quasi-Collocation indicator field in the detected PDCCH/ePDCCH with DCI format 2D for determining the PDSCH resource element mapping and antenna port quasi-collocation, provided that the UE is configured with Type B quasi-collocation. For PDSCH without a corresponding PDCCH/ePDCCH, the UE uses the parameter set indicated in the PDCCH/ePDCCH with DCI format 2D corresponding to the associated semi-persistent scheduling activation for determining the PDSCH resource element mapping and antenna port quasi-collocation [5].

le 9.31 Characteristics of PDSCH Modes when Configured with C-RNTI [5]

nsmission Mode	Release	DCI Format	PDCCH Search Space	Transmission Format of PDSCH	Number of Codewords	CSI Feedback from UE	Number of Layers	Number of Antenna Ports	Type of Reference Signals
1	8	1A	Common and UE-specific	Single-antenna port	1	CQI	1	1 (AP0)	Cell-specific
		1	UE-specific	Single-antenna port			1	1 (AP0)	Cell-specific
2	8	1A	Common and UE-specific	Transmit diversity	1	CQI	2,4	2,4	Cell-specific
		1	UE-specific	Transmit diversity			2,4	2,4	Cell-specific
3	8	1A	Common and UE-specific	Transmit diversity	1	CQI,RI	2,4	2,4	Cell-specific
		2A	UE-specific	Large delay CDD or transmit diversity	2		2,4	2,4	Cell-specific
4	8	1A	Common and UE-specific	Transmit diversity	1	CQI,RI,PMI	2,4	2,4	Cell-specific
		2	UE-specific	Closed-loop spatial multiplexing or transmit diversity	2		2,4	2,4	Cell-specific
5	8	1A	Common and UE-specific	Transmit diversity	1	CQI,PMI	2	2	Cell-specific
		1D	UE-specific	Multi-user MIMO	1		2	2	Cell-specific
6	8	1A	Common and UE-specific	Transmit diversity	1	CQI,PMI	2	2	Cell-specific
		1B	UE-specific	Closed-loop spatial multiplexing using a single transmission layer	1		1	1	UE-specific

(*Continued*

le 9.31 (Continued)

nsmission Mode	Release	DCI Format	PDCCH Search Space	Transmission Format of PDSCH	Number of Codewords	CSI Feedback from UE	Number of Layers	Number of Antenna Ports	Type of Reference Signals
7	8	1A	Common and UE-specific	If the number of PBCH antenna ports is one, single-antenna on AP0; otherwise, transmit diversity	1	CQI	1,2,4	1 (AP0) or 2,4	Cell-specific
		1	UE-specific	Single-antenna	1		1	1 (AP5)	UE-specific
8	9	1A	Common and UE-specific	If the number of PBCH antenna ports is one, single-antenna on AP0; otherwise, transmit diversity	1,2	CQI	1,2,4	1 (AP0) or 2,4	UE-specific
		2B	UE-specific	Dual layer transmission on AP7 and AP8 or single-antenna on AP7 or AP8	1,2		1,2	1 (AP7 or AP8) 2 (AP7 and AP8)	UE-specific
9	10	1A	Common and UE-specific	Non-MBSFN subframe: if the number of PBCH antenna ports is one, single-antenna on AP0; otherwise, transmit diversity MBSFN subframe: single-antenna on AP7	1,2	CQI (PTI, PMI, and RI if configured by the eNB)	Non-MBSFN: 1,2,4 MBSFN: 1	Non-MBSFN: 1 (AP0),2,4 MBSFN: 1 (AP7)	UE-specific

		2C	UE-specific	Up to eight layer transmission on AP7-14 or single-antenna on AP7 or AP8	1,2		Up to 8	Up to 8 (AP7-AP14) 1 (AP7 or AP8)	UE-specific
10	11	1A	Common and UE-specific	Non-MBSFN subframe: if the number of PBCH antenna ports is one, single-antenna on AP0; otherwise, transmit diversity MBSFN subframe: single-antenna on AP7	1,2	CQI (PTI, PMI, and RI if configured by the eNB)	Non-MBSFN: 1,2,4 MBSFN: 1	Non-MBSFN: 1 (AP0),2,4 MBSFN: 1 (AP7)	UE-specific
		2D	UE-specific	Up to eight layer transmission on AP7-14 or single-antenna on AP7 or AP8	1,2		Up to 8	Up to 8 (AP7-AP14) 1 (AP7 or AP8)	UE-specific

9.16 OFDM signal generation

The continuous time-domain signal $s_l^p(t)$ on antenna port p in OFDM symbol l in a downlink slot (assuming 12 sub-carriers per resource block) is defined as follows [3]:

$$s_l^p(t) = \sum_{k=-\lfloor 6N_{RB}^{DL} \rfloor}^{-1} a_{k^-}^p(l) e^{j2\pi k \Delta f (t - N_{CP}(l)T_s)} + \sum_{k=1}^{\lceil 6N_{RB}^{DL} \rceil} a_{k^+}^p(l) e^{j2\pi k \Delta f (t - N_{CP}(l)T)} \quad (9.104)$$

where $0 \le t < (N_{CP} + N_{FFT})T_s$, $k^- = k + \lfloor 12N_{RB}^{DL} \rfloor$, $k^+ = k + \lfloor 6N_{RB}^{DL} \rfloor - 1$, and the parameter $N_{FFT} = 2048$ for $\Delta f = 15$ kHz sub-carrier spacing. The OFDM symbols in a slot are transmitted in increasing order of index l starting at $l = 0$, where OFDM symbol $l > 0$ starts at time $\sum_{l'=0}^{l-1}[N_{CP}(l') + N_{FFT}]T_s$ within the slot. In the case where the first OFDM symbol(s) in a slot use a normal cyclic prefix and the remaining OFDM symbols use an extended cyclic prefix (e.g., MBSFN subframes), the starting position of the OFDM symbols with an extended cyclic prefix are identical to those in a slot where all OFDM symbols use an extended cyclic prefix. Therefore, there will be a part of the time slot between the two cyclic prefix regions where the transmitted signal is not specified.

9.17 Timing advance

In LTE/LTE-Advanced, when a UE attempts to establish a RRC connection with the eNB, it transmits a random access preamble from which the eNB estimates the transmission time (Figure 9.133). Then the eNB transmits a random access response which includes a timing advance command, based on which the UE adjusts the uplink frame transmit timing. Transmission of the uplink radio frame number k from the UE starts $(N_{TA} + N_{TA}^{offset})T_s$ seconds before the start of the corresponding downlink radio frame at the UE where $0 \le N_{TA} \le 20,512$ (samples) and $N_{TA}^{offset} = 0$ for frame structure type 1, and $N_{TA}^{offset} = 624$ (samples) for frame structure type 2. Note that there may be no transmission in some slots in a radio frame. Upon reception of a timing advance command, the UE is required to adjust its uplink transmission timing for the PUCCH, PUSCH, and SRS of the primary cell. The timing advance command indicates the change of the uplink timing

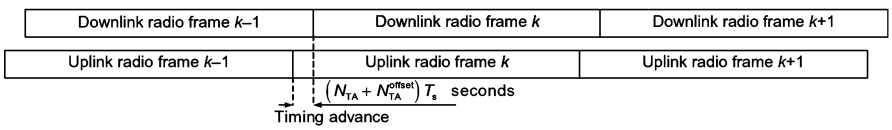


FIGURE 9.133

Relative timing of uplink and downlink radio frames [3,5].

relative to the current uplink timing in multiples of $16T_s$ (samples). The start timing of the random access preamble is discussed in Chapter 10. In LTE Rel-10, the uplink transmission timing for the PUSCH and SRS of a secondary cell is the same as the primary cell.

In the case of random access response, an 11-bit timing advance command parameter $T_A = 0, 1, 2, ..., 1282$ indicates the N_{TA} values where the amount of the timing alignment is given by $N_{TA} = 16T_A$. In other cases, a 6-bit timing advance command parameter $T_A = 0, 1, 2, ..., 63$ indicates the adjustment of the current N_{TA} value where the updated N_{TA} value is given by $N_{TA}^{New} = N_{TA}^{Old} + 16(T_A - 31)$. Note that the adjustment of N_{TA} value by a positive or a negative offset indicates advancing or deferring the uplink transmission time by a given amount, respectively. For a timing advance command received in subframe n, the timing adjustment is applied from subframe $n + 6$ onward. When the UE's uplink PUCCH, PUSCH, and SRS transmissions in subframe n and subframe $n + 1$ are overlapping due to the timing adjustment, the UE transmits the entire subframe n and does not transmit the overlapped portion of subframe $n + 1$. If the received downlink timing changes and the uplink–downlink offset is not compensated or is only partially compensated by the uplink timing adjustment without timing advance command, the UE is expected to change N_{TA} accordingly [3].

After initial uplink synchronization via the random access channel, the UE maintains its uplink frame timing alignment using the timing advance control element sent as part of a MAC PDU on PDSCH that allows the eNB to periodically send timing commands to the UE. The network may also send timing commands as part of a random access response message. The timing advance is a 6-bit number that adjusts timing (advances or delays) relative to the current UE timing, when sent as a MAC control element. The timing advance for random access response is an 11-bit quantity that specifies absolute uplink frame timing relative to the downlink frame. The eNB calculates the desired timing advance based on uplink transmissions from the UE. The eNB also configures the UE's timing alignment timer. If the UE does not receive a timing advance MAC control element before the timing alignment timer expires, it is no longer considered to be uplink-synchronized with the serving cell. When this timer expires, the UE must flush HARQ buffers and reinitialize the HARQ processes. The UE must further release all PUCCH and SRS resources. If any uplink transmission is required, the UE must perform a contention-based random access procedure in order to obtain timing alignment and an uplink grant before it can transmit in the uplink [5].

After the UE has been synchronized to the downlink transmissions from the eNB, the initial timing advance is set by means of the random access procedure in order to establish uplink synchronization. This requires the UE to transmit a random access preamble from which the eNB estimates the uplink timing and responds with an 11-bit initial timing advance command contained in the random access response message. This allows the timing advance to be configured by the eNB with a granularity of 0.52 μs from 0 up to a maximum of 0.67 ms corresponding to a cell radius of 100 km. A cell range of 100 km is sufficient for most

practical rural deployment scenarios. The granularity of 0.52 μs enables the uplink transmission timing to be set with an accuracy within the length of the uplink SC-FDMA symbol cyclic prefix. This granularity is also significantly finer than the length of a cyclic shift of the uplink reference signals. Studies have shown that timing misalignment of up to 1 μs does not cause any significant degradation in system performance due to increased intra-cell interference. Thus the granularity of 0.52 μs is sufficiently fine to allow for additional timing errors arising from the uplink timing estimation in the eNB and the accuracy with which the UE sets its initial transmission timing.

Following the initial timing advance establishment for each UE in the cell, it will then need to periodically update the uplink timing in order to counter changes in the arrival time of the uplink signals at the eNB. Such changes may arise from different causes including the movement of a UE (causing the propagation delay to change at a rate dependent on the relative velocity of the UE and the eNB), sudden changes in propagation delay due to changing multipath profile (such changes typically occur most frequently in dense urban environments as the UEs move around the corners of buildings), oscillator frequency drifts in the UE, and Doppler frequency shift arising from the movement of the UE especially in LoS propagation, resulting in an additional frequency offset of the uplink signals received at the eNB [14].

The update of the timing advance to mitigate the aforementioned effects is performed by a closed-loop mechanism whereby the eNB measures the received uplink timing and issues timing advance update commands to instruct the UE to adjust its transmission timing accordingly relative to its current transmission timing. In deriving the timing advance update commands, the eNB may measure any uplink signal including sounding reference signals, demodulation reference signals, channel quality indicators, HARQ feedback, or uplink data transmissions. In general, wideband uplink signals enable more accurate timing estimation. The advantage of accurate timing estimation has to be traded off with the uplink overhead caused by use of wideband reference signals. In addition, cell-edge UEs are typically power-limited, and thus are bandwidth-limited for a given uplink SINR. The timing estimation accuracy of narrowband uplink signals can be improved through averaging multiple measurements over time and interpolating the resulting power delay profile.

The eNB must balance the overhead of sending regular timing advance update commands to all the UEs in the cell against the UE's ability to transmit as soon as data becomes available in its transmit buffer. The eNB thus configures a timer for each UE, which the UE restarts each time a timing advance update command is received; if the UE does not receive another timing advance update command before the timer expires, it must then consider its uplink to have lost synchronization. In that case, in order to avoid the risk of generating interference to uplink transmissions from other UEs, the UE is not permitted to make another uplink transmission without first transmitting a random access preamble to reinitialize the uplink timing adjustment. One further use of timing advance is to create a

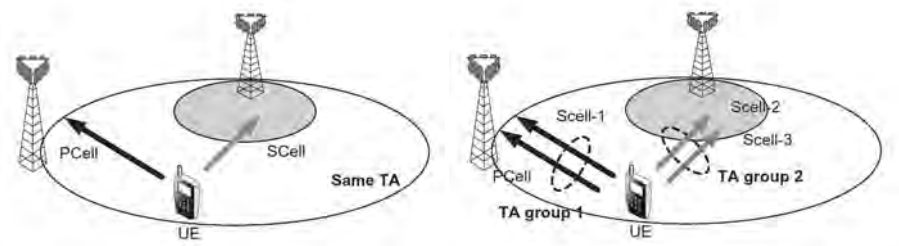


FIGURE 9.134

Illustration of single and multiple timing advance concepts [165].

switching time between uplink reception at the eNB and downlink transmission for TDD and H-FDD operation. This switching time can be generated by applying an additional timing advance offset to the uplink transmissions, to increase the amount of timing advance beyond what is required to compensate for the round-trip propagation delay [14].

Heterogeneous networks (HetNet) comprising macro-cells and small-cell base stations are considered as a key enabler for performance improvement in future cellular networks. LTE Rel-11 has specified carrier aggregation with multiple uplink timing advances and other enhancements to support non-collocated cells, e.g., multiple uplink power control instances or improved sounding reference signals. One of the scenarios of interest in HetNet is the use of remote radio heads (or geographically separated RF transceivers) connected via fiber-optic to a central baseband processing unit. A variation of this scenario, shown in Figure 9.134, would involve macro-cells which overlay small cells where both simultaneously use the same component carriers, requiring inter-cell interference coordination for proper operation (e.g., using an LTE Rel-10 feature known as enhanced inter-cell interference coordination (eICIC) which utilizes almost blank subframes (ABS)). In these scenarios, multiple timing advances are required to support non-collocated cells with carrier aggregation. Assuming synchronization to the macro-cell's PCell is already obtained, the UE has to subsequently synchronize to the SCell of the other site. Therefore, the PCell eNB will request a RACH on the SCell immediately after SCell activation (see Figure 9.134). The RACH request is sent via PDCCH signaling from the PCell. Enhanced MAC control elements are used to signal multiple timing advance values using previously reserved fields in the header.

In the case where several RF carriers require the same timing advance, these carriers will be grouped in what is known as a timing advance group (TAG).[33] The TAGs will be configured by eNB RRC signaling. This will reduce the overall

[33]A timing advance group is a group of serving cells that is configured by RRC signaling and uses the same timing reference cell and the same timing advance value. Primary timing advance group (pTAG) is the timing advance group containing the PCell. Secondary timing advance group (sTAG) is the timing advance group not containing the PCell [7].

complexity and signaling overhead for multiple TAs in case carrier aggregation is being used at each transmission point. For each TAG, a reference carrier for synchronization will be configured as a timing reference.

In carrier aggregation scenarios where two or more component carriers are aggregated to provide wider transmission bandwidths, a UE may simultaneously receive or transmit on one or multiple cells depending on its capabilities. A UE with single timing advance capability for carrier aggregation can simultaneously receive and/or transmit on multiple RF carriers corresponding to multiple serving cells sharing the same timing advance (multiple serving cells grouped in one TAG). A UE with multiple timing advance capability for carrier aggregation can simultaneously receive and/or transmit on multiple RF carriers corresponding to multiple serving cells with different timing advances (multiple serving cells grouped in multiple TAGs). It must be noted that each TAG contains at least one serving cell with configured uplink where the mapping of each serving cell to a TAG is configured by RRC signaling. A UE which does not support carrier aggregation can transmit/receive on a single RF carrier corresponding to one serving cell (one serving cell in one TAG).

LTE-Advanced supports contiguous and non-contiguous aggregation of component carriers where each RF carrier is limited to a maximum of 110 resource blocks in the frequency-domain using the LTE Rel-8/9 OFDM parameters. It is possible to configure a UE to aggregate a different number of RF carriers originating from the same eNB and of possibly different bandwidths in the uplink and the downlink. The number of TAGs that can be configured depends on the TAG capability of the UEs.

The UE is required to adjust uplink transmission timing for PUCCH/PUSCH/SRS of the primary cell based on the received timing advance command for a TAG containing the primary cell. The uplink transmission timing for PUSCH/SRS of a secondary cell is the same as the primary cell, if the secondary cell and the primary cell belong to the same TAG. In the case where a carrier-aggregation-enabled UE receives a timing advance command for a TAG that does not contain the primary cell, then it is required to adjust its uplink transmission timing for PUSCH/SRS of all the secondary cells in the TAG based on the received timing advance command.

9.18 Power control and link adaptation

Power control is a mechanism where the transmit power of the downlink or uplink control or traffic channels are adjusted at the base station or at mobile stations, based on instructions from the serving base station such that with minimal impact on the reliability of the downlink/uplink transmissions and throughput, the inter-user/inter-cell interference among users and base stations are reduced. Therefore, power control can be considered as a link adaptation

mechanism that is utilized for interference mitigation in cellular systems. While increasing the transmit power over a communication link has certain advantages such as higher SNR at the receiver, which reduces the bit error rate and allows higher data rate and results in greater spectral efficiency as well as more protection against signal attenuation over fading channels, a higher transmit power; however, it has several drawbacks including increased power consumption of the transmitting device, reducing the UE battery life, and increased interference to other users in the same or adjacent frequency bands. The following sections describe the power control algorithms that are incorporated in LTE/LTE-Advanced.

9.18.1 Basic concepts

LTE/LTE-Advanced provide uplink power control mechanisms to compensate the effects of path loss, shadowing, fast fading, and implementation loss. The uplink power control is further used to mitigate inter-cell and intra-cell interference, thereby enhancing the overall throughput and reducing the effective UE power consumption. The uplink power control includes open-loop and closed-loop power control. The base station transmits necessary power control information through transmission of power control messages. The parameters of the power control algorithm are optimized on a system-wide basis by the eNB, and are broadcast periodically or trigged by events. The UE provides the necessary information through higher-layer control messages to the serving eNB in order to enable uplink power control. The eNB can exchange necessary information with neighboring base stations through backhaul to support uplink power control to facilitate the handover process.

The power control scheme may not be effective in high mobility scenarios for compensating the effects of a fast fading channel due to variation of the channel impulse response. As a result, the power control is used to mitigate the distance-dependent path loss, shadowing, and implementation loss. The uplink power control takes into consideration the MIMO transmission mode and whether a single user or multiple users are supported on the same resource at the same time. The open-loop power control compensates the channel variations and implementation loss without requiring frequent interactions with the serving eNB. The UE can determine the transmit power based on the transmission parameters sent by the eNB, uplink channel quality, downlink CSI, or the interference knowledge obtained from downlink transmissions. The open-loop power control provides a coarse initial transmit power setting for the mobile station before establishing connection with the base station.

It is perceived that rate control is more efficient than power control under certain conditions [15]. Rate control in principle implies that the power amplifier is always transmitting at full power and therefore is efficiently utilized. On the other hand, power control often results in inefficient utilization of the power amplifier because the transmission power is often less than its maximum. In practice, the

radio-link data rate is controlled by adjusting the modulation scheme and/or the channel coding rate. In good channel conditions, the value of E_b/N_0 at the receiver is high, and the main limitation of the data rate is the bandwidth of the radio link. In such conditions, use of higher-order modulation, e.g., 16QAM or 64QAM, together with a high coding rate, is more appropriate for link adaptation. Similarly, in the case of poor channel conditions, the use of QPSK and low-rate coding is preferred. Link adaptation by means of rate control is referred to as adaptive modulation and coding (AMC).

A power control mechanism takes into consideration the serving eNB target link SINR and/or interference level to other cells/sectors for mitigating inter-cell interference. In order to achieve the target SINR, the serving eNB path loss can be fully or partially compensated based on a trade-off between overall system throughput and cell-edge performance. The mobile station's transmit power is adjusted in order to ensure the level of interference is less than the permissible interference level. The closed-loop power control, on the other hand, compensates channel variations through periodic power-control commands from the serving eNB. The base station measures the uplink CSI and interference level using uplink data and/or control channel transmissions, and sends power control commands to the mobile stations. Upon receiving the power control command from the eNB, the UE adjusts its uplink transmit power. The closed-loop power control is active during data and control channel transmissions.

A UE is expected to maintain the transmit power density (i.e., total transmit power normalized by transmission bandwidth) for each data and control channel below a certain level that is determined by the maximum permissible power level for the UE, emission mask, and other regulatory constraints. In other words, when the number of active logical resource units assigned to a particular user is reduced, the total transmitted power must be reduced proportionally by the UE in the absence of any additional change of power control parameters. When the number of resource blocks is increased, the total transmitted power must be proportionally increased such that the transmitted power level does not exceed the permissible power levels specified by 3GPP and the regulatory specifications [1]. For interference level control, the information about the current interface level of each eNB may be shared among the base stations via backhaul.

The (uplink) transmit power control (TPC) in mobile communication systems is meant to balance the transmitted energy per bit in order to maintain the link quality corresponding to the minimum QoS requirements, to minimize interference to other users in the system, and to minimize the power consumption of the mobile terminal. In achieving these goals, the power control has to adapt to the characteristics of the propagation channel, including path loss, shadowing, and fast fading, as well as overcoming interference from other users both within the same cell and in neighboring cells.

The requirements for uplink interference management in LTE are different from the operational requirements in the uplink of the CDMA systems. In CDMA

systems, the uplink is not orthogonal and the primary source of interference which has to be managed is intra-cell interference among different intra-cell users. The uplink users in a CDMA system share the same time–frequency resources and they all contribute to a rise in interference above thermal noise level at the base station receiver; a phenomenon typically known as rise over thermal (RoT),[34] that ought to be carefully controlled and shared among users. The primary mechanism for increasing the uplink data rate for a given user is to reduce the spreading factor and increase the transmission power accordingly, consuming a larger proportion of the total available RoT in the cell. In contrast, in LTE, the uplink is theoretically orthogonal and intra-cell interference management is consequently less critical than that of a CDMA system.

The main mechanisms for adapting the uplink data rate in LTE are varying the transmit bandwidth and varying the MCS, while the transmit power spectral density (PSD) could typically remain approximately constant for a given MCS. Moreover, in a CDMA system, the power control was primarily designed with continuous transmission in mind for circuit-switched services, while in LTE, different UEs are scheduled at 1 ms transmission time intervals. This is reflected in the fact that power control in CDMA is periodic, while LTE allows for larger power steps which do not have to be periodic, with a minimum loop delay of about 5 ms. With the above considerations in mind, the power control scheme provided in LTE employs a combination of open-loop and closed-loop control, which theoretically requires less feedback than a closed-loop scheme, because the closed-loop feedback is only needed to compensate for cases when the UE's own estimate of the required power setting is not adequate.

A typical mode of operation for power control in LTE involves setting a coarse operating point for the transmit PSD based on path loss estimation. This would provide a suitable PSD setting for reference MCS in the prevailing path loss and shadowing conditions. Faster adaptation can then be applied around the open-loop operating point by closed-loop power control. This can control the interference level and adapt the transmit power setting according to the channel conditions including fast fading. However, due to the orthogonal nature of the LTE uplink, the LTE closed-loop power control does not need to be as fast as in CDMA systems. In fact, the fastest and most frequent adaptation of the uplink transmissions is by means of the uplink scheduling grants, which vary the transmission bandwidth and accordingly the total transmit power, together with setting the MCS, in order to attain the desired uplink data rate. With this combination,

[34]In CDMA systems, the rise over thermal (RoT) indicates the ratio between the total power received from wireless sources at a base station and the thermal noise. In order to decode the data at the base station, a minimum SINR at the receiver must be guaranteed. RoT metric is a measure of the uplink loading. By increasing the number of transmitting UEs and their transmit power, the level of interference in the uplink increases. This interference is perceived by the base station as noise adversely affecting the SINR. The base station controls the interference level by adjusting the UE uplink assignments and transmit-power [40].

the power control scheme in LTE in practice provides support for more than one mode of operation [14].

The initial open-loop operating point for the transmit power per resource block depends on a number of factors including the inter-cell interference and cell loading. It can be further divided into two components: (1) a semi-static base-level value comprising a nominal power level that is common for all UEs in the cell (measured in dBm per resource block) and a UE-specific offset; and (2) an open-loop path loss compensation component. Different base levels can be configured for PUSCH data transmissions depending on the scheduling mode (i.e., dynamic or semi-persistent scheduling), which in principle allows different BLER operating points to be used for each scheduling mode. One possible use for different BLER operating points is to achieve a lower probability of retransmission when using semi-persistent scheduling. The UE-specific offset component of the base level enables the eNB to correct the systematic offsets of the UE's transmission power setting. The path loss compensation component is based on the UE's estimate of the downlink path loss, which can be derived from the UE's measurement of RSRP and the known transmit power of the downlink reference signals, which is broadcast by the eNB.

In order to obtain a reasonable indication of the uplink path loss, the UE should filter the downlink path loss estimate with a suitable time window to remove the effect of fast fading but not shadowing. Typical filter lengths are between 100 and 500 ms [14]. For PUSCH and SRS, the degree to which the uplink PSD is adapted to compensate for the path loss can be set by the eNB, on a scale from "no compensation" to "full compensation." This is an important feature of power control in LTE and is known as fractional power control; it is configured by means of a fractional path loss compensation factor, referred to as α. In principle, the combination of the base level and the path loss compensation component allow the eNB to configure the degree to which the UE responds to the path loss. At one extreme, the eNB could configure the base level to the lowest level (−126 dBm) and rely entirely on the UE's path loss measurement to raise the power toward the cell-edge; alternatively, the eNB can set the base level to a higher value, possibly in conjunction with only partial path loss compensation. Disregarding the UE-specific offset, the range of the base level for PUSCH (−126 to +23 dBm per RB) is intended to cover the full range of target SINR values for different degrees of path loss compensation, transmission bandwidths, and interference levels. For example, the highest value of the base level corresponds to the maximum permissible transmission power of an LTE UE and would typically only be used, if path loss compensation was not used. The lowest value of base level for PUSCH −126 dBm is relevant to a case when full path loss compensation is used and the uplink transmission and reception conditions are optimal. In general, the maximum path loss that can be compensated depends on the required SINR and the transmission bandwidth. Note that this assumes full path loss compensation and ignores the dynamic offset [14]. In other

words, LTE defines the setting of the UE transmit power for PUSCH by the following equation [5]:

$$P_{\text{PUSCH}} = \min[P_{T_{\max}}, 10\log_{10} M + P_0 + \alpha P_{\text{L}} + \delta_{\text{MCS}} + f(\Delta_i)]\ [\text{dBm}] \qquad (9.105)$$

where

- $P_{T_{\max}}$ is the maximum allowed transmit power that depends on the UE power class (i.e., UE maximum output power which is typically + 23 dBm for a power class 3 UE).
- M denotes the number of physical resource blocks assigned for uplink transmission.
- P_0 is a cell/UE-specific base level parameter that is signaled through higher-layer signaling.
- α represents the path loss compensation factor, which is a 3-bit cell-specific parameter in the range of [0 1] which is signaled through higher-layer signaling, where $\alpha = 0$ corresponds to "no compensation" and $\alpha = 1$ corresponds to "full compensation."
- P_{L} is the downlink path loss estimate, which is calculated by the UE based on RSRP measurement.
- δ_{MCS} is cell/UE-specific MCS defined in 3GPP specifications for LTE.
- $f(\Delta_i)$ is a UE-specific function which indicates the use of accumulated or absolute correction value and Δ_i is the closed-loop correction value. The UE-specific TPC commands can operate in two different modes: (1) accumulative TPC commands that are available for PUSCH, PUCCH, and SRS; and (2) absolute TPC commands that are available for PUSCH only. In the case of PUSCH, the switch between these two modes is configured semi-statically for each UE by RRC signaling, i.e., the mode cannot be changed dynamically. With the accumulative TPC commands, each TPC command signals a power step relative to the previous level. This is the default mode and is particularly suitable for fine-tuning of the transmission power, and for situations where a UE receives power control commands in groups of successive subframes. This mode is similar to the closed-loop power control in CDMA systems.

In the above equation $P_0 = \alpha(\text{SNR}_0 + P_{\text{N}}) + (1 - \alpha)(P_{T_{\max}} - 10\log_{10} M_0)$ [dBm] where:

- SNR_0 is the open-loop SNR target.
- P_{N} is the noise power per resource block.
- M_0 defines the number of resource blocks for which the SNR target is set with full-power.

Some of the parameters in the above equations are broadcast by the eNB, i.e., the same value for all the users in the cell. The UE sets its initial transmission power based on received parameters from the eNB and path loss calculated by the UE per se. It is worthwhile to note that Δ_i is signaled by the eNB

to any UE after it sets its initial transmit power, i.e., Δ_i has no contribution in the initial setting of the UE transmit power. The expression, based on which a UE sets its initial power, can be obtained from the above equations by ignoring δ_{MCS} and the closed-loop correction factor, thus $P_{PUSCH} = 10\log_{10} M + P_0 + \alpha P_L$ [dBm] where the parameter M denotes the total number of resource blocks scheduled by the eNB. The power assignment for transmission at the UE is performed on the basis of per resource block, and each resource block contains an equal amount of power. Therefore, by neglecting M, the expression used by the UE to assign power to each resource block is given by $PSD_{TX} = P_0 + \alpha P_L$ [dBm/RB].

The above equation represents the transmit SPD of a resource block expressed in dBm/RB. Note that each resource block corresponds to the 180 kHz subchannel. The parameter PSD_{TX} is a helpful means to explain the basic difference between conventional and fractional power control schemes. Setting the path loss compensation factor α to one (full compensation of path loss) leads to conventional power control, whereas the use of values $0 < \alpha < 1$ (fractional compensation of path loss) leads to fractional power control. Note that $\alpha = 0$ (no compensation of path loss) leads to no power control, i.e., all users will use the maximum allowed transmission power.

When $\alpha = 1$, it can be shown that $P_0 = SNR_0 + P_N$ [dBm], thus $PSD_{TX} = P_0 + P_L$ [dBm/RB], taking into account the path loss, the received PSD at the eNB is given by $PSD_{RX} = P_0$ [dBm/RB]. It is important to note that the received PSD at the eNB is equal to P_0; therefore, a conventional power control scheme leads all users to equal SPD. This scheme is widely used in cellular systems which are not using an orthogonal transmission scheme in the uplink, such as CDMA-based systems. One of the advantages of this power control scheme is that it removes the near-far problem which is typical of CDMA systems, since it equalizes the power level of the UEs before receiving at the base station. It can be shown that under these conditions, the received PSD is the same for all users independent of their path loss for a given target SNR.

The fractional power control scheme allows the user terminals' transmissions to be received with variable PSD depending on their path loss, i.e., the user with good channel conditions will be received with high PSD. Using $0 < \alpha < 1$, the PSD is given by $PSD_{TX} = P_0 + \alpha P_L = \alpha(SNR_0 + P_N) + (1 - \alpha)P_{max} + \alpha P_L$ [dBm/RB]. In contrast to conventional power control, which allows full compensation of path loss, fractional power control compensates for a fraction of the path loss and this is the reason for the term "fractional power control." The received PSD can be obtained by taking path loss into account and is given by $PSD_{RX} = P_0 + (\alpha - 1)P_L$ [dBm/RB]. The received PSD in the case of a conventional power control scheme results in P_0, whereas the case of a fractional power control scheme has an additional term, where both P_0 and $(\alpha - 1)P_L$ are cell-specific and broadcast to the UEs by the eNB, which means that they are the same for all the UEs. Thus, P_L is the key factor that allows users to be received with different power spectral densities [15].

The fractional path loss compensation factor α can be seen as a tool to trade off the fairness of the uplink scheduling against the total cell capacity. Full path loss compensation maximizes fairness for cell-edge UEs. However, when considering multiple cells together as a system, the use of partial path loss compensation can increase the total system capacity in the uplink, as fewer resources are consumed, ensuring the success of transmissions from cell-edge UEs and less inter-cell interference is caused to neighboring cells. Path loss compensation factors around 0.7–0.8 typically give a close-to-maximal uplink system capacity (typically around 15–25% greater than can be achieved with full path loss compensation) without causing significant degradation to the cell-edge data rate. The target received PSD for a given MCS is reduced as the path loss increases, so that cell-edge UEs cause less inter-cell interference. Inter-cell interference is of particular concern for UEs located near the edge of a cell, as they may disrupt the uplink transmissions in the neighboring cells [14].

Open-loop power control is the ability of the UE transmitter to set its uplink transmit power to a specified value suitable for the eNB receiver. The uplink transmit power is given as $P_{\mathrm{OL}} = \min[P_{\max}, 10\log_{10} M + P_0 + \alpha P_{\mathrm{L}}]$ [dBm], which is set by open-loop power control. The choice of α depends on whether a conventional or fractional power control scheme is used. Using $\alpha = 1$ leads to conventional open-loop power control, while $0 < \alpha < 1$ leads to fractional open-loop power control. The estimate of the path loss is obtained after measuring the RSRP and then the calculation for transmission power is performed based on the above equation.

Closed-loop power control is the ability of the UE to adapt the uplink transmit-power according to the closed-loop correction values received from the eNB, also known as TPC commands. The TPC commands are transmitted by the eNB to the UE, based on the closed-loop target signal-to-interference and noise ratio and the measured received SINR. In a closed-loop power control mechanism, the uplink receiver at the eNB estimates the SINR of the received signal and compares it with the target SINR value. When the received SINR value is below the target SINR, a TPC command is transmitted to the UE to instruct the UE to increase the transmitter power. Otherwise, the TPC command will cause the UE transmitter power to decrease. The LTE closed-loop power control mechanism operates around the open-loop operation point. The UE adjusts its uplink transmission power based on the TPC commands it receives from the eNB when the uplink power setting is performed at the UE using open-loop power control.

For the low-rate PUCCH (carrying HARQ ACK/NACK and CQI reports), full path loss compensation is always applied, because PUCCH transmissions from different UEs are code-division-multiplexed. Full path loss compensation facilitates good control of the interference among different users, and thereby helps to maximize the number of users which can be accommodated simultaneously on PUCCH. A different base level value P_0 is also provided for PUCCH compared to that used for PUSCH, assuming a reference PUCCH format. For SRS, an

additional semi-static offset relative to the PUSCH power operating point may be configured through RRC signaling.

Downlink power control determines the energy per resource element (EPRE) which implies the energy prior to cyclic prefix insertion. The term resource element energy also implies the average energy taken over all constellation points for the modulation scheme in use. Uplink power control determines the average power over a SC-FDMA symbol, on which the physical channel is transmitted.

9.18.2 Downlink power allocation

The cell-specific reference signals are embedded into the overall system bandwidth at certain resource elements. In the frequency-domain, every sixth sub-carrier carries a reference signal. The reference signal pattern is a pseudo-random sequence, whose generation depends on the cell identity and the cyclic prefix. Furthermore, a frequency shift is applied that is based on a modulo-6 operation. The mapping of the pattern is therefore cell-specific, but the spacing is always six sub-carriers. In the time-domain, every fourth OFDM symbol carries reference symbols. Due to their importance, reference signals are the highest powered components within the downlink signal. The power level for the reference signals is signaled within system information to the device. The reference signal power level is cell-specific and is in the range of -60 to $+50$ dBm per 15 kHz. It is a requirement that the LTE base station transmits all reference signals with constant power over the entire bandwidth. The power of all other signal components (synchronization signals, PBCH, PCFICH, PDCCH, PDSCH, and PHICH) is set relative to this value. As mentioned above, there are OFDM symbols that contain reference signals and there are OFDM symbols that do not contain reference signals.

The relative PDSCH power for these symbols is given by two different parameters ρ_A and ρ_B (see Figure 9.135). For the majority of cases, ρ_A corresponds to the parameter P_A that is signaled via higher layers. Only for some special cases such as transmit diversity with four antennas or MU-MIMO, ρ_A is computed differently. Parameter P_A is device specific, is obtained as part of the *RRCConnectionSetup* message, and can take one out of eight different values. In Figure 9.135, it is assumed that P_A is -4.77 dB. Parameter P_B is related to the cell-specific reference signal power and cannot be changed dynamically. It can take one out of four integer values. Depending on the number of transmit antennas (1, 2, or 4), each value corresponds to a certain ratio and thus a power offset. LTE Rel-8/9 networks that are currently deployed worldwide are supporting a 2×2 MIMO baseline configuration. Let us assume $P_B = 3$, in that case, the resource element carrying data in that OFDM symbol where reference signals are present is transmitted with an additional offset of 3 dB compared to

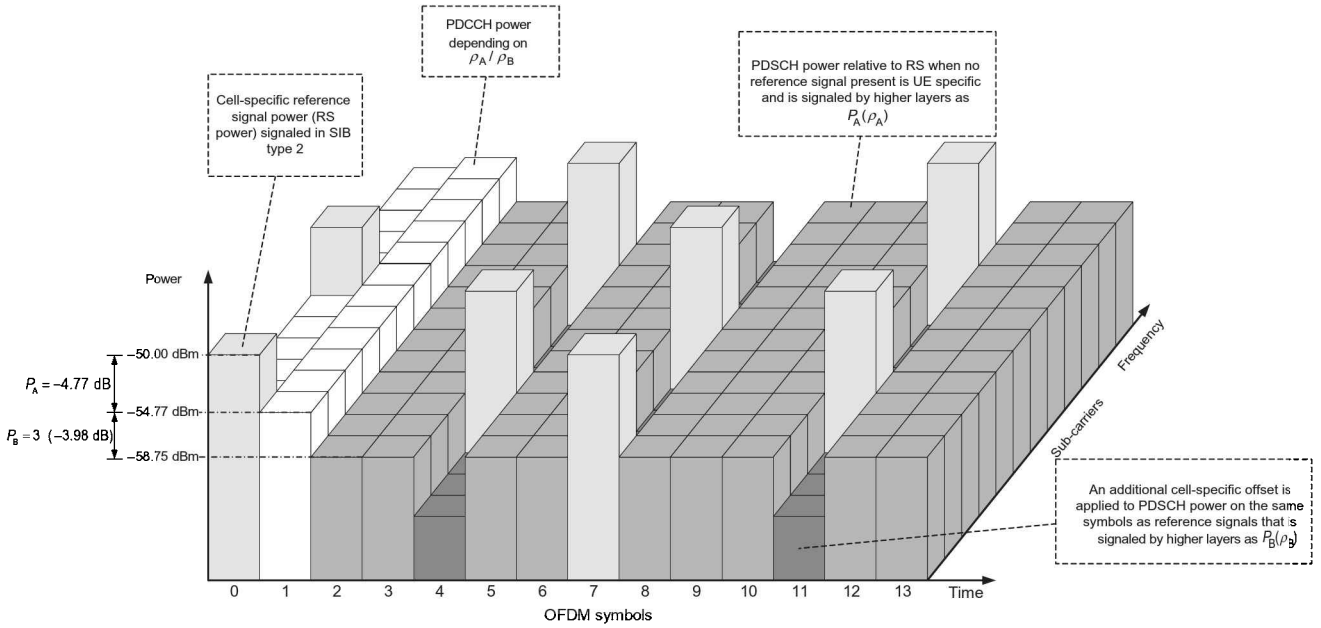


FIGURE 9.135

Example of LTE Rel-8 downlink power allocation [134].

symbols without reference signals. For only one transmit antenna (SISO), $P_B = 3$ translates to -3.98 dB as shown in Figure 9.135. The overall goal is to have a constant power for all OFDM symbols to avoid power variations at the UE receiver. With less PDSCH power given by P_B, the boost of reference signals is compensated, compared to OFDM symbols that do not contain reference signals. The PDSCH power always depends on the resource allocation, i.e., the number of allocated resource blocks. Resource allocation can change from one subframe to another subframe, thus P_A can also change on a subframe-by-subframe basis. While incorporating P_A and P_B, it is ensured that the overall OFDM symbol power remains constant, even when the PDSCH allocation is changed [5].

9.18.3 PUSCH power control

Starting with LTE Rel-10 and in the case of carrier aggregation, it is possible to transmit multiple PUSCH on different component carriers simultaneously, or it is possible to simultaneously transmit PUSCH and PUCCH on the same or different component carriers (this was not allowed in previous releases of LTE). In principle, each physical channel is separately and independently power controlled. However, in the case of concurrent transmission of multiple physical channels from the same terminal, the total transmit power for all physical channels may occasionally exceed the maximum terminal output power setting according to the terminal power class. The basic strategy is to ensure reliable transmission of all control signaling by allocating an appropriate amount of power to the respective physical channels. The remaining power is then assigned to the rest of the physical channels. For each uplink component carrier configured for a terminal, there is also an associated and explicitly configured maximum per-carrier transmit power, which may be different for different component carriers. This may be explained as follows. In many practical scenarios, the terminal will not be scheduled for uplink transmission on all of its configured component carriers, so the terminal should be able to transmit with its maximum output power. As will be seen in this section, the power control of each physical channel explicitly ensures that the total transmit power for a given component carrier does not exceed maximum per-carrier transmit-power; however, separate power-control algorithms do not ensure that the total transmit power for all component carriers does not exceed the maximum terminal output power. The latter limitation will be enforced by a subsequent power scaling applied to the physical channels that are concurrently transmitted. This power scaling is carried out in a way that control signaling channels have higher priority relative to traffic channels during concurrent transmission in the uplink.

If the UE does not simultaneously transmit PUSCH and PUCCH on the serving cell k, i.e., the kth component carrier because component carriers are referred

to as cells in LTE-Advanced, then the UE transmit power $P^k_{\text{PUSCH}}(i)$ for PUSCH transmission in the ith subframe is given as follows:

$$P^k_{\text{PUSCH}}(i) = \min \left\{ \underbrace{P^k_{\max}(i)}_{\substack{\text{Maximum Power per} \\ \text{Component Carrier}}}, \underbrace{10\log_{10}(M^k_{\text{PUSCH}}(i))}_{\text{Bandwidth Factor}} + \underbrace{P^k_0(j) + \alpha_k(j)P^k_L}_{\substack{\text{Basic Open-loop} \\ \text{Operating Point}}} + \underbrace{\Delta^k_{\text{TF}}(i) + f_k(\Delta_{\text{TPC}}(i))}_{\text{Dynamic Offset}} \right\} \text{[dBm]} \quad (9.106)$$

If the UE transmits PUSCH and PUCCH simultaneously on the serving cell k, then the UE transmit power $P^k_{\text{PUSCH}}(i)$ when transmitting PUSCH in the ith subframe is given as follows:

$$P^k_{\text{PUSCH}}(i) = \min \left\{ \underbrace{10\log_{10}(\hat{P}^k_{\max}(i) - \hat{P}_{\text{PUCCH}}(i))}_{\substack{\text{Maximum available power per} \\ \text{component carrier}}}, \underbrace{10\log_{10}(M^k_{\text{PUSCH}}(i))}_{\text{Bandwidth Factor}} + \underbrace{P^k_0(j) + \alpha_k(j)P^k_L}_{\substack{\text{Basic open-loop} \\ \text{operating point}}} + \underbrace{\Delta^k_{\text{TF}}(i) + f_k(\Delta_{\text{TPC}}(i))}_{\text{Dynamic offset}} \right\} \text{[dBm]} \quad (9.107)$$

In the above equations:

- $P^k_{\max}(i)$ is the configured UE transmit power in the ith subframe for serving cell k, implying that this channel power is calculated and set for every subframe.
- $\hat{P}_{\text{PUCCH}}(i)$ is the linear value of $P_{\text{PUCCH}}(i)$.
- $M^k_{\text{PUSCH}}(i)$ is the number of resource blocks allocated for the UE in the ith subframe and on serving cell k.
- $P^k_0(j)$ is a parameter composed of the sum of a component $P^k_{0\ NOMINAL\_PUSCH}(j)$ provided through higher-layer signaling for $j = 0,1$ and another component $P^k_{0\ UE\_PUSCH}(j)$ that is provided by higher layers for $j = 0,1$ in the kth serving cell.
- P^k_L is the downlink path loss estimated via RSRP measurements by the UE on serving cell k in dB.
- $\Delta^k_{\text{TF}}(i)$ is the MCS-dependent component, referred to as transport format in the LTE specifications, which allows the transmitted power per resource block to be adapted according to the information data rate.
- i denotes the subframe number.
- $j = 0$ or 1.
- $f_k(\Delta_{\text{TPC}}(i))$ denotes a TPC command and $f_k(.)$ represents an accumulation function in the case of accumulative TPC commands.

The above power control equation allows the UE's transmit power to be controlled with a granularity of 1 dB within the range of -40 dBm to $+23$ dBm. The maximum transmission power of a UE may be subject to additional

restrictions, such as the maximum transmission power in a cell may be restricted to a lower level by RRC signaling, or in some configurations, reductions in maximum output power may be applied in order to satisfy in-band and out-of-band emission requirements. For LTE Rel-10 UEs which support carrier aggregation in the uplink, the transmission power of an individual carrier or uplink physical channel may have to be scaled down according to certain rules in order to satisfy the total output power constraints. The UE accuracy in setting total transmission power depends on the length of time since the last uplink transmission has taken place, and the size of the required change in transmission power [15].

It was mentioned above that if the total transmit power of the UE exceeds $\hat{P}_{T_{\max}}(i)$, the UE scales down the power of PUSCH for the serving cell k in the ith subframe $\hat{P}^k_{\text{PUSCH}}(i)$ such that the condition $\sum_k a_i \hat{P}^k_{\text{PUSCH}}(i) \leq (\hat{P}_{T_{\max}}(i) - \hat{P}_{\text{PUCCH}}(i))$ is satisfied where $\hat{P}_{\text{PUCCH}}(i)$ is the linear value of $P_{\text{PUCCH}}(i)$, $\hat{P}^k_{\text{PUSCH}}(i)$ is the linear value of $P^k_{\text{PUSCH}}(i)$, $\hat{P}_{T_{\max}}(i)$ is the linear value of the UE total configured maximum output power $P_{T_{\max}}$ in the ith subframe, and $0 \leq a_i \leq 1$ is the scaling factor of $\hat{P}^k_{\text{PUSCH}}(i)$ for serving cell k. In a hypothetical scenario, when the UE is scheduled to transmit PUSCH containing UCI on the jth serving cell and transmit PUSCH without UCI in the remaining serving cells, and if the total transmit power of the UE is going to exceed $\hat{P}_{T_{\max}}(i)$, the UE is required to scale $\hat{P}^k_{\text{PUSCH}}(i)$ for the serving cells without UCI in the ith subframe such that the following condition is satisfied [5]:

$$\sum_{k \neq j} a_i \hat{P}^k_{\text{PUSCH}}(i) \leq (\hat{P}_{T\max}(i) - \hat{P}^j_{\text{PUSCH}}(i)) \tag{9.108}$$

where $\hat{P}^j_{\text{PUSCH}}(i)$ is the PUSCH transmit power for the cell with UCI and a_i is the scaling factor of $\hat{P}^k_{\text{PUSCH}}(i)$ for serving cell k containing no UCI. In this case, the power scaling is not applied to $\hat{P}^j_{\text{PUSCH}}(i)$ unless $\sum_{k \neq j} a_i \hat{P}^k_{\text{PUSCH}}(i) = 0$ and the total transmit power of the UE would still exceed $\hat{P}_{T_{\max}}(i)$. Note that values of the scaling coefficients a_i are the same across serving cells when $a_i > 0$, but for certain serving cells a_i may be set to zero. In another hypothetical scenario where the UE has simultaneous PUCCH and PUSCH transmission with UCI on the jth serving cell and PUSCH transmission without UCI in the remaining serving cells, if the total transmit power of the UE would exceed $\hat{P}_{T_{\max}}(i)$, the UE sets $\hat{P}^k_{\text{PUSCH}}(i)$ according to the following criteria [5]:

$$\hat{P}^j_{\text{PUSCH}}(i) = \min[\hat{P}^j_{\text{PUSCH}}(i), \hat{P}_{T_{\max}}(i) - \hat{P}_{\text{PUCCH}}(i)] \tag{9.109}$$

and

$$\sum_{k \neq j} a_i \hat{P}^k_{\text{PUSCH}}(i) \leq (\hat{P}_{T\max}(i) - \hat{P}_{\text{PUCCH}}(i) - \hat{P}^j_{\text{PUSCH}}(i)) \tag{9.110}$$

9.18.4 PUCCH power control

The appropriate level of the received signal power for PUCCH transmission is simply determined by the minimum power necessary to achieve a satisfactory block error rate in decoding of the uplink control signaling information transmitted on PUCCH. However, it is important to take the following considerations into account: (1) In general, decoding performance is not determined by the received signal strength, but rather by the received signal-to-interference-plus-noise ratio, i.e., an appropriate received signal power depends on the interference level at the receiver, where the interference level depends on network deployment and is time-variant since loading of the network varies across time. (2) There are different PUCCH formats which are used to carry a different types of uplink control information. The different PUCCH formats thus carry a different number of information bits per subframe and the information they carry may also have different error-rate requirements. The required received SINR may differ in different PUCCH formats, something that needs to be taken into account when setting the PUCCH transmit power in a given subframe. The power control for PUCCH can be described by the following expression [5]:

$$P_{\text{PUCCH}}(i) = \min\left\{ \underbrace{P^{k}_{\text{max}}(i)}_{\substack{\text{Maximum available} \\ \text{power for PUCCH}}} \;,\; \underbrace{P_0 + P^{k}_{\text{L}}}_{\substack{\text{Basic operating} \\ \text{point}}} + \underbrace{\Delta_{\text{PUCCH-FORMAT}}}_{\substack{\text{Format-dependent} \\ \text{power adjustment} \\ \text{factor}}} + \underbrace{\delta_{\text{PUCCH}}(i)}_{\substack{\text{Dynamic adjustment} \\ \text{based on TPC} \\ \text{commands}}} \right\} \tag{9.111}$$

where:

- $P^{k}_{T_{\text{max}}}(i)$ is the configured UE transmit power in the ith subframe on serving cell k.
- $P_{\text{PUCCH}}(i)$ denotes the PUCCH transmit power in a given subframe.
- P^{k}_{L} represents the downlink path loss as estimated by the terminal based on downlink RSRP.
- P_0 is a cell-specific parameter that is broadcast as part of the system information and can be seen as the desired or target received signal power. As discussed earlier, the required received power depends on the receiver noise and uplink interference level.
- $\delta_{\text{PUCCH}}(i)$ is a UE-specific dynamic correction value, also referred to as a TPC command, included in the PDCCH with DCI format 1/1A/1B/1D/2/2A/2B/2C/2D for the primary cell or sent jointly coded with other UE-specific PUCCH correction values on a PDCCH with DCI format 3/3A whose CRC is scrambled with TPC-PUCCH-RNTI.

In practice, it is not possible to allow P_0 to vary with interference level for the following reasons: (1) the terminal does not read the system information continuously and thus it does not have access to the updated P_0 value; and (2) the uplink path loss estimate is derived from downlink measurements and is not fully accurate due to differences between the instantaneous downlink and uplink propagation conditions, as well as measurement inaccuracies. Thus, in practice, P_0 may reflect the average interference level or the relatively constant noise level.

In order to reflect different SINR requirements for various PUCCH formats, the PUCCH power control expression includes the term $\Delta_{\text{PUCCH-FORMAT}}$ which adds a format-dependent power offset to the transmit power. The power offsets are defined such that a baseline PUCCH format, i.e., the format corresponding to the transmission of a single HARQ ACK/NACK, 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 acknowledgment and used in the case of downlink spatial multiplexing, should have a power offset of 3 dB, reflecting the fact that twice as much power is needed to communicate two acknowledgment bits as opposed to one. 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 $\delta_{\text{PUCCH}}(i)$ in the above expression. The power control commands are accumulative, i.e., each received TPC command increases or decreases the term $\delta_{\text{PUCCH}}(i)$ by a certain amount. The power control commands for PUCCH can be provided to the terminal via two different means: (1) by including a power control command in each downlink assignment (i.e., the terminal receives a TPC command every time it is scheduled on the downlink); and (2) the power control commands can also be sent using DCI format 3/3A that simultaneously provides power control commands to multiple terminals. In practice, such TPC commands are typically transmitted on a regular basis and can be used to adjust the PUCCH transmit power prior to periodic uplink channel-state reports. The power control command carried within the uplink scheduling grant consists of 2 bits, corresponding to four different update steps ($-1, 0, +1, +3$ dB). The same is true for the power control command carried on DCI format 3/3A designated for power control. On the other hand, when the PDCCH is configured to use DCI format 3, each TPC 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. The reason for including a 0 dB value (no change of PUCCH transmit power) as a power-control step is that the power control command is included in every downlink scheduling assignment and sometimes it is desirable not to update the PUCCH transmit power for each assignment.

9.18.5 SRS power control

The UE sets the transmit power of the sounding reference signal in the ith subframe and the kth component carrier as follows [5]:

$$P^k_{\text{SRS}}(i) = \min[P^k_{\max}(i), P^k_{\text{SRS\_OFFSET}}(m) + 10\log_{10}(M^k_{\text{SRS}}) + P^k_0(j) + \alpha_k(j)P^k_L + f_k(i)] \text{ [dBm]} \quad (9.112)$$

where:

- $P^k_{\max}(i)$ is the configured UE transmit power in the ith subframe on serving cell k.
- $P^k_{\text{SRS\_OFFSET}}(m), m = 0, 1$ is a 4-bit parameter semi-statically configured by higher layers for the serving cell.
- M^k_{SRS} denotes the bandwidth of the SRS transmission in subframe i for serving cell k in number of resource blocks.
- $f_k(i)$ is the current PUSCH power control adjustment state for serving cell k.
- $P^k_0(j)$ and $\alpha_c(j)$ terms are the same as PUSCH power control with $j = 1$.

If the total transmit power of the UE for the sounding reference signal is going to exceed $\hat{P}^k_{\max}(i)$, i.e., the maximum allotted power for the kth component carrier, the UE scales down $\hat{P}^k_{\text{SRS}}(i)$ in subframe i such that $\sum_k a_i \hat{P}^k_{\text{SRS}}(i) \leq \hat{P}^k_{\max}(i)$ is satisfied where $\hat{P}^k_{\text{SRS}}(i)$ denotes the linear value of $P^k_{\text{SRS}}(i)$ and $\hat{P}^k_{\max}(i)$ is the linear value of $P^k_{\max}$ defined in 3GPP TS 36.211 [3] in the ith subframe, and $0 < a_i \leq 1$ are the scaling factor of $\hat{P}^k_{\text{SRS}}(i)$. Note that α_i values are the same across serving cells.

9.18.6 PRACH power control

The UE must be synchronized in the uplink before it can send signaling or data. Given that the UEs within a cell's coverage area are geographically distributed at different distances from the eNB location, each will exhibit a different propagation delay when accessing the network. Without uplink synchronization, subframes transmitted from different UEs would likely overlap and interfere with each other.

To overcome this problem, the UE performs uplink synchronization procedures on the RACH. The process is initiated by sending RACH preambles using the physical random access channel (PRACH). The random access procedure may be contention-based or contention-free. In the contention-based procedure, the UE selects a preamble from an available set of two groups. If two UEs select the same preamble in the same RACH resource, a collision can result that must be resolved before either UE can access the system. Contention-based RACH is generally used during initial access or if the UE has lost uplink synchronization. With contention-free synchronization, the UE uses a dedicated preamble and it is possible that it had already connected to the network and must resynchronize following a handover to a new cell. During the random access procedure, the preamble index, the target preamble received power (denoted by *Preamble_Received_Target_Power* parameter), the corresponding RA-RNTI, and the PRACH resource are signaled by the higher layers. The preamble transmission power P_{PRACH} is determined by $P_{\text{PRACH}} = \min[P^k_{\max}(i),\ Preamble\_Received\_Target\_Power + P^k_L]$ [dBm] where $P^k_{\max}(i)$ is the configured UE transmit power for the ith subframe of the primary cell, and P^k_{L} is the downlink path loss estimate calculated by the UE for the primary cell.

During the RACH procedure, the UE transmits a preamble randomly selected from 64 possible preamble sequences divided into two groups. The group from which the UE selects the preamble depends on the size of the message that it would like to transmit and on the calculated path loss to the eNB. Once the preamble is transmitted, the UE monitors the downlink after three subframes following the end of the preamble transmission for a duration controlled by a configurable window. If no downlink response is detected, then the preamble is retransmitted with a power that is increased by a configurable parameter. This process continues until the maximum number of preamble transmissions is reached. If the preamble is successfully detected by the eNB, then the downlink random access response (RAR) message is sent allocating resources for the UE to transmit its message or enforcing a back-off. In the final step, contention resolution must be performed for a contention-based access attempt. The UE sets its initial power level to the following value: *Preamble_Initial_Received_Target_Power* + *Delta_Preamble* (see Table 9.32 for the permissible values). The parameter *Delta_Preamble* depends on the type of preamble format. The UE selects the first available RACH resource for the transmission. If there is no response within (three subframes + *RA_Response_Window* subframes), the UE assumes that the transmission has been unsuccessful. It selects another RACH preamble and performs the back-off procedure. The power level is increased by the amount *Power_Ramp_Step*. This process can be repeated until the *Preamble_Transmission_Counter* timer expires, in that case, an access failure is declared to the upper layers.

9.18.7 Power headroom

In order to assist the eNB to schedule/allocate the uplink resources to different UEs, it is important that the UE can report its available power headroom to the eNB. The eNB can use the power headroom reports (PHRs) to determine how much uplink bandwidth per subframe the UE can use. This can help to avoid allocating uplink transmission resources to UEs which are unable to properly use them.

The UE PHR MAC control element indicates the power headroom available in the UE. The PHR is encoded in 6 bits with a reporting range from +40 to −23 dB in steps of 1 dB. Positive values indicate the difference between the maximum UE transmit power and current UE transmit power. Negative values

Table 9.32 *Delta_Preamble* Values [8]

Preamble Format	Delta_Preamble Value
0	0 dB
1	0 dB
2	−3 dB
3	−3 dB
4	8 dB

indicate the difference between the maximum UE transmit power and the calculated UE transmit power (Figure 9.136). The UE transmit power is calculated based on whether the UE were to transmit according to the current grant with allocated HARQ and RV configuration. The PHR can be configured to be sent periodically or when the downlink path loss changes by a specified amount. For periodic reporting, a report is trigged by the expiration of the *Periodic_PHR_Timer*, which can be configured with values between 10 ms and infinity. For threshold reporting, a PHR is triggered when the path loss changes by the value of *DL_PathlossChange* parameter (1, 3, 6, or infinity dB), provided that the configurable timer *Prohibit_PHR_Timer* has expired. The *Prohibit_PHR_Timer* starts when a PHR has been transmitted and can have values between 0 and 1000 ms.

A PHR can only be sent in subframes in which a UE has an uplink grant, i.e., the report is related to the subframe in which it is transmitted. Therefore, the PHR is a prediction rather than a direct measurement. The UE cannot directly measure its actual transmission power headroom for the subframe in which the report is transmitted and relies on reasonably accurate calibration of the UE's power amplifier output, especially at high output powers when information about power headroom is critical to the system performance.

There are two different types of PHRs. A Type 1 PHR report indicates the power headroom assuming PUSCH-only transmission on the component carrier, while a Type-2 PHR report assumes joint PUSCH and PUCCH transmission, which is a feature added in LTE Rel-10.

9.18.8 Link adaptation

In cellular systems, the quality of the signal received by a UE depends on the channel quality from the serving cell, the level of interference from other cells, and the noise level. To optimize system capacity and coverage for a given transmission power, the transmitter attempts to adapt (and optimize) the information rate and the transmission format, as well as the multi-antenna mode for each user

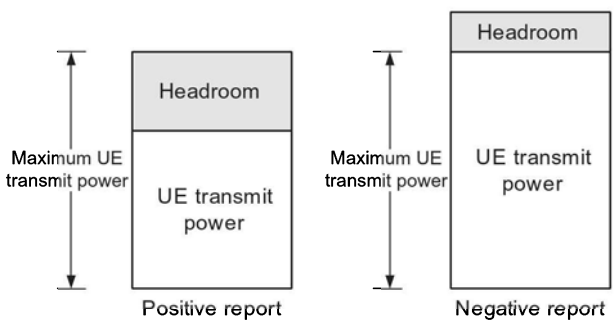


FIGURE 9.136

Power headroom interpretation.

according to the variations in received signal quality at the UE. This is commonly referred to as link adaptation and is typically based on AMC, power control, and adaptation of multi-antenna modes. Low-order modulation such as BPSK and QPSK are more robust and can tolerate higher levels of interference, but provide lower bit rates. High-order modulation such as 16QAM and 64QAM offer higher bit rates, but nonetheless are more prone to errors due to their higher sensitivity to interference, noise, and channel estimation errors; therefore, they are more useful, when the SINR is sufficiently high.

For a given modulation scheme, the code rate can be chosen depending on the channel conditions. A lower code rate can be used in poor channel conditions and a higher code rate can be used when SINR is sufficiently high. The adaptation of the code rate is achieved by applying puncturing to increase the code rate or code repetition to reduce the code rate. In LTE, the modulation and channel coding rates are constant over the allocated frequency resources for a given user, and time-domain channel-dependent scheduling and AMC are supported instead. In addition, when multiple transport blocks are transmitted to one user in a given subframe using multi-layer MIMO, each transport block can use an independent MCS. Furthermore, the UE can be configured to report CQIs to assist the eNB in selecting an appropriate MCS to use for the downlink transmissions. The CQI reports are derived from the downlink received signal quality, typically based on measurements of the downlink reference signals. It is important to note that the reported CQI is not a direct indication of SINR. Instead, the UE reports the highest MCS that it can decode with a BLER not exceeding 10%. Thus the information received by the eNB takes into account the characteristics of the UE's receiver and not just the prevailing channel quality. A UE that is designed with advanced receivers such as MLD or SIC may report a higher channel quality and depending on the characteristics of the eNB's scheduler, can receive a higher data rate [15].

A simple method by which a UE can choose an appropriate CQI value could be based on a set of BLER thresholds, as illustrated in Figure 9.137. The UE will periodically or aperiodically report the CQI value corresponding to the MCS that ensures a BLER less than 10% based on the measured received signal quality.

The CQI metric is generally used in selecting the best MCS under current channel conditions. The CQI measurement typically consists of four basic steps: measuring SINR over the bandwidth of interest (sub-band or wideband), introducing measurement error to the measured SINR, converting the SINR value to discrete CQI values, and finally reporting the CQI using PUCCH or PUSCH. The linear SINR is calculated for each set of resource blocks using the received cell-specific reference signal power and the total interference in every measurement period. The measured linear SINR value for the nth resource block is converted to logarithmic scale as follows $\mathrm{SINR}_{\mathrm{dB}}(n) = 10\log_{10}[\mathrm{SINR}_{\mathrm{linear}}(n)] + \varepsilon_{\mathrm{dB}}$ where $\varepsilon_{\mathrm{dB}}$ is a Gaussian distributed error with zero-mean and parameterized variance, which is added to the measured (ideal) SINR. The SINR values are quantized and converted to discrete CQI values $\mathrm{CQI}_{\mathrm{dB}}(n) = \Delta_Q \lfloor \mathrm{SINR}_{\mathrm{dB}}(n)/\Delta_Q + 0.5 \rfloor$. The CQI is measured at integer multiplies of transmission time interval. The

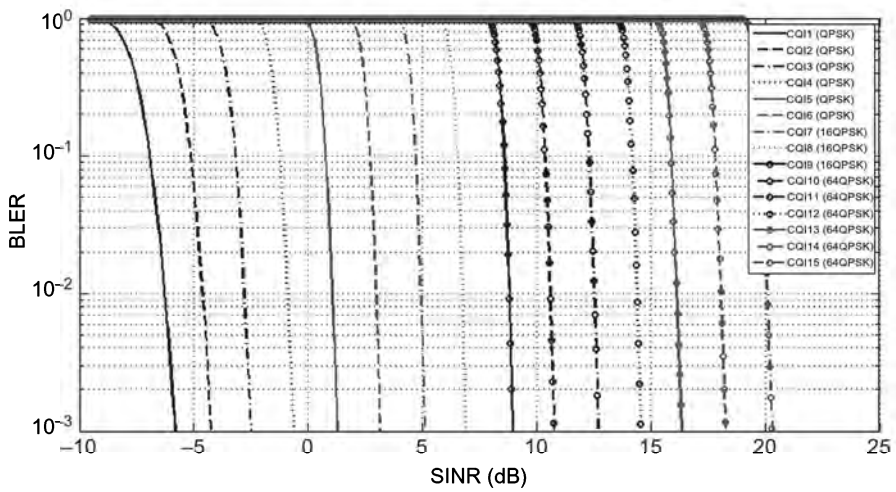


FIGURE 9.137

Link-level BLER as a function of SINR with CQI (MCS) as a parameter [142].

measured CQI values are reported with a certain delay using a specific CQI reporting scheme. It is possible to change the granularity of the basic reporting scheme by changing the number of CQI reports per TTI. Full feedback reporting is done by measuring and reporting individual CQI values for all resource blocks. The minimum granularity is achieved by wideband CQI, which is an average value calculated over all resource blocks.

The granularity of the CQI reports is determined by defining a number of sub-bands $N_{\text{sub-band}}$, each consists of $k_{\text{sub-band}}$ contiguous physical resource blocks. The value of $k_{\text{sub-band}}$ depends on the type of CQI report and is a function of the system bandwidth (see Table 9.28). In each case, the number of sub-bands spans the entire system bandwidth and is given by $N_{\text{sub-band}} = \lceil N_{\text{RB}}^{\text{DL}} / k_{\text{sub-band}} \rceil$ where $N_{\text{RB}}^{\text{DL}}$ is the number of resource blocks over the entire system bandwidth. The CQI reports can be wideband, eNB-configured sub-band, or UE-selected sub-band. In the case of multiple transmit antennas at the eNB, the CQI values may be reported for both codewords.

The eNB scheduler needs to know the number of resource blocks and MCS corresponding to each CQI value in order to properly allocate the resources for each UE. With the modulation scheme obtained from Table 9.33, one would derive a certain range of MCS that can be used for each CQI index. However, in order to obtain a specific MCS and the number of resource blocks, the code rate, given in Table 9.33, is required. Nevertheless, there is not a single formula that would give you a unique value for MCS and the number of resource blocks. Therefore, there is a set of MCS and number of resource blocks that meet the modulation scheme and code rate requirement in Table 9.33. An example is provided in Table 9.34, where the CIF is assumed to be two OFDM symbols for the

Table 9.33 CQI Values and the Corresponding Code Rates [5]

CQI Index	Modulation Scheme	Coding Rate × 1024	Information Bits per Symbol
0		Out of Range	
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	616	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

Table 9.34 Example Mapping of CQI Values to Number of Resource Blocks [158]

CQI Index	Modulation Scheme	Bits/ Symbol	REs/ PRB	N_{RB}	I_{MCS}	Transport Block Size	Code Rate
0			Out of Range				
1	QPSK	2	138	20	0	536	0.101449
2	QPSK	2	138	20	0	536	0.101449
3	QPSK	2	138	20	2	872	0.162319
4	QPSK	2	138	20	5	1736	0.318841
5	QPSK	2	138	20	7	2417	0.442210
6	QPSK	2	138	20	9	3112	0.568116
7	16QAM	4	138	20	12	4008	0.365217
8	16QAM	4	138	20	14	5160	0.469565
9	16QAM	4	138	20	16	6200	0.563768
10	64QAM	6	138	20	20	7992	0.484058
11	64QAM	6	138	20	23	9912	0.600000
12	64QAM	6	138	20	25	11448	0.692754
13	64QAM	6	138	20	27	12576	0.760870
14	64QAM	6	138	20	28	14688	0.888406
15	64QAM	6	138	20	28	14688	0.888406

control region, i.e., 138 resource elements per resource block pair in each subframe, and further it is assumed that there are 20 resource blocks assigned to the UE. The transport block size (excluding the 24-bit CRC) is then determined for various values of the CQI and modulation schemes.

As we mentioned earlier, power control dynamically adjusts the transmit power to compensate for variations in the instantaneous channel condition. The objective of these adjustments is to maintain an approximately constant E_b/N_0 at the receiver in order to successfully transmit data with an acceptable block error rate. In principle, TPC increases the power at the transmitter when the radio link experiences poor radio conditions, and vice versa. Therefore, the transmit power is in essence inversely proportional to the channel quality. This results in an approximately constant data rate, regardless of the channel variations. For services such as circuit-switched voice, this is a desirable property. Transmit-power control can be seen as one type of link adaptation, i.e., the adjustment of transmission parameters, in this case the transmit power, is adapted to differences and variations in the instantaneous channel conditions to maintain the received E_b/N_0 at a desired level. However, in mobile communication scenarios and in particular packet-data transmission, there is no desire to provide a constant data rate over a radio link at all times. In such cases, an alternative to transmit-power control is link adaptation by means of dynamic rate control. Rate control does not aim at keeping the instantaneous radio-link data rate constant, regardless of the instantaneous channel conditions. Instead, with rate control, the data rate is dynamically adjusted to compensate for the varying channel conditions, i.e., the data rate is increased when good channel conditions prevail, and vice versa. Thus, rate control maintains $E_b/N_0 \sim (P_{TX}, R)$ at the desired level, not by adjusting the transmission power P_{TX}, but rather by adjusting the data rate R. It can be shown that rate control is more efficient than power control. Rate control in principle implies that the power amplifier is always transmitting at full power and therefore is efficiently utilized. On the other hand, power control results in inefficient utilization of the power amplifier, as the transmission power is typically less than its maximum value. In practice, the link-level data rate is controlled by adjusting the modulation scheme and/or the channel coding rate. In the case of good channel conditions, the energy per bit E_b at the receiver is high and the main limitation of the data rate is going to be the bandwidth of the radio link [15].

The list of modulation schemes and code rates which can be signaled by means of a CQI value is given in Table 9.33. The AMC can exploit the UE feedback by assuming that the channel fading is sufficiently slow. This requires the channel coherence time to be greater than the CQI measurement delay (i.e., the time interval between the UE's measurement of the downlink reference signals and the subframe containing the correspondingly adapted downlink transmission on PDSCH), which is typically 7–8 ms (depending on the UE speed and the operating frequency). However, a trade-off needs to be made between the amount of CQI information reported by the UEs and the accuracy with which the link adaptation can track the prevailing channel conditions. More frequent reporting of the

CQI in the time-domain allows better tracking of the channel and interference variations, while finer resolution in the frequency-domain enables better use of frequency selective scheduling gain. However, both lead to increased feedback overhead in the uplink. Therefore, the eNB can configure both the time-domain update rate and the frequency-domain resolution of the CQI to optimize the overall performance of the system.

References

3GPP Specifications[35]

[1] 3GPP TS 36.101, Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception (Release 11), December 2012.

[2] 3GPP TS 36.104, Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception (Release 11), December 2012.

[3] 3GPP TS 36.211, Evolved Universal Terrestrial Radio Access (E-UTRA), Physical Channels and Modulation (Release 11), December 2012.

[4] 3GPP TS 36.212, Evolved Universal Terrestrial Radio Access (E-UTRA) Multiplexing and Channel Coding (Release 11), December 2012.

[5] 3GPP TS 36.213, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 11), December 2012.

[6] 3GPP TS 36.214, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer – Measurements (Release 11), December 2012.

[7] 3GPP TS 36.300, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN) Overall description, Stage 2 (Release 11), December 2012.

[8] 3GPP TS 36.321, Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol specification (Release 11), December 2012.

[9] 3GPP TS 36.331 Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 11), December 2012.

[10] 3GPP TR 36.912 Feasibility Study for Further Advancements for E-UTRA (LTE-Advanced) (Release 11), September 2012.

[11] 3GPP TR 36.814 Feasibility Study for Further Advancements for E-UTRA (LTE-Advanced) (Release 9), March 2010.

[12] 3GPP TR 36.819 Coordinated Multi-Point Operation for LTE Physical Layer Aspects (Release 11), December 2011.

Books

[13] C. Johnson, Long Term Evolution in Bullets, second ed., Create Space Independent Publishing Platform, 2012.

[14] S. Sesia, I. Toufik, M. Baker, LTE, the UMTS Long Term Evolution: From Theory to Practice, second ed., John Wiley & Sons, 2011.

[35]3GPP Specifications can be found at the following URL: http://www.3gpp.org/ftp/Specs/archive/36_series/

[15] E. Dahlman, S. Parkvall, J. Skold, 4G: LTE/LTE-Advanced for Mobile Broadband, Academic Press, 2011.
[16] H. Holma, A. Toskala, LTE Advanced: 3GPP Solution for IMT-Advanced, John Wiley & Sons, 2012.
[17] H. Holma, A. Toskala, LTE for UMTS: Evolution to LTE-Advanced, John Wiley & Sons, 2011.
[18] S. Ahmadi, Mobile WiMAX: A Systems Approach to Understanding IEEE 802.16m Radio Access Technology, Academic Press, 2010.
[19] B. Sklar, Digital Communications: Fundamentals and Applications, second ed., Prentice Hall, 2001.
[20] A. Paulraj, et al., Introduction to Space−Time Wireless Communications, Cambridge University Press, 2008.
[21] W.C. Jakes, Microwave Mobile Communications, second ed., Wiley-IEEE Press, 1994.
[22] G. Savo, Glisic, Advanced Wireless Communications: 4G Cognitive and Cooperative Broadband Technology, second ed., Wiley Inter-science, 2007.
[23] D.E. Dudgeon, R.M. Mersereau, Multidimensional Digital Signal Processing, Prentice Hall, 1984.
[24] D. Tse, P. Viswanath, Fundamentals of Wireless Communication, Cambridge University Press, 2005.
[25] E. Biglieri, et al., MIMO Wireless Communications, Cambridge University Press, 2010.
[26] A. Goldsmith, Wireless Communications, Cambridge University Press, 2005.
[27] S. Lin, D.J. Costello, Error Control Coding, second ed., Prentice Hall, 2004.
[28] S.L. Marple, Digital Spectral Analysis with Applications, Prentice Hall, 1987.
[29] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., Johns Hopkins University Press, 1996.
[30] A. Papoulis, Probability, Random Variables and Stochastic Processes, fourth ed., McGraw Hill Higher Education, 2002.
[31] L. Rade, Mathematics Handbook for Science and Engineering, Springer, 2010.
[32] G.A. Korn, T.M. Korn, Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review, Revised Edition, Dover Publications, 2000.
[33] C.E. Shannon, W. Weaver, The Mathematical Theory of Communication, University of Illinois Press, 1998.
[34] F. Dowla, Handbook of RF and Wireless Technologies, Newnes, 2003.
[35] G. Arfken, *Bessel functions*, Mathematical Methods for Physicists, third ed., Academic Press, 1985.
[36] W.G. Bickley, Bessel Functions and Formulae, Cambridge University Press, Cambridge, England, 1957.
[37] F. Bowman, Introduction to Bessel Functions, Dover, 2010.
[38] H.G. Myung, D. Goodman, Single Carrier FDMA: A New Air Interface for Long Term Evolution, Wiley & Sons, 2008.
[39] T.M. Cover, J.A. Thomas, Elements of Information Theory, John Wiley & Sons, 1991.
[40] J.G. Proakis, Digital Communications, fifth ed., McGraw-Hill, November 2007.
[41] N. Costa, S. Haykin, Multiple-Input Multiple-Output Channel Models: Theory and Practice, John Wiley & Sons, 2010.
[42] C. Oestges, B. Clerckx, MIMO Wireless Communications: From Real-World Propagation to Space−Time Code Design, Academic Press, 2007.

[43] A. Sibille, C. Oestges, A. Zanella, MIMO: From Theory to Implementation, Academic Press, 2010.
[44] I. Stakgold, M.J. Holst, Green's Functions and Boundary Value Problems, third ed., John Wiley & Sons, 2011.
[45] R.G. Gallager, Principles of Digital Communication, Cambridge University Press, 2008.

Articles

[46] G. Bauch, G. Dietl, MIMO for 3GPP LTE-Advanced and Beyond, Global Telecommunications Conference (GLOBECOM), 2010.
[47] C. Mehlfuhrer, et al., The vienna LTE simulators—enabling reproducibility in wireless communication research, EURASIP J. Adv. Signal Process. 2011 (1) (2011).
[48] B. Sklar, Rayleigh fading channels in mobile digital communication systems. I. Characterization, IEEE Commun. Mag. 35 (7) (1997).
[49] B. Sklar, Rayleigh fading channels in mobile digital communication systems. II. Mitigation, IEEE Commun. Mag. 35 (7) (July 1997).
[50] Y. Okumura, E. Ohmori, K. Fukuda, Field strength and its variability in VHF and UHF land mobile radio service, Rev. Electr. Commun. Lab. 16 (1968).
[51] M. Hata, Empirical formulae for propagation loss in land mobile radio services, IEEE Trans. Veh. Technol. VT-29 (3) (1980).
[52] Y. Wu, W.Y. Zou, Orthogonal frequency division multiplexing: a multi-carrier modulation scheme, IEEE Trans. Consum. Electron. 41 (3) (1995).
[53] H. Su, et. al., Analysis of tone reservation method for WiMAX system, International Symposium on Communications and Information Technologies, ISCIT'06, 2006.
[54] A. Osseiran, J.C. Guey, Hopping pilot pattern for interference mitigation in OFDM, IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, 2008.
[55] R. Nilsson, O. Edforst, M. Sandellt, P. Ola Borjesson, An analysis of two-dimensional pilot-symbol assisted modulation for OFDM, 1997 IEEE International Conference on Personal Wireless Communications Publication, 1997.
[56] H. Sampath, P. Stoica, A. Paulraj, Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion, IEEE Trans. Commun. 49 (12) (2001).
[57] Y.S. Choi, P.J. Voltz, F.A. Cassara, On channel estimation and detection for multicarrier signals in fast and selective rayleigh fading channels, IEEE Trans. Commun. 49 (8) (2001).
[58] S.M. Alamouti, A simple transmit diversity technique for wireless communications, IEEE J. Sel. Areas Commun. 16 (8) (1998).
[59] A. Nosratinia, T.E. Hunter, A. Hedayat, Cooperative communication in wireless networks, IEEE Commun. Mag. (2004).
[60] J.G. Andrews, Interference cancellation for cellular systems: a contemporary overview, IEEE Wirel. Commun. (2005).
[61] D. Gesbert, et al., From theory to practice: an overview of MIMO space−time coded wireless systems, IEEE J. Sel. Areas Commun. 21 (3) (2003).
[62] R.D. Murch, K. Ben Letaief, Antenna systems for broadband wireless access, IEEE Commun. Mag. (2002).
[63] S. Catreux, et al., Adaptive modulation and MIMO coding for broadband wireless data networks, IEEE Commun. Mag. (2002).

[64] A.J. Paulraj, D.A. Gore, R.U. Nabar, H. Bölcskei, An *overview of MIMO communications—a key to gigabit wireless*, Proc. IEEE 92 (2) (2004).
[65] M.G. Parker, K.G. Patersonyand, C. Tellambura, Golay complementary sequences, in: J.G. Proakis (Ed.), Wiley Encyclopedia of Telecommunications, John Wiley & Sons, 2003.
[66] D.J. Love, R.W. Heath Jr., Grassmannian beamforming for multiple-input multiple-output wireless systems, IEEE Trans. Inf. Theory 49 (10) (2003).
[67] V. Tarokh, H. Jafarkhani, On the computation and reduction of the peak-to-average power ratio in multicarrier communications, IEEE Trans. Commun. 48 (1) (2000).
[68] T.M. Schmid, D.C. Cox, Robust frequency and timing synchronization for OFDM, IEEE Trans. Commun. 45 (12) (1997).
[69] M. Pukkila, Channel Estimation Modeling, Postgraduate Course in Radio-communications, Fall 2000.
[70] M.S. Akram, Pilot-based channel estimation in OFDM systems, Master Thesis, 2007.
[71] Freescale Semiconductor Application Note, Channel Estimation in OFDM Systems, AN3059 Rev. 0, 2006.
[72] V. Srivastava, et al., Robust MMSE channel estimation in OFDM systems with practical timing synchronization, IEEE Wireless Communications and Networking Conference 2004.
[73] L. Zheng, D.N.C. Tse, Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels, IEEE Trans. Inf. Theory 49 (5) (2003).
[74] H. Weingarten, Y. Steinberg, S. Shamai, The capacity region of the gaussian multiple-input multiple-output broadcast channel, IEEE Trans. Inf. Theory 52 (9) (2006).
[75] F. Kaltenberger, et al., Capacity of linear multi-user MIMO precoding schemes with measured channel data, EURECOM, Sophia-Antipolis, France.
[76] V. Stankovic, M. Haardt, Multi-user MIMO downlink precoding for users with multiple antennas, in: Proceedings of the 12th Meeting of the Wireless World Research Forum (WWRF), Toronto, Canada, 2004.
[77] Q.H. Spencer, A.L. Swindlehurst, M. Haardt, Zero-Forcing methods for downlink spatial multiplexing in multiuser MIMO channels, IEEE Trans. Signal Process. 52 (2) (2004).
[78] N. Jindal, MIMO broadcast channels with finite-rate feedback, IEEE Trans. Inf. Theory 52 (11) (2006).
[79] Q.H. Spencer, C.B. Peel, A. Lee Swindlehurst, M. Haardt, An introduction to the multi-user MIMO downlink, IEEE Commun. Mag. (2004).
[80] D. Gesbert, et al., Shifting the MIMO paradigm, IEEE Signal Process. Mag. (2007).
[81] L. Wei, Capacity of Hybrid Open-Loop and Closed-Loop MIMO with Channel Uncertainty at Transmitter, Helsinki University of Technology, Espoo, 2008.
[82] D.J. Love, R.W. Heath Jr., Grassmannian precoding for spatial multiplexing systems, in: Proceedings of the Allerton Conference on Communication Control and Computing, Monticello, 2003.
[83] T. Abe, G. Bauch, Differential codebook MIMO precoding technique, 2007 IEEE Global Telecommunications Conference, 2007.
[84] B. Bandemer, M. Haardt, S. Visuri, Linear MMSE multi-user MIMO downlink precoding for users with multiple antennas, 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 2006.

[85] G. Mitchell, F.R. Kschischang, An augmented orthogonal code design for the non-coherent MIMO channel, 2008 24th Biennial Symposium on Communications, 2008.
[86] Y. Fan, P. Lundén, M. Kuusela, M. Valkama, Efficient semi-persistent scheduling for VoIP on EUTRA downlink, IEEE 68th Vehicular Technology Conference (VTC), Fall 2008.
[87] Goldsmith, et al., Capacity limits of MIMO channels, IEEE J. Sel. Areas Commun. 21 (2003).
[88] H. Weingarten, Y. Steinberg, S. Shamai, The capacity region of the gaussian MIMO broadcast channel, in: Proceedings of the Conference Information Sciences and Systems (CISS), Princeton, NJ, 2004.
[89] M. Costa, Writing on dirty paper, IEEE Trans. Inf. Theory 29 (1983).
[90] T. Svantesson, A.L. Swindlehurst, A performance bound for prediction of a multi-path MIMO channel, in: Proceedings of the 37th Asilomar Conference on Signals, Systems, and Computers, Session: Array Processing for Wireless Communications, Pacific Grove, CA, 2003.
[91] C.B. Peel, B.M. Hochwald, A.L. Swindlehurst, A vector-perturbation technique for near-capacity multi-antenna multi-user communication, IEEE Trans. Commun. (2003).
[92] M. Bengtsson, B. Ottersten, Optimal and suboptimal beamforming, in: L.C. Godara (Ed.), Handbook of Antennas in Wireless Communications, CRC Press, 2001.
[93] M. Schubert, H. Boche, Solution of the multiuser downlink beamforming problem with individual SINR constraints, IEEE Trans. Veh. Technol. 53 (2004).
[94] Q.H. Spencer, A.L. Swindlehurst, M. Haardt, Zero-forcing methods for downlink spatial multiplexing in multi-user MIMO channels, IEEE Trans. Signal Process. 52 (2004).
[95] G. Caire, S. Shamai, On the achievable throughput of a multi-antenna gaussian broadcast channel, IEEE Trans. Inf. Theory 49 (2003).
[96] U. Erez, S. Shamai, R. Zamir, Capacity and lattice strategies for cancelling known interference, in: Proceedings International Symposium Information Theory and its Applications, 2000.
[97] C. Windpassinger, R.F.H. Fischer, J.B. Huber, Lattice-reduction-aided broadcast precoding, in: Proceedings of Fifth ITG Conference Source and Channel Coding, 2004.
[98] Berrou, Glavieux, Thitimajshima, Near shannon limit error-correcting coding and decoding: turbo-codes, in: Proceedings of 1993 International Communication Conference, 1993.
[99] Q.H. Spencer, A.L. Swindlehurst, Channel allocation in multi-user MIMO wireless communications systems, in: Proceedings of 2004 International Communication Conference, 2004.
[100] Perez, Seghers, Costello, A distance spectrum interpretation of turbo codes, IEEE Trans. Inf. Theory 42 (6) (1996).
[101] J. Rothweller, Turbo codes, IEEE Potentials 18 (I) (1999).
[102] C. Berrou, A. Glavieux, P. Thitimajshima, Near Shannon limit error-correcting coding and decoding: turbo codes, in: Proceedings of 1993 International Communication Conference, 1993.
[103] C. Berrou, A. Glavieux, Near optimum error correcting coding and decoding: turbo-codes, IEEE Trans. Commun. 44 (10) (1996).

[104] C. Heegard, Stephen B. Wicker, Turbo Coding, Springer, 2010.
[105] C. Berrou, R. Pyndiah, P. Adde, C. Douillard, R. Le Bidan, An overview of turbo codes and their applications, The European Conference on Digital Wireless Technology, 2005.
[106] H. Ma, J. Wolf, On tail biting convolutional codes, IEEE Trans. Commun. COM-34 (2) (1986).
[107] C. Weiss, C. Bettstetter, S. Riedel, Code construction and decoding of parallel concatenated tail-biting codes, IEEE Trans. Inf. Theory 47 (1) (2001).
[108] Y.E. Wang, R. Ramesh, To bite or not to bite—a study of tail bits versus tail-biting, Seventh IEEE International Symposium Personal, Indoor and Mobile Radio Communications, PIMRC'96, vol. 2, 1996.
[109] B.E. Priyanto, H. Codina, Initial performance evaluation of DFT-Spread OFDM Based SC-FDMA for UTRA LTE Uplink, IEEE 65th Vehicular Technology Conference (VTC), 2007.
[110] J. Lee, J.-K. Han, J. Zhang, MIMO Technologies in 3GPP LTE and LTE-Advanced, EURASIP J. Wireless Commun. Networking (2009).
[111] S. Catreux, P.F. Driessen, L.J Greenstein, Attainable throughput of an interference-limited multiple-input multiple-output (MIMO) cellular system, IEEE Trans. Commun. 49 (8) (2001).
[112] I. Telatar, Capacity of multi-antenna Gaussian channels, AT&T Technical Memorandum, 1995.
[113] G.J. Foschini, M.J. Gans, On limits of wireless communications in a fading environment when using multiple antennas, Wireless Pers. Commun 1998.
[114] M.-S. Alouini, A. Goldsmith, Capacity of Nakagami multipath fading channels, Proceedings Vehicular Technology Conference, 1997.
[115] D. Chizhik, G.J. Foschini, R.A. Valenzuela, Capacities of multi-element transmit and receive antennas: correlations and keyholes, Electron. Lett. 36 (2000).
[116] D. Gesbert, H.B.̈olcskei, D. Gore, A. Paulraj, MIMO wireless channels: capacity and performance prediction, Global Telecommunications Conference (GLOBECOM), 2000.
[117] D.A. Gore, R.U. Nabar, A. Paulraj, Selecting an optimal set of transmit antennas for a low rank matrix channel, Proc. ICASSP 2000 49 (8) (2000).
[118] S. Sandhu, R.U. Nabar, D.A. Gore, A. Paulraj, Near-optimal selection of transmit antennas for a MIMO channel based on Shannon capacity, in: Proceedings of the 34th Asilomar Conference on Signals, Systems and Computers, 2000.
[119] A.F. Molisch, M.Z. Win, J.H. Winters, A. Paulraj, Capacity of MIMO systems with antenna selection, in: Proceedings of the IEEE International Conference on Communications (ICC), 2001.
[120] Z. Shen, et al., Overview of 3GPP LTE-advanced carrier aggregation for 4G wireless communications, IEEE Commun. Mag. 50 (2) (2012).
[121] Y.-H. Nam, et al., Evolution of reference signals for LTE-advanced systems, IEEE Commun. Mag. 50 (2) (2012).
[122] L. Liu, et al., Downlink MIMO in LTE-advanced: SU-MIMO vs. MU-MIMO, IEEE Commun. Mag. 50 (2) (2012).
[123] D. Lee, et al., Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges, IEEE Commun. Mag. 50 (2) (2012).
[124] D. Bai, et al., LTE-Advanced modem design: challenges and perspectives, IEEE Commun. Mag. 50 (2) (2012).

[125] J. Lee, Y. Kim, et al., Coordinated multipoint transmission and reception in LTE-Advanced systems, IEEE Commun. Mag. 50 (2) (2012).
[126] Z. Shen, et al., Dynamic uplink-downlink configuration and interference management in TD-LTE, IEEE Commun. Mag. 50 (2) (2012).
[127] C.S. Park, et al., Evolution of uplink MIMO for LTE-advanced, IEEE Commun. Mag. 49 (2) (2011).
[128] R. Irmer, et al., Coordinated multipoint: concepts, performance, and field trial results, IEEE Commun. Mag. 49 (2) (2011).
[129] M. Vu, A. Paulraj, MIMO Wireless Linear Precoding, IEEE Signal Process. Mag. 24 (5) (2007).
[130] R. Gold, Optimal binary sequences for spread spectrum multiplexing, IEEE Trans. Inf. Theory 13 (4) (1967).
[131] X. Cai, G.B. Giannakis, Bounding performance and suppressing inter-carrier interference in wireless mobile OFDM, IEEE Trans. Commun. 51 (12) (2003).
[132] J. Lee, H.-L. Lou, D. Toumpakaris, J.M. Cioffi, Effect of carrier frequency offset on OFDM systems for multipath fading channels, Global Telecommunications Conference (GLOBECOM), 2004.

Miscellaneous

[133] 3G Americas, MIMO and smart antennas for 3G and 4G wireless systems. <http://www.3gamericas.org>, 2010.
[134] Rohde, Schwarz Application Notes, LTE-Advanced technology introduction, 2010.
[135] 3G Americas, 3GPP mobile broadband innovation path to 4G: Release 9, Release 10 and beyond: HSPA + , SAE/LTE and LTE-Advanced, 2010.
[136] 3G Americas, HSPA to LTE-Advanced: 3GPP broadband evolution to IMT-Advanced (4G), 2009.
[137] 3G Americas, The mobile broadband evolution: 3GPP release 8 and beyond HSPA + , SAE/LTE and LTE-Advanced, 2009.
[138] IEEE 802.16m-07/244r1, Y. Song, L. Cai, K. Wu, H. Yang, Collaborative MIMO, 2007.
[139] Agilent Technologies, 3GPP long term evolution: system overview, product development, and test challenges. <http://www.agilent.com>, 2009.
[140] IEEE C802.16m-08/153, B.-C. Ihm, J. Choi, W. Lee, Pilot related to DL-MIMO, 2008.
[141] MATLAB CENTRAL, An open exchange for the MATLAB and SIMULINK user community. <http://www.mathworks.com/matlabcentral>.
[142] Vienna University of Technology, Institute of Telecommunications, LTE link-level and system-level simulators. <http://www.nt.tuwien.ac.at/about-us/staff/josep-colom-ikuno/lte-simulators>, 2012.
[143] 3GPP R1-060385 Motorola, Cubic metric in 3GPP-LTE, 2006.
[144] The Coded Modulation Library (The Iterative Solutions Coded Modulation Library (ISCML) is an open source toolbox for simulating capacity approaching codes in Matlab). <http://www.iterativesolutions.com>.
[145] 4G Americas, 4G mobile broadband evolution: release 10, release 11 and beyond—HSPA, SAE/LTE and LTE-Advanced, 2012.
[146] 4G Americas, MIMO and smart antennas for mobile broadband networks, 2012.
[147] 4G Americas, Developing and integrating a high performance HET-NET, 2012.

[148] 4G Americas, 4G mobile broadband evolution: 3GPP release 10 and beyond—HSPA +, SAE/LTE and LTE-Advanced, 2011.
[149] 3GPP R1-113297, NTT DOCOMO, Enhanced PDCCH for DL MIMO in Rel-11, 2011.
[150] 3GPP R1-112928, Ericsson, ST-Ericsson, On enhanced PDCCH design, 2011.
[151] 3GPP R1-113175, Renesas Mobile, Link-level evaluation of E-PDCCH design aspects, 2011.
[152] 3GPP R1-113194, LG Electronics, Consideration on transmit diversity scheme for enhanced PDCCH transmission, 2011.
[153] 3GPP R1-113202, Intel Corporation, Performance analysis of enhanced downlink control signaling, 2011.
[154] 3GPP R1-113238, Research in Motion, E-PDCCH transmission with transmit diversity, 2011.
[155] 3GPP R1-114302, NTT DOCOMO, DM-RS design for E-PDCCH in Rel-11, 2011.
[156] Anritsu, Future technologies and testing for fixed mobile convergence, SAE and LTE in cellular mobile communications. <http://www.anritsu.com/search/en-US/downloadssearch.aspx>, 2008.
[157] Wireless World Initiative New Radio—WINNER +, Enabling Techniques for LTE-A and beyond, Deliverable D2.2. <http://projects.celtic-initiative.org/winner+/deliverables_winnerplus.html>, 2010.
[158] LTE System Description. <http://www.sharetechnote.com>.
[159] DSPLOG Signal Processing for Communication <http://www.dsplog.com>.
[160] 4G Evolution Lab—LTE and LTE-Advanced Toolbox and Block-set: a simulation library for the 3GPP LTE and LTE Advanced specification for the Mathworks MATLAB and Simulink. <http://www.steepestascent.com>.
[161] R. Duncan, Supporting Operators with LTE Roaming, CDG International Roaming Team, August 2012.
[162] Samsung Electronics, LTE PHY specification. <http://www.slideshare.net/allabout4g/lte-rel-8-physical-layer>, 2008.
[163] Motorola White Paper, TD-LTE, Exciting alternative global Momentum. <www.motorola.com>, 2010.
[164] B. Schulz, LTE Transmission modes and beamforming, Rohde & Schwarz White Paper, 2011.
[165] E. Seidel, LTE-A Carrier Aggregation Enhancements in Release 11, NOMOR Research GmbH, Munich, Germany, 2012.
[166] B. Holter, On the Capacity of the MIMO Channel- A Tutorial Introduction, Norwegian University of Science and Technology, 2001.
[167] S. Giraud, et al., Bulk acoustic wave filters synthesis and optimization for multi-standard communication terminals, IEEE Trans Ultrason Ferroelectr Freq Control 57 (I) (2010).
[168] L.A. Coldren, R. Rosenberg, Surface-acoustic-wave resonator filters, Proc. IEEE 67 (1) (1979).
[169] Return loss, from Wikipedia, < http://en.wikipedia.org/wiki/Return_loss >.
[170] Masaaki Koiwa, et al., Multiband Mobile Terminals, NTT DoCoMo Tech. J. 8 (2) (2006).
[171] Siemens White Paper, TD-SCDMA: the Solution for TDD bands, March 2004 < http://www.tdscdma-forum.org/ >.

[172] D.C. Chu, Polyphase Codes With Good Periodic Correlation Properties, IEEE Trans Inf Theory (1972).
[173] Maximum length sequence, from Wikipedia, <http://en.wikipedia.org/wiki/Maximum_length_sequence>.
[174] Jung-In Kim, et al., SSS Detection Method for Initial Cell Searchin 3GPP LTE FDD/TDD Dual Mode Receiver, Ninenth International Symposium on Communication and Information Technologies, 2009.
[175] J. van de Beek, et al., Low-complex frame synchronization in OFDM systems, Fourth IEEE International Conference on Universal Personal Communications. 1995.
[176] ARTIST 4G Project, D2.2 Advanced receiver signal processing techniques – evaluation and characterization, 2011.
[177] Hidekazu Taoka, et al., MIMO and CoMP in LTE-Advanced, DoCoMo Techn. J. 12 (2) (2010).
[178] Sigen Ye, et al., Enhanced Physical Downlink Control Channel in LTE Advanced Release 11, IEEE Communications Magazine, 2013.
[180] Yusuke Ohwatari, et al., Performance of Advanced Receiver Employing Interference Rejection Combining to Suppress Inter-cell Interference in LTE-Advanced Downlink, IEEE Vehicular Technology Conference (VTC), 2011.
[181] Ernesto Zimmermann, et al., On the Complexity of Sphere Decoding, in Proceedings International Symposium on Wireless Personal Multimedia Communication, 2004.
[182] Federico Boccardi, et al., Precoding Schemes for the MIMO-GBC, Int. Zurich Seminar on Communications (IZS), 2006.

CHAPTER

Uplink Physical Layer Functions

10

CHAPTER CONTENTS

10.1 Overview of uplink physical layer processing 724
10.2 Principles of SC-FDMA 724
10.3 SC-FDMA and clustered SC-FDMA 734
10.4 Uplink frame structure 736
10.5 Physical resource structure and channel mapping 738
10.5.1 Physical resource blocks 742
10.5.2 Uplink resource allocation 743
10.5.2.1 Uplink resource allocation type 0 747
10.5.2.2 Uplink resource allocation type 1 748
10.6 Uplink reference signals and channel estimation 753
10.6.1 Uplink demodulation reference signals 754
10.6.2 Sounding reference signals 762
10.7 Uplink physical channel processing 769
10.7.1 Physical uplink control channel 774
10.7.2 PUCCH formats 778
10.7.2.1 PUCCH formats 1, 1a, 1b 780
10.7.2.2 PUCCH formats 2, 2a, 2b 785
10.7.2.3 PUCCH format 3 797
10.7.3 Physical uplink shared channel 802
10.7.4 Physical random-access channel 821
10.8 SC-FDMA signal generation 837
10.9 LTE/LTE-Advanced uplink multi-antenna transmission schemes 838
10.9.1 Transmit antenna selection 839
10.9.2 Virtual MU-MIMO (cooperative MIMO) 840
10.9.3 Closed-loop spatial multiplexing 841
10.9.4 Transmit diversity 844
10.9.5 Uplink MIMO receivers 846
10.9.5.1 LMMSE receiver 849
10.9.5.2 MRC receiver 849
10.9.5.3 MMSE-IRC receiver 849
10.9.5.4 SIC receiver 850
10.10 Uplink transmission modes 851

© 2014 Elsevier Inc. All rights reserved.

Ex.1005.00752

10.11 Uplink feedback in LTE/LTE-Advanced 852
10.12 Uplink transmitter and receiver chain 861
References 866

This chapter describes the physical layer protocols and functional processing in the LTE/LTE-Advanced uplink direction. As shown in Figure 10.1, the physical layer is the lowest protocol layer in baseband signal processing that interfaces with the physical media (in this case air-interface) through which the signal is transmitted and received. The physical layer receives MAC PDUs (transport blocks) and processes them through channel coding, interleaving, baseband modulation, layer mapping for multi-antenna transmission, eigen-precoding, resource element, and antenna mapping. The choice of the appropriate modulation and coding scheme and multi-antenna transmission mode are critical to achieve the

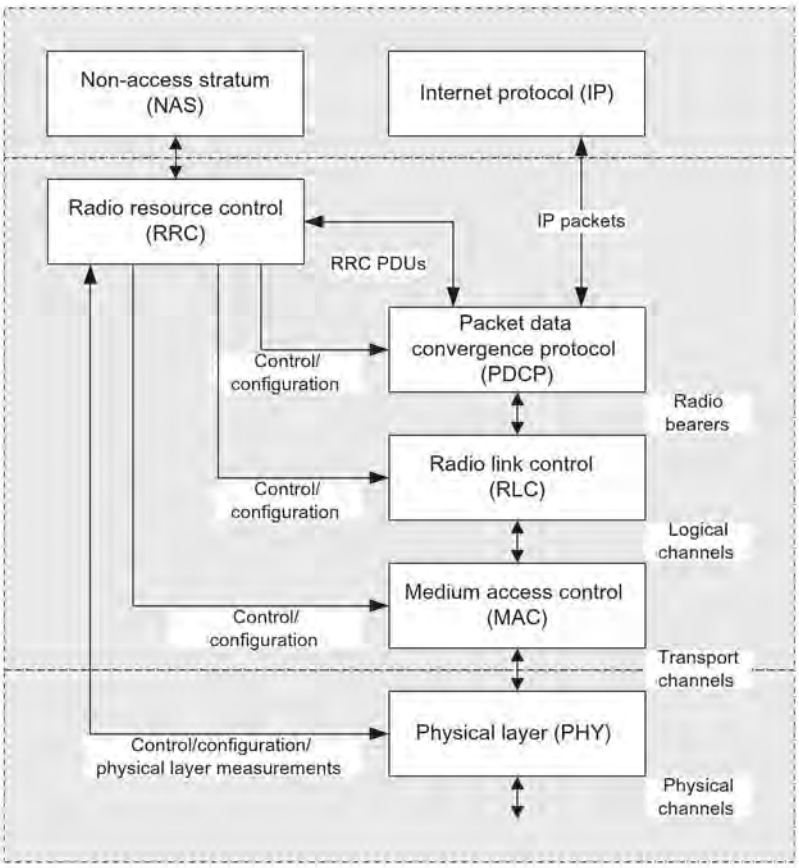


FIGURE 10.1

Overall view of LTE/LTE-Advanced physical layer.

desired reliability/robustness and system/user throughput in mobile wireless data communications. Typical mobile radio channels tend to be dispersive and time variant and exhibit severe Doppler effects, multipath delay variation, intra-cell and inter-cell interference, and fading.

A good and robust design of the physical layer ensures that the system can normally operate and overcome the above deleterious effects, and can provide the maximum throughput and lowest latency under various operating conditions. The chapters on physical layer in this book are dedicated to the systematic design of physical layer protocols and functional blocks of Rel-10/11 of LTE systems, the theoretical background on physical layer procedures, and performance evaluation of physical layer components. The theoretical background is provided to ensure that the reader understands the underlying theory governing the operation of various functional blocks and procedures. While the focus is mainly on the techniques that were incorporated in the design of the LTE/LTE-Advanced physical layer, the author has attempted to take a more generic and systematic approach to the design of physical layer for the systems beyond IMT-Advanced so that the reader can understand and apply the learnings to the design and implementation of any MIMO-OFDM-based physical layer irrespective of the radio access technology. Physical layer processes both control-plane and user-plane signals; however, due to different design requirements and reliability and performance criteria, the procedures tend to be different.

While the requirements for the design of the LTE/LTE-Advanced uplink physical layer are mainly similar to those of the downlink, there are some design attributes that are specific to the uplink physical layer including the desire for orthogonal uplink transmission by different UEs to minimize intra-cell interference and maximize the capacity, flexible data rate adaptation as a function of SINR, sufficiently low peak-to-average power ratio (PAPR) of the transmitted signal to avoid excessive cost, size, and power consumption of the UE transmit circuitry, ability to exploit the frequency diversity attainable by the wider bandwidths even when transmitting at low data rates, support for frequency-selective scheduling, and support for advanced multiple-antenna techniques in order to exploit spatial diversity and to enhance uplink capacity. The multiple-access scheme selected for the LTE uplink to satisfy the above design requirements is single-carrier frequency division multiple access (SC-FDMA). A major advantage of SC-FDMA over the direct-sequence code-division multiple access (DS-CDMA) scheme used in UMTS is that it achieves intra-cell orthogonality even in frequency-selective channels. The SC-FDMA scheme avoids strong intra-cell interference linked to DS-CDMA, which significantly reduces system capacity and limits the use of adaptive modulation. A code-division multiplexed uplink further suffers from an increased PAPR, if multi-code transmission is used from a single UE.

The use of OFDM for the LTE uplink would have been attractive given the maximal uplink/downlink functional commonality and ease of scheduling/implementation. In principle, an OFDM scheme similar to the LTE downlink could

satisfy all the uplink design criteria listed above, except the low PAPR requirement. The SC-FDMA combines the desirable characteristics of OFDM with the low PAPR of single-carrier transmission schemes. Similar to OFDM, the SC-FDMA divides the transmission bandwidth into multiple parallel sub-carriers, with the orthogonality between the sub-carriers being maintained in frequency-selective channels by the use of a cyclic prefix or guard period. The use of a cyclic prefix prevents inter-symbol interference (ISI) between SC-FDMA symbols. It transforms the linear convolution of the multipath channel into a circular convolution, enabling the receiver to equalize the channel simply by scaling each sub-carrier by a complex gain factor. However, unlike OFDM, where the data symbols directly modulate each sub-carrier independently such that the amplitude of each sub-carrier at a given time instant is set by the constellation points of the baseband modulation scheme, in SC-FDMA the signal modulated on a given sub-carrier is a linear combination of the data symbols transmitted at the same time instant. Therefore, in each symbol period, the transmitted sub-carriers of an SC-FDMA signal carry a component of each modulated data symbol. This gives SC-FDMA its crucial single-carrier property, which results in the PAPR being significantly lower than pure multi-carrier transmission schemes such as OFDM [39].

10.1 Overview of uplink physical layer processing

This section provides an overview of physical layer processing in the LTE/LTE-Advanced uplink. The processing steps and the associated control signaling are depicted in Figure 10.2. The main criteria in the design of the LTE/LTE-Advanced physical layer were to increase the application throughput and capacity, reduce access latency, support higher user mobility, minimize intra-cell and inter-cell interference, improve reliability of control and data channel coverage especially at the cell-edge, and reduce the complexity and signaling overhead. The overall performance evaluations provided in Chapter 11 suggest that the above requirements have been successfully met. In the next sections, we describe in detail the functional blocks and protocols of the uplink physical layer and their interactions based on the order the information is processed, as shown in Figure 10.2.

10.2 Principles of SC-FDMA

The LTE uplink uses SC-FDMA with a cyclic prefix in order to achieve inter-user orthogonality and to enable efficient frequency-domain equalization at the receiver. The discrete Fourier transform (DFT)-spread OFDM (DFT-S OFDM) is a form of the single-carrier transmission technique, where the signal is generated in the frequency-domain similar to OFDMA, and is illustrated in Figure 10.3.

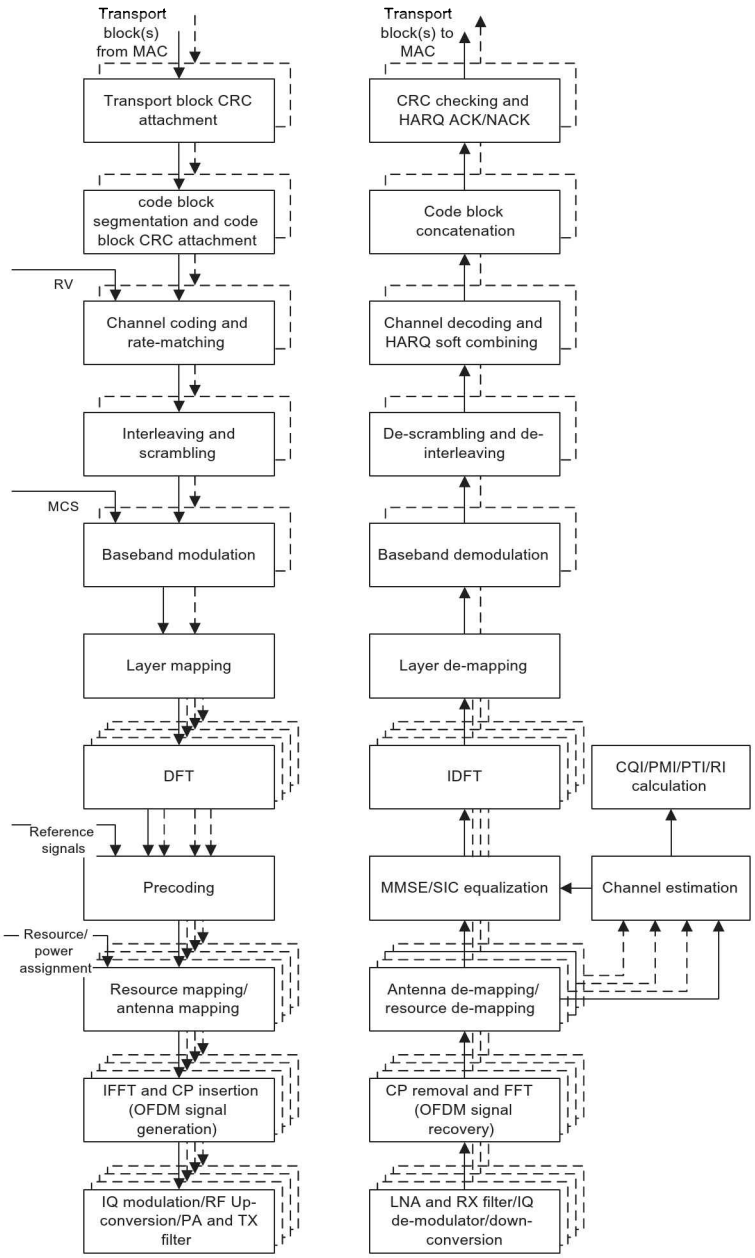


FIGURE 10.2

The LTE/LTE-Advanced uplink physical layer processing blocks [3,14,15,16].

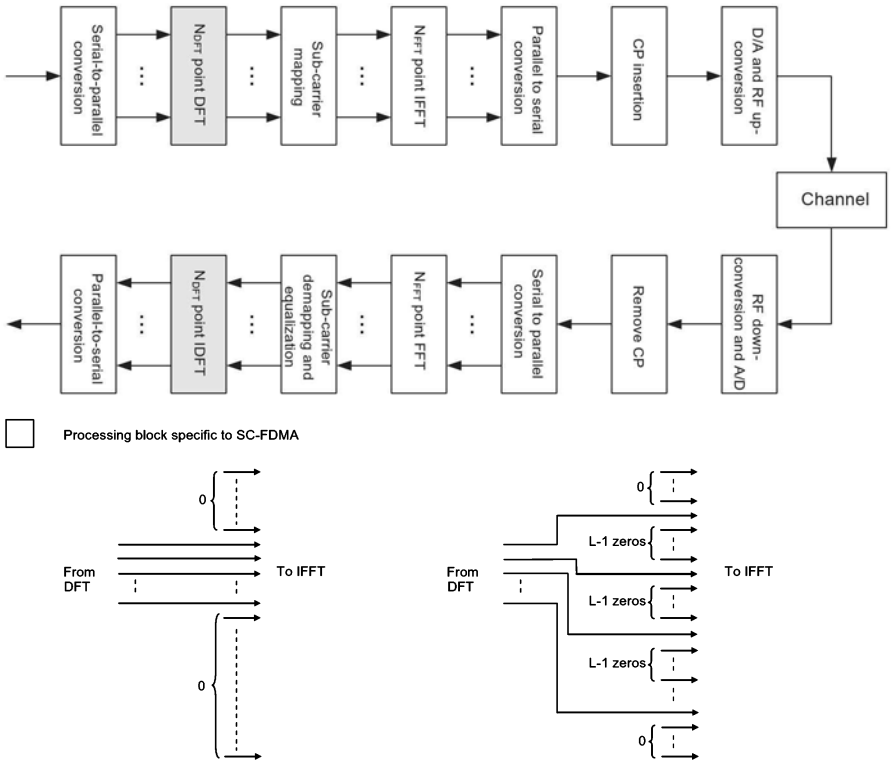


FIGURE 10.3

Transmitter/receiver structure for SC-FDMA with localized and distributed sub-carrier mapping schemes (note that $N_{DFT} < N_{FFT}$) [39].

In Figure 10.3, the common processing blocks in OFDMA and SC-FDMA are distinguished from those that are specific to SC-FDMA. This allows for a relatively high degree of commonality with the downlink OFDMA baseband processing using the same parameters, e.g., clock frequency, sub-carrier spacing, and FFT/IFFT size. The use of SC-FDMA in the uplink is mainly due to relatively inferior PAPR properties of OFDMA that result in worse uplink coverage compared to SC-FDMA. The PAPR characteristics are important for cost-effective design of a mobile station's power amplifiers (PAs).

The principles of SC-FDMA signal processing can be explained as follows. The *i*th transmitted symbol in an SC-FDMA system without cyclic prefix in single transmit/receive antenna case can be expressed as a vector of length N_{FFT} samples defined by $\mathbf{y} = \mathbf{F\Theta Dx}$, where $\mathbf{x} = (x_1, x_2, \ldots, x_{N_{DFT}})^T$ is an $N_{DFT} \times 1$ vector with N_{DFT} QAM-modulated symbols (the superscript "T" denotes matrix transpose operation), $\mathbf{D}$ is an $N_{DFT} \times N_{DFT}$ matrix which performs N_{DFT}-point DFT operation, $\mathbf{\Theta}$ is the $N_{FFT} \times N_{DFT}$ mapping matrix for sub-carrier assignment, and

F performs N_{FFT}-point IFFT operation. After the propagation through the multipath fading channel and addition of the additive white Gaussian noise and removing the cyclic prefix and going through the N_{FFT}-point FFT module, the received signal vector in the frequency-domain can be expressed as $\mathbf{z} = \mathbf{H\Theta Dx} + \mathbf{w}$, where **H** is the diagonal channel matrix and **w** is the receiver noise vector. Note that the maximum excess delay of the channel is assumed to be shorter than the cyclic prefix; therefore, the ISI can be mitigated by the cyclic prefix. The amplitude and phase distortion in the received signal due to the multipath channel is compensated by a frequency-domain equalizer (FDE) and the signal at the FDE output can be described as $\mathbf{v} = \mathbf{Cz}$, where $\mathbf{C} = \text{diag}(c_1, c_2, \ldots, c_{N_{\text{FFT}}})$ is the diagonal matrix of FDE coefficients. The FDE complex coefficients can be derived using minimum mean-square error (MMSE) criterion as follows:

$$c_k = \frac{H_k^*}{|H_k|^2 + \sigma_{\text{n}}^2/\sigma_{\text{s}}^2} \tag{10.1}$$

where k denotes the sub-carrier index, σ_{n}^2 denotes the variance of the additive noise, and σ_{s}^2 is the variance of the transmitted pilot symbol. Following the sub-carrier de-mapping function and IDFT de-spreading, an $N_{\text{DFT}} \times 1$ vector $\hat{\mathbf{x}}$ containing N_{DFT} QAM-modulated symbols as an estimate to the input vector **x** is obtained at the receiver. The IDFT de-spreading block in the receiver averages the noise over each sub-carrier. A particular sub-carrier may experience deep fading in a frequency-selective fading channel. The IDFT de-spreading averages and spreads the fading effect, which results in a noise enhancement to all the QAM symbols. Therefore, the IDFT de-spreading makes DFT-S OFDM more sensitive to the noise.

As shown in Figure 10.3, the modulation symbols in blocks of N_{DFT} symbols are processed through an N_{DFT}-point DFT processor, where N_{DFT} denotes the number of sub-carriers assigned to the transmission of the data/control block. The rationale for the use of DFT precoding is to reduce the cubic metric (CM) of the transmitted signal. From an implementation point of view, the DFT size should ideally be a power of 2. However, such a constraint would limit the scheduler flexibility in terms of the amount of resources that can be assigned for an uplink transmission. In LTE, the DFT size and the size of the resource allocation is limited to products of the integers 2, 3, and 5. For example, the DFT sizes of 60, 72, and 96 are allowed, but a DFT size of 42 is not allowed [16]. Therefore, the DFT can be implemented as a combination of relatively low-complex radix-2, radix-3, and radix-5 FFT processing blocks.

The sub-carrier mapping in SC-FDMA determines which part of the spectrum is used for transmission by inserting a number of zeros (i.e., null sub-carriers inserted between the data sub-carriers) in the upper and/or lower end of the frequency region, as shown in Figure 10.3. Distributed mapping can be obtained by inserting $L - 1$ zeros between each DFT output sample. A mapping with $L = 1$ (as shown in Figure 10.3) corresponds to localized transmissions, i.e., transmissions

where the DFT outputs are mapped to consecutive sub-carriers. In localized sub-carrier mapping, a group of N_{DFT} adjacent sub-carriers are allocated to a user, where $N_{DFT} < N_{FFT}$ results in zeros being appended to the output of the DFT spreading which results in an interpolated version of the original N_{DFT} QAM data symbols at the IFFT output of the OFDM modulator. The transmitted signal is similar to a narrowband single carrier with a cyclic prefix, which is equivalent to time-domain generation with a repetition factor of one ($L = 1$) and sinc($\cdot$) function pulse shaping. Distributed transmission, on the other hand, allocates N_{DFT} equally spaced sub-carriers (i.e., every Lth sub-carrier). Therefore, $L - 1$ zeros are inserted between the N_{DFT} DFT outputs and additional zeros are appended to either side of the DFT output prior to the IFFT operation. Similar to the localized case, the zeros appended on either side of the DFT output provide sinc($\cdot$) function interpolation effect, while the zeros inserted between the DFT outputs produce waveform repetition in the time-domain. This results in a transmitted signal similar to time-domain interleaved FDMA (IFDMA) with repetition factor L and pulse shaping. Unlike the standard OFDM where the each data symbol is carried by the individual sub-carriers, the SC-FDMA transmitter carries data symbols over a group of sub-carriers transmitted simultaneously. In other words, the group of sub-carriers that carry each data symbol can be viewed as one frequency band carrying data sequentially in a standard FDMA. It can be mathematically shown that the SC-FDMA baseband time-domain samples after IDFT are the original data symbol set repeated in time-domain over the symbol period (Figure 10.4).

The performance of SC-FDMA with localized and distributed sub-carrier mapping was simulated to demonstrate the potential gain that can be obtained via sub-carrier distribution across frequency, particularly in higher mobility conditions. Figure 10.5 shows the block error rate (BLER) as a function of signal-to-noise

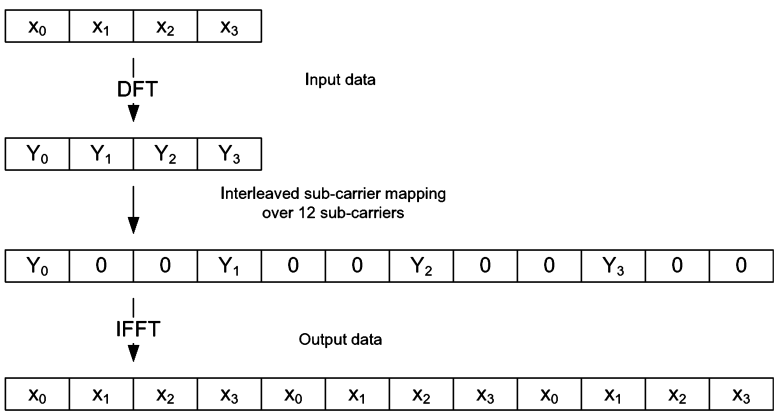


FIGURE 10.4

Time-domain representation of interleaved SC-FDMA.

ratio (SNR) for localized and distributed sub-carrier mapping. It is assumed that $N_{DFT} = 60, N_{FFT} = 512,$ and $N_{CP} = 20$ (number of samples constituting the cyclic prefix). Furthermore, the evaluation is conducted with ITU-PedA and ITU-VehA channel models, MMSE equalization at the receiver, QPSK modulation, and raised-cosine pulse shaping at the transmitter. It can be observed that the distributed sub-carrier mapping is more suited for high-mobility conditions and the gain is increased with increasing mobility.

The goal of equalization is to compensate the effects of channel distortion due to frequency-selectivity and to restore the original signal. One approach to signal equalization is in the time-domain using a linear equalizer, which consists of a linear filter with an impulse response $w(t)$ operating on the received signal. By selecting different filter impulse responses, different receiver/equalizer schemes can be implemented. For example, in DS-CDMA systems, a RAKE receiver is typically used. The RAKE receiver is an equalization method in which the filter impulse response is selected to match the channel impulse response $w(t) = h^*(-t)$. Therefore, in a RAKE receiver, the filter impulse response is selected as the complex conjugate of the time-reversed channel impulse response. Selecting the receiver filter according to the maximal ratio combining (MRC) criterion as above, i.e., a channel-matched filter, the filter maximizes the post-processing SNR. However, MRC-based filtering does not provide any compensation for any radio channel frequency selectivity and no equalization. Thus, MRC-based receiver filtering is appropriate when the received signal is mainly impaired by noise or interference from other transmissions. Another alternative is to select the receiver filter to compensate the radio channel frequency selectivity. This can be achieved by selecting the receiver filter impulse response to satisfy $w(t) * h(t) = 1$, where the

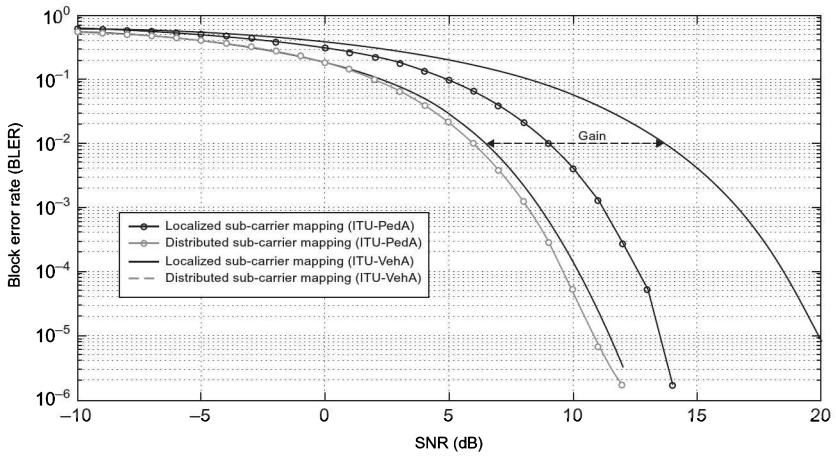


FIGURE 10.5

Comparison of localized and distributed sub-carrier mapping in SC-FDMA [136].

operator "*" denotes linear convolution. This method of filtering is known as zero-forcing (ZF) equalization, which compensates any channel frequency selectivity. However, the ZF equalization may lead to a significant increase in the noise level after equalization, degrading the overall link performance. This will be the case especially when the channel has large variations in its frequency response. Another alternative is to select a filter which provides a trade-off between signal distortion due to channel frequency selectivity and the corruption due to noise/interference, resulting in a filter impulse response that minimizes the mean-square error between the equalizer output and the transmitted signal. The linear equalizers are typically implemented as a time-discrete finite impulse response (FIR) filter with a certain number of taps. In general, the complexity of such a discrete-time equalizer increases with increasing bandwidth of the signal [16].

An alternative to time-domain equalization is frequency-domain equalization which can significantly reduce the complexity of linear equalization. In this method, the equalization is performed on a block of data. The received signal is transformed to the frequency-domain using a DFT operation. The equalization is done as a frequency-domain filtering operation, where the frequency-domain filter $W(k)$ is the DFT of the corresponding time-domain impulse response $w(n)$. The equalized frequency-domain signal is then transformed to the time-domain using an IDFT operator. For processing of each signal block of size $N = 2^m$, the frequency-domain equalization would include two N-point DFT/IDFT operations and N complex multiplications.

With the introduction of a cyclic prefix, the channel would appear as a circular convolution over a receiver processing block of size N. Therefore, there would be no need for overlap-and-discard in the receiver processing. Furthermore, the frequency-domain filter taps can now be calculated directly from an estimate of the sampled channel frequency response. Similar to the OFDM case, the drawback of using a cyclic prefix in conjunction with single-carrier transmission is the overhead in terms of extra power and bandwidth consumption. One method to reduce the relative cyclic-prefix overhead is to increase the block size N of the FDE. However, the accuracy of block equalization requires the channel to be approximately constant over a time period corresponding to the size of the processing block.

The detection procedure for an SC-FDMA signal is illustrated in Figure 10.6, and is compared with that of an OFDM waveform. The transmission through a time-dispersive or equivalently a frequency-selective channel will distort the SC-FDMA signal and an equalizer is needed to compensate for the effects of channel frequency-selectivity. However, as shown in Figure 10.6, a simple one-tap equalizer can be applied to each sub-carrier in OFDM, whereas in the case of SC-FDMA, the frequency-domain equalization function is applied to the complex-valued symbols at the output of the sub-carrier de-mapping and prior to the IDFT operation (Figure 10.7).

It must be noted that in OFDM downlink parameterization, the DC sub-carrier is unused in order to support direct conversion receiver architectures. In contrast, nulling DC sub-carrier is not possible in SC-FDMA since it affects the low

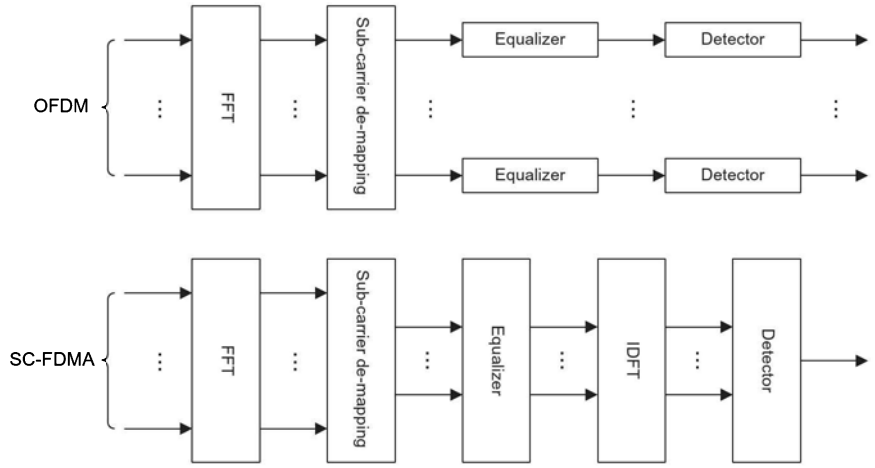


FIGURE 10.6

Illustration of different equalization/detection aspects of SC-FDMA and OFDM [39].

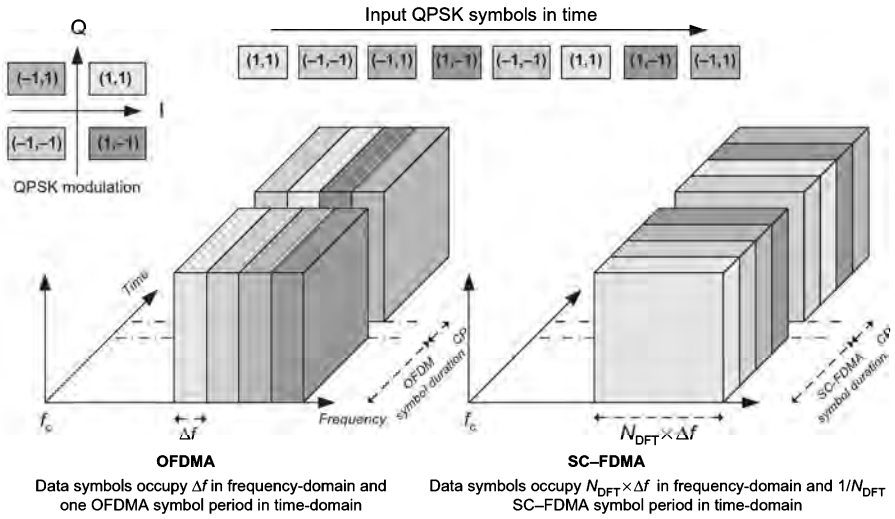


FIGURE 10.7

Comparison of OFDMA and SC-FDMA operation using QPSK modulation with $N_{DFT} = 4$ and $L = 1$. The input QPSK symbols, shown on top, are fed into the modulator from left to right [134].

CM/PAPR property of the transmit signal. Direct conversion transmitters and receivers can introduce distortion at the carrier frequency (zero frequency or DC sub-carrier in baseband) due to local oscillator leakage. In the LTE/LTE-Advanced downlink, this issue is avoided by inclusion of an unused DC sub-carrier. However, for the uplink, the same solution may adversely impact the low CM property of the transmitted signal. In order to minimize the impact of such distortion on the packet error rate and the CM/PAPR, the DC sub-carrier of the SC-FDMA signal is modulated in the same way as all the other sub-carriers but the sub-carriers are all frequency-shifted by half a sub-carrier spacing ($\Delta f/2$), resulting in an offset of 7.5 kHz relative to the DC sub-carrier. Therefore, two sub-carriers straddle the DC location; hence, the amount of distortion affecting any individual resource block is reduced by half.

In order to demonstrate the similarities and differences between OFDMA and SC-FDMA operation, let us assume that one wishes to transmit a sequence of eight QPSK symbols as shown in Figure 10.7 [134]. In the OFDMA case, assuming $N_{\text{DFT}} = 4$, four QPSK symbols would be processed in parallel, each of them modulating its own sub-carrier at the appropriate QPSK phase. After one OFDM symbol period, a guard period or cyclic prefix is inserted to mitigate the multipath effects. For SC-FDMA, each symbol is transmitted sequentially. With $N_{\text{DFT}} = 4$, there are four data symbols transmitted in one SC-FDMA symbol period. The higher-rate data symbols require four times the bandwidth, and so each data symbol occupies $N_{\text{DFT}} \times \Delta f$ Hz of the spectrum assuming a sub-carrier spacing of Δf Hz. After four data symbols, the cyclic prefix is inserted. Note the OFDMA and SC-FDMA symbol periods are the same [134].

As mentioned earlier, the PAPR of OFDM is intrinsically inferior to SC-FDMA. Figure 10.8 shows the comparison of complementary cumulative distribution functions (CCDF)[1] of OFDM and SC-FDMA PAPRs. It can be seen that the PAPR of SC-FDMA is approximately 3 dB better than that of OFDM with a probability of 0.99. In the case of 16QAM modulation, the PAPR of SC-FDMA increases relative to that of SC-FDMA with QPSK modulation, whereas in the case of OFDM, the PAPR distribution is more or less independent of the modulation scheme because the OFDM signal is the sum of a large number of independently modulated sub-carriers, thus the instantaneous power has an approximately exponential distribution, regardless of the modulation scheme applied to the different sub-carriers [16].

A new study suggests that the PAPR of an OFDM signal does not precisely predict the PA power de-rating or power capability as accurately as the CM criterion. The CM criterion has been adopted by 3GPP as a method to determine PA power de-rating because of its accuracy over a wide range of devices and signals.

[1]The CDF of the real-valued random variable X is defined as $x \rightarrow F_X(x) = P(X \leq x), \forall x \in \mathbb{R}$, where the right-hand side represents the probability that the random variable X takes on a real value less than or equal to x. The CDF of X can be defined in terms of the probability density function $f(x)$ as $F(x) = \int_{-\infty}^{x} f(x)\mathrm{d}x$. The CCDF, on the other hand, is defined as $P(X > x) = 1 - F_X(x)$ [31–33].

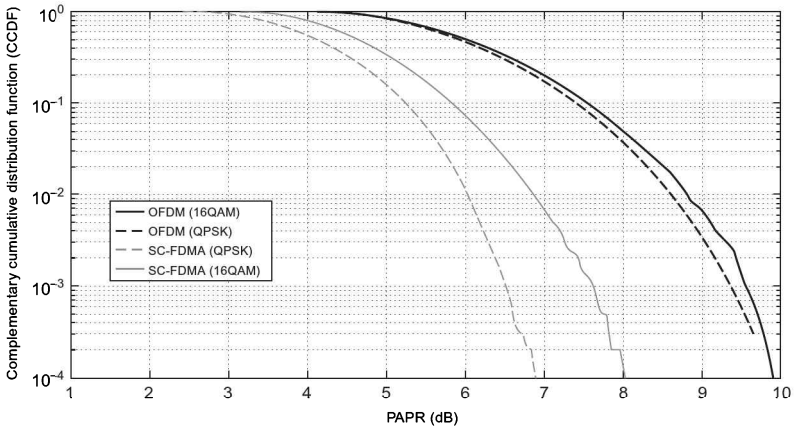


FIGURE 10.8

Comparison of OFDM and SC-FDMA PAPRs (5 MHz bandwidth) [136].

This method has proven to be more accurate for WCDMA signals compared to methods that use the statistical PAPR to predict power de-rating based on the information that has been collected from several devices for a variety of signals. The CM can be calculated using the following equation [138]:

$$\mathrm{CM} = \frac{20\log(s^3_{\mathrm{norm_{RMS}}}) - 20\log(s^3_{\mathrm{norm\text{-}reference_{RMS}}})}{K} \tag{10.2}$$

where $s_{\mathrm{norm_{RMS}}}$ denotes the root mean-square (RMS) value of the normalized voltage waveform of the input signal, and $s_{\mathrm{norm\text{-}reference_{RMS}}}$ is the RMS value of the normalized voltage waveform of the reference signal (i.e., signal generated using 12.2kbps adaptive multi-rate (AMR) speech codec [ref]). The CM is used to model the impact of nonlinearity in the PA on the adjacent-channel leakage ratio. A small value of the CM represents a small PA backoff. The CM is insensitive to the highest power values of the distribution that only have very low probabilities. Thus, methods to reduce the PAPR may not have any major impact on the CM. In practice, the value of parameter K was empirically determined to be 1.85 for a set of WCDMA signals. In other words, this equation computes the cubic power of a signal $s(t)$ and, compares it to a reference signal $s_{\mathrm{reference}}(t)$, and uses the empirical slope factor K to estimate the CM value. Although the signal vectors are shown as a function of time, the timescale is effectively removed in the RMS operation and is not a part of the computation. Thus, the raw CM result is not a function of symbol rate, or alternatively not a function of signal bandwidth. Similarly, a change in chip rate has no effect on PAPR. The bandwidth effects will be significant in this study since it includes signals with 3 dB bandwidths larger than the 3.84 MHz used for WCDMA. Table 10.1 gives the CM values corresponding to some example input signals such as OFDM, DFT-S OFDM, and

Table 10.1 Example CM Values Corresponding to Slope Factor *K* [138]

Signal					
Type	Modulation	RAW CM (dB)	*K*	Bandwidth Offset (dB)	CM (dB)
WCDMA	Voice	1.52	–	0.00	0
OFDM	16QAM	7.75	1.56	0.77	4.76
DFT-S OFDM	QPSK	3.44	1.56	0.77	2.00
DFT-S OFDM	16QAM	4.85	1.56	0.77	2.90
DFT-S OFDM	64QAM	5.18	1.56	0.77	3.11
IFDMA	QPSK	2.40	1.56	0.00	0.56
IFDMA	16QAM	4.36	1.56	0.00	1.82
IFDMA	64QAM	4.64	1.56	0.00	2.00

IFDMA which is a variant of SC-FDMA. It can be seen that DFT-S OFDM has an advantage relative to other signals shown in the table. In Table 10.1, the bandwidth offset in decibels denotes a de-rating increase of 0.77 dB from the CM regression line that can be used to estimate the actual de-rating of higher bandwidth signals. The *K* factor is empirically derived based on the measurements on various PAs.

10.3 SC-FDMA and clustered SC-FDMA

The LTE Rel-8 uplink is based on single-carrier frequency division multiple access scheme, which allocates carriers across a contiguous block of spectrum, thus limiting scheduling flexibility. LTE-Advanced introduced clustered SC-FDMA in the uplink, allowing frequency-selective scheduling of component carriers for better link-level performance. The physical uplink shared channel (PUSCH) and the physical uplink control channel (PUCCH) can be scheduled together to reduce latency. However, clustered SC-FDMA increases PAPR, leading to transmitter non-linearity issues. Furthermore, the presence of multi-carrier signals increases opportunities for co-channel and adjacent-channel interference.

LTE-Advanced enhances the uplink multiple access scheme by adopting clustered SC-FDMA. This scheme is similar to SC-FDMA but has the advantage that it allows non-contiguous (clustered) groups of sub-carriers to be allocated for transmission by a single UE, thus enabling uplink frequency-selective scheduling and improved link performance. Clustered SC-FDMA was chosen with preference to pure OFDM to avoid a significant increase in PAPR. It helps satisfy the requirement for increased uplink spectral efficiency while maintaining backward

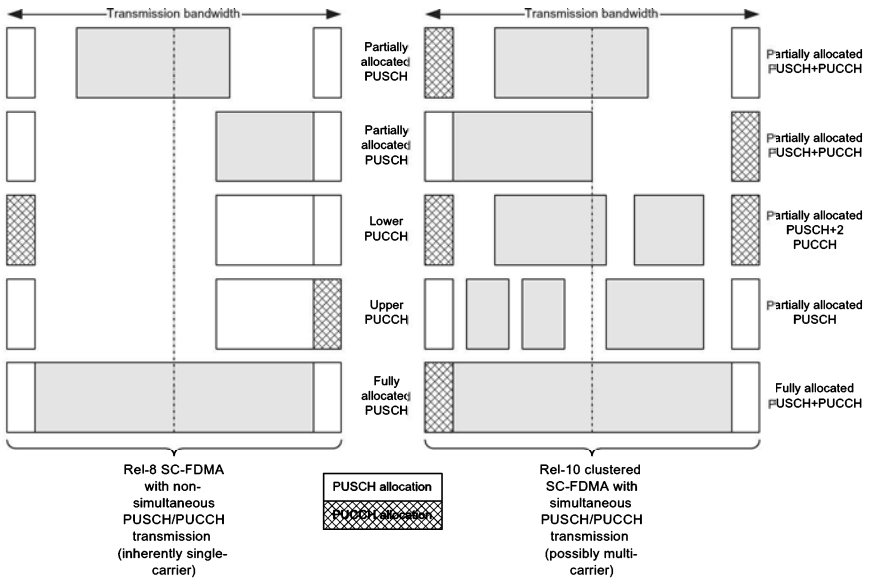


FIGURE 10.9

Comparison of LTE and LTE-Advanced uplink transmission [161].

compatibility with LTE Rel-8. There is only one transport block and one HARQ entity per scheduled component carrier. Each transport block is mapped to a single component carrier and a UE may be scheduled over multiple component carriers simultaneously using carrier aggregation.

Examples of different LTE Rel-8 and Rel-10 uplink configurations are shown in Figure 10.9. The main observation is that all LTE Rel-8 configurations are single carrier, which means that the PAPR is no greater than the underlying QPSK or 16QAM modulation format, whereas in LTE Rel-10, it is possible to transmit more than one carrier, which makes the PAPR higher than the Rel-8 cases. Note that the carriers mentioned above in the context of clustered SC-FDMA and simultaneous PUCCH/PUSCH transmission are contained within one component carrier and should not be confused with multiple component carriers and carrier aggregation. The initial specifications limit the number of SC-FDMA clusters to two, which will provide some improved spectral efficiency over a single cluster when transmitting through a frequency-selective channel with more than one distinct peak.

The clustered SC-FDMA must be distinguished from multiple SC-FDMA in the sense that clustered SC-FDMA retains a single DFT operation, but modifies the resource element mapping at the output of the DFT operation from a single cluster as used for SC-FDMA to multiple clusters which are multiplexed with $N_{\text{FFT}} - N_{\text{DFT}}$ zeros to form the input of the IFFT operation over N_{FFT} virtual subcarriers. The resulting waveform is no longer single-carrier but still has a low CM

value. On the other hand, multiple SC-FDMA consists of a number of DFT operations, where the kth DFT block is of size N_{DFT}^{k}; the output symbols are then multiplexed with $N_{\text{FFT}} - \sum_{k=1}^{N_{\text{cluster}}} N_{\text{DFT}}^{k}$ zeros to fit the N_{FFT}-point IFFT operation. This waveform is no longer single carrier and experiences worse CM compared to that of clustered SC-FDMA for the same number of clusters [16].

10.4 Uplink frame structure

The LTE frame structure for the FDD and TDD schemes was discussed in Chapter 9. The downlink and uplink frame structure parameters are the same, except that the time-domain underlying symbols in the downlink are OFDM symbols, whereas in the uplink the underlying symbols are SC-FDMA symbols due to the use of different multiple access methods in the uplink and downlink. In this chapter, some of the uplink-specific aspects of frame structure are repeated to set the ground for further uplink physical layer discussions as well as physical channel mappings. As mentioned in Chapter 9, LTE/LTE-Advanced downlink and uplink transmissions are organized into radio frames of duration 10 ms consisting of 10 equally-sized consecutive subframes of 1 ms each, consisting of a number of consecutive OFDM/SC-FDMA symbols. Each subframe is further divided into two equally sized slots of 0.5 ms. The number of OFDM/SC-FDMA symbols within one subframe is either 14 for a normal cyclic prefix or 12 for an extended cyclic prefix (Figure 10.10).

In an FDD scheme, all ten subframes within a radio frame contain either downlink or uplink subframes depending on the link direction. The uplink and downlink transmitters use separate frequency bands for transmissions; therefore, both downlink and uplink transmitters can transmit at all times and at the same time [3]. In a TDD scheme, the bandwidth is shared between uplink and downlink, with the sharing being performed by allotting different periods of time to uplink and downlink (i.e., time multiplexing). In LTE/LTE-Advanced, there are seven different patterns of uplink/downlink switching known as uplink–downlink configurations 0 through 6 that are shown in Table 10.2 [3].

As shown in Figure 10.10 and Table 10.2, the special subframe (subframe 1 in every uplink–downlink configuration and subframe 6 in uplink–downlink configurations 0, 1, 2, and 6) contains a portion of downlink transmission at the start of the subframe (downlink pilot time slot or DwPTS), a portion of unused symbols in the middle of the subframe (guard period), and a portion of uplink transmission at the end of the subframe (uplink pilot time slot or UpPTS). The lengths of DwPTS, GP, and UpPTS can take one of the ten combinations of values given in Table 10.3. The LTE standard specifies the lengths in terms of the sampling time T_s, but the lengths can be interpreted in terms of the number of OFDM/SC-FDMA symbols, as shown in Table 10.3.

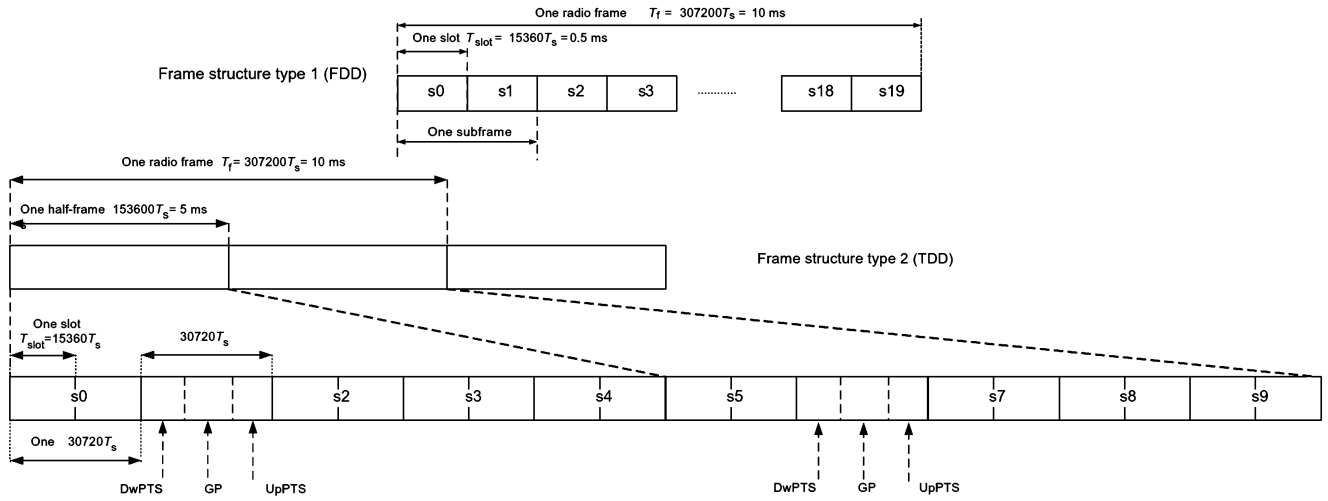


FIGURE 10.10

LTE/LTE-Advanced FDD/TDD frame structures [3].

Table 10.2 LTE/LTE-Advanced TDD Frame Configurations [3]

Uplink–Downlink Configuration	Multiplexing of Downlink, Uplink, and Special Subframes in One Radio Frame
0	DL SP UL UL UL DL SP UL UL UL
1	DL SP UL UL DL DL SP UL UL DL
2	DL SP UL DL DL DL SP UL DL DL
3	DL SP UL UL UL DL DL DL DL DL
4	DL SP UL UL DL DL DL DL DL DL
5	DL SP UL DL DL DL DL DL DL DL
6	DL SP UL UL UL DL SP UL UL DL

10.5 Physical resource structure and channel mapping

The smallest resource unit for uplink transmissions is a resource element (one sub-carrier). An uplink physical channel corresponds to a set of resource elements carrying information originating from the higher layers. The following uplink physical channels are defined:

- The physical uplink shared channel (PUSCH) carries user data and upper layer signaling as well as control information such as HARQ ACK/NACK, channel

Table 10.3 Special Subframe Configurations in TDD [3]

Special Subframe Configuration	Normal Cyclic Prefix in Downlink			Extended Cyclic Prefix in Downlink		
	DwPTS	UpPTS		DwPTS	UpPTS	
		Normal Cyclic Prefix in Uplink	Extended Cyclic Prefix in Uplink		Normal Cyclic Prefix in Uplink	Extended Cyclic Prefix in Uplink
0	3			3		
1	9			8		
2	10	1	1	9	1	1
3	11			10		
4	12			3		
5	3			8	2	2
6	9			9		
7	10	2	2	–	–	–
8	11			–	–	–
9	6			–	–	–

quality indicator (CQI), precoding matrix index (PMI), and rank indication (RI).

- The physical uplink control channel (PUCCH) carries control information such as HARQ ACK/NACK, CQI, and RI. In LTE Rel-8, the UE transmits control information on the PUCCH, when no PUSCH allocations are available to the UE. Starting from LTE Rel-10, PUCCH and PUSCH can be simultaneously transmitted by the UE. The PUCCH carries uplink control information (UCI). Each PUCCH uses one resource block in each of the two slots of the subframe. As shown in Figure 10.11, the PUCCH resource blocks are always allocated at the edge of the channel. This helps reduce out-of-band (OOB) emissions from data transmissions on inner resource blocks. Allocating the PUCCH in the center of the band would further create non-contiguous PUSCH resources, which limits contiguous resources that can be allocated to a single UE, and thus would limit the achievable peak throughput. The PUCCH resource blocks are transmitted with a hopping pattern from one edge to the other edge in each slot of the subframe. Frequency hopping provides frequency diversity for PUCCH. A maximum of four resource blocks are reserved for PUCCH in this example. The physical resources used for PUCCH depend on parameters that are provided by the higher layers. The PUCCH allocations also contain demodulation reference signals (DM-RSs) for coherent demodulation.

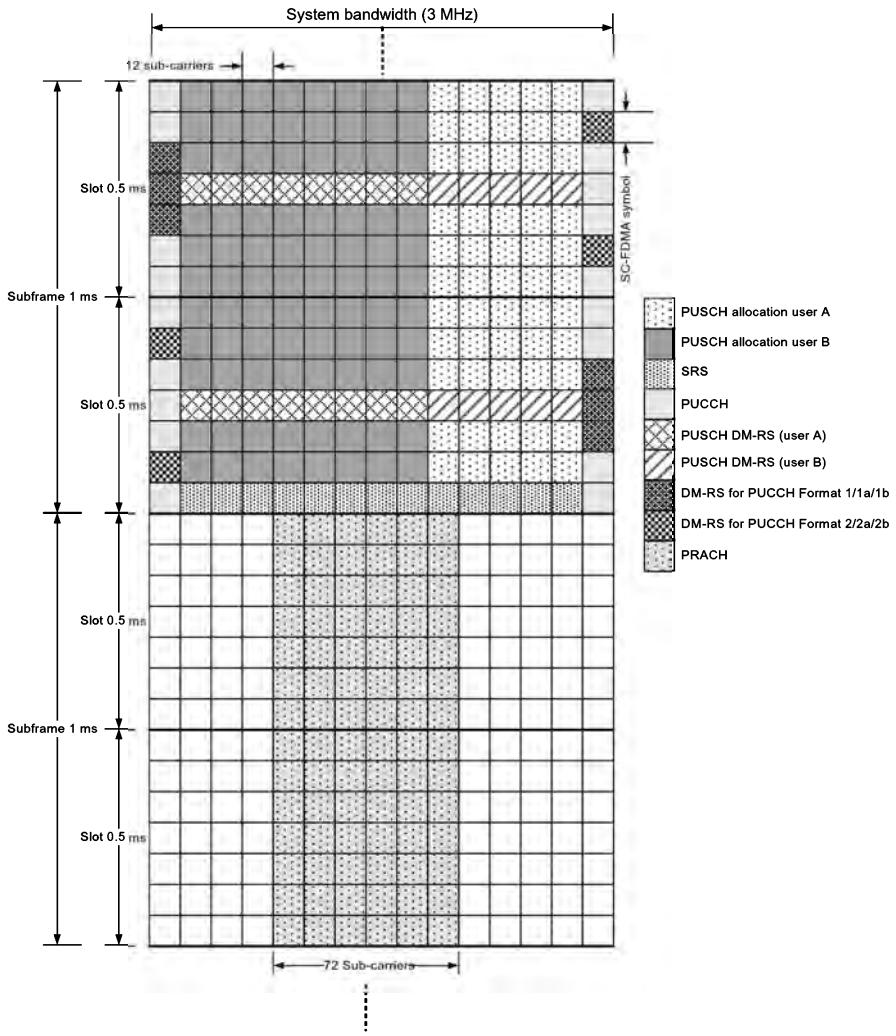


FIGURE 10.11

Example of uplink frame structure and location of the physical signals/channels.

The PUCCH allocation may contain one, two, or three DM-RS symbols. The DM-RS for each UE is generated using a cyclic shift of base Zadoff–Chu sequences. The RRC sub-layer assigns cyclic shift parameters (i.e., cyclic shift and base sequence group) to the UE during the RRC connection setup. The location of DM-RS differs based on formats and the length of the cyclic prefix.

- The physical random-access channel (PRACH) is used by non-synchronized UEs to access the network. It is further used during initial system access,

handovers, and radio link failure recovery scenarios. A preamble sequence, which may be randomly selected from a group of sequences or assigned (dedicated) by the eNB, is transmitted over PRACH. The sequences are Zadoff–Chu sequences of length 839. There are 64 different cyclic shifts generated from each root Zadoff–Chu sequence with zero correlation between them. Therefore, different cyclic shifts are essentially orthogonal to each other. The PRACH resource consists of six consecutive resource blocks. The location of PRACH is indicated by an offset parameter signaled by higher-layer signaling. There is at most one PRACH resource per subframe in FDD mode. Multiple UEs can access the PRACH resource by using different preambles. The UEs may select preambles randomly, which may result in collision. Therefore, contention resolution mechanisms are necessary. The PRACH may not be present in every subframe and every radio frame. The *PRACH configuration index* parameter indicates frame number and subframe numbers in which the PRACH resource is available. The sub-carrier spacing for PRACH transmission is different from that of other channels in the uplink. The number of sub-carriers is equal to the length of preamble, which is 839 (due to longer OFDM symbol length for PRACH). The sub-carrier spacing is 1.25 kHz, resulting in a longer symbol duration of 800 μs. The choice of sub-carrier spacing also results in additional guard bands on both sides because the entire 1.08 MHz is not used for the PRACH transmission.

In addition, there are two types of reference signals in the uplink as follows:

- The demodulation reference signal is a known signal transmitted in PUSCH and/or PUCCH whenever they are transmitted and provides a coherent demodulation reference for uplink transmissions. The DM-RS provides the reference for channel estimation and coherent demodulation of uplink transmissions from the UE on PUSCH or PUCCH. The DM-RS is transmitted only in the resource blocks transmitted from the UE for PUSCH or PUCCH. It is transmitted during a specific symbol and time-division multiplexed with the resource blocks allocated to the UE for data or control information transmission. The DM-RS signal is based on Zadoff–Chu sequences with a length of up to the maximum number of resource blocks. Thirty groups of base sequences are defined where each group contains three base sequences. Group hopping patterns and sequence hopping patterns are defined, resulting in different base sequences in different slots. The base sequences are generated differently based on whether they are more or less than three resource blocks. From these base sequences, different cyclic shifts are defined. The cyclic shifts are derived from UE-specific parameters assigned in DCI format 0 and cell identity. This creates the UE-specific cyclic shifts necessary to achieve orthogonality for MU-MIMO, where the same resource block is allocated to multiple users. The cyclic shift is also derived from the slot number, resulting in different cyclic shifts in different slots for the same UE.

- The sounding reference signal (SRS) is a wideband signal transmitted by the UE for several purposes. The eNB can use SRS transmission to select the best resource blocks to allocate to the UE (i.e., frequency-selective scheduling). It may also be used for uplink power control. If uplink MIMO is supported, it may be further used for adaptive antenna switching for uplink transmit diversity. The SRS sequences are selected from the same Zadoff–Chu root sequences used for DM-RS. The cyclic shifts are configured for each UE when SRS transmission is configured. If configured, the SRS transmission always occurs in the last SC-FDMA symbol of a subframe. The SRS transmission may not be allowed in every subframe. The periodicity of subframes containing SRS may be limited to 2, 5, 10, 20, 40, 80, 160, or 320 ms [5]. The bandwidth for SRS transmission is specified in terms of number of resource blocks and may extend from 4 to 36 resource blocks, if the maximum number of resource blocks in the uplink is in the range of 6-40. The SRS bandwidth may extend from 4 to 96 resource blocks, if the maximum number of resource blocks in the uplink is in the range of 80–110 [3]. The SRS parameters (periodicity, bandwidth, etc.) can be cell-specific and broadcast from each cell. Additionally, UE-specific parameters that override cell-specific parameters may be configured through higher-layer signaling. Frequency hopping may be defined for SRS transmission, if the SRS bandwidth is smaller than the system bandwidth.

The uplink unit of resource allocation is a resource block. A resource block is a two-dimensional array of all SC-FDMA symbols in a slot and 12 sub-carriers in the frequency-domain. There are 72 (6 symbols by 12 sub-carriers) or 84 (7 symbols by 12 sub-carriers) resource elements in a resource block corresponding to extended cyclic prefix or normal cyclic prefix, respectively. The number of resource blocks depends on the channel bandwidth. There is a minimum of six resource blocks for the 1.4 MHz channel and 100 resource blocks for the 20 MHz channel. A UE is allocated a number of contiguous or non-contiguous resource blocks. Figure 10.11 illustrates an example of mapping of uplink physical channels and signals to the physical resources within a subframe for an uplink bandwidth of 3 MHz (15 resource blocks).

10.5.1 Physical resource blocks

The signal transmitted in each slot is described by one or several resource grids of $12 \times N_{RB}^{UL}$ sub-carriers and N_s^{UL} SC-FDMA symbols. The uplink resource grid is illustrated in Figure 10.12. The parameter N_{RB}^{UL} denotes the number of uplink resource blocks, which depends on the uplink transmission bandwidth, configured in the cell whose value falls in the range of $\min(N_{RB}^{UL}) \leq N_{RB}^{UL} \leq \max(N_{RB}^{UL})$, where $\min(N_{RB}^{UL}) = 6$ and $\max(N_{RB}^{UL}) = 110$ represent the minimum and maximum uplink bandwidths, respectively. The number of SC-FDMA symbols in a slot N_s^{UL} depends on the cyclic prefix length which is configured by the higher layer signaling. Similar to the downlink, there are seven SC-FDMA symbols for the normal cyclic prefix and six SC-FDMA symbols for extended cyclic prefix in each slot.

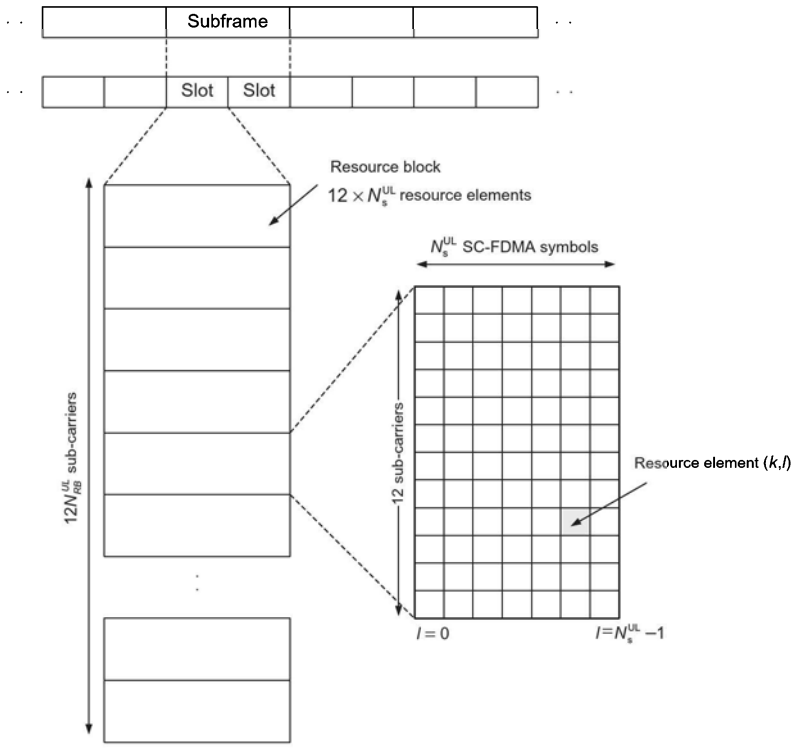


FIGURE 10.12

Uplink physical resource structure [3].

Each resource element in the resource grid within a slot is uniquely defined by the index pair (k, l), where $k = 0, 1, \ldots, 12 \times N_{\mathrm{RB}}^{\mathrm{UL}} - 1$ and $l = 0, 1, \ldots, N_{\mathrm{s}}^{\mathrm{UL}} - 1$ are the frequency and time-domain indices, respectively. The resource element (k, l) on antenna port p corresponds to the complex-valued symbol a_{kl}^{p}. A physical resource block (PRB) is defined as $N_{\mathrm{s}}^{\mathrm{UL}}$ consecutive SC-FDMA symbols in the time-domain and 12 consecutive sub-carriers in the frequency-domain. Note that the number of sub-carriers in a resource block corresponding to sub-carrier spacing 7.5 kHz is 24; however, throughout this book, we assume a 15 kHz sub-carrier spacing. A PRB in the uplink thus consists of $12 \times N_{\mathrm{s}}^{\mathrm{UL}}$ resource elements, corresponding to one slot in the time-domain and 180 kHz in the frequency-domain. The PRB number in the frequency-domain n_{PRB} and the resource element (k, l) in a slot are related as $n_{\mathrm{PRB}} = \lfloor k/12 \rfloor$.

10.5.2 Uplink resource allocation

The concept of virtual resource blocks (VRBs) or distributed resource allocation and the mapping of VRBs to PRB were discussed in Chapter 9 for downlink resource allocations. The distributed resource allocation in the downlink consists

of two separate steps that include mapping from VRB pairs to PRB pairs such that frequency-consecutive VRB pairs are not mapped to frequency-consecutive PRB pairs, as well as splitting of each resource-block pair such that the two resource blocks of the resource-block pair are transmitted with a certain frequency gap over the slots of each subframe. The latter step can be seen as frequency hopping on a slot basis. In the absence of the frequency hopping, each VRB is directly mapped to a PRB. Note that localized transmissions without use of frequency hopping can achieve some degree of frequency-selective-scheduling gain, whereas localized transmission using frequency hopping yields frequency-diversity gain as well as inter-cell interference randomization [16].

The VRB concept can also be used in the uplink, allowing use of frequency-diversity transmission in the uplink. However, in the uplink, where transmission from a terminal should always be over a set of consecutive sub-carriers in the absence of clustered SC-FDMA, distribution of resource-block pairs in the frequency-domain is not possible. Therefore, uplink distributed resource allocation is only similar to the second step of downlink distributed transmission, i.e., inserting a frequency gap between the PRBs in the first and second slots of a subframe. This is referred to as PUSCH frequency hopping in the literature. There are two types of uplink frequency hopping defined for PUSCH as follows: (1) sub-band hopping according to cell-specific hopping/mirroring patterns (as shown in Figure 10.13); and (2) hopping based on explicit hopping information contained in the scheduling grant. The uplink frequency hopping is not supported for clustered SC-FDMA transmission, since sufficient frequency diversity gain can be achieved by properly assigning the location of the clusters.

In order to support sub-band frequency hopping according to cell-specific hopping/mirroring patterns, a set of consecutive sub-bands with predetermined size are defined as illustrated in Figure 10.13. It should be noted that the sub-bands do not cover the total uplink frequency band due to the fact that a number of resource blocks at the edges of the uplink frequency band are designated for transmission of PUCCH. As an example, if the overall uplink bandwidth contains 25 resource blocks and there are a total of four sub-bands, each consisting of five resource blocks, then five resource blocks at both edges of the band are not included in the hopping bandwidth and can be used for PUCCH transmission (see Figure 10.13). In the case of sub-band hopping, the set of VRBs provided in the scheduling grant are mapped to a corresponding set of PRBs according to a cell-specific hopping pattern. The physical resources to be used for transmission are obtained by shifting the VRBs provided in the scheduling grant by a number of sub-bands according to the hopping pattern, where the hopping pattern can provide different shifts for each slot. The hopping pattern is cell specific, which implies that it is the same for all terminals within a cell; however, different terminals will transmit on non a_{kl}^{p} overlapping physical resources as long as they are assigned non a_{kl}^{p} overlapping virtual resources [16].

In addition to the hopping pattern, there is also a cell-specific mirroring pattern defined in a cell. The mirroring pattern controls, on a slot basis, if mirroring

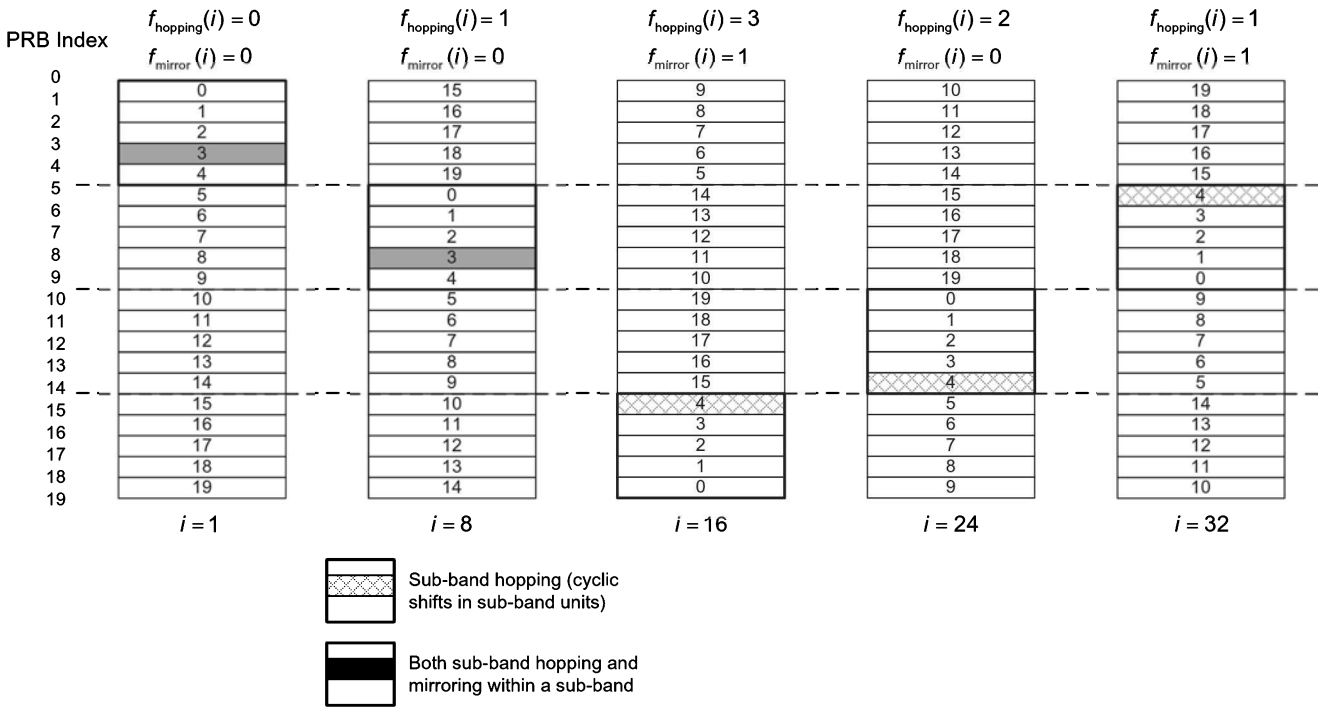


FIGURE 10.13

Illustration of an example PUSCH frequency hopping (20 resource blocks, $L_{sb} = 4$) [156].

within each sub-band should be applied to the assigned resource. In principle, mirroring suggests that the resource blocks within each sub-band are numbered right to left instead of left to right. Figure 10.13 illustrates an example of mirroring in conjunction with hopping within a sub-band. The mirroring pattern is not applied to the first slot, while mirroring is applied to the second slot. Both hopping and mirroring patterns depend on the physical layer cell identity and are typically different in neighboring cells. Furthermore, the periodicity of the hopping/mirroring patterns is one radio frame.

An alternative approach to uplink cell-specific hopping/mirroring is frequency hopping according to explicit information provided in the scheduling grant. In such cases, the scheduling grant includes information about the resource allocation for uplink transmission in the first slot and information related to the amount of offset applicable to the resource to be used for uplink transmission in the second slot relative to the resource location in the first slot. Selection between hopping according to cell-specific hopping/mirroring patterns as described above, or hopping according to explicit information in the scheduling grant, can be made dynamically. More specifically, for cell bandwidths less than 50 resource blocks, there is a flag in the scheduling grant indicating whether frequency hopping is based on cell-specific hopping/mirroring patterns or according to information in the scheduling grant. In the latter case, the hopping is always half of the hopping bandwidth. In the case of larger bandwidths in excess of 50 resource blocks, there are 2 bits in the scheduling grant. One of the combinations indicates that hopping should be according to the cell-specific hopping/mirroring patterns, while the three remaining combinations indicate hopping of 1/2, +1/4, and −1/4 of the hopping bandwidth [16].

The uplink also supports transmission time interval (TTI) bundling where four consecutive subframes are bundled together following a downlink grant. Four HARQ processes are necessary, the downlink PHICH transmissions correspond to the last subframe of the bundle, and the retransmissions are non-dynamic. The purpose of TTI bundling is to enable successful reception of the uplink transmissions in coverage-limited scenarios. The TTI bundling is configured by the RRC sub-layer.

The uplink resource allocation scheme in LTE Rel-8 is single-cluster allocation, which is contiguous resource allocation in the frequency-domain. LTE-Advanced has enabled multi-cluster uplink transmissions. The single-cluster allocations use uplink resource allocation type 0, which is identical to downlink resource allocation type 2, except that the single-bit flag indicating localized/distributed transmission is replaced by a single-bit hopping flag. The resource allocation field in the DCI provides the set of VRBs to be used for uplink transmission. The set of PRBs to use in the two slots of a subframe is controlled by the hopping flag.

The multi-cluster allocations with up to two clusters were introduced in LTE Rel-10 and use uplink resource allocation type 1. In resource allocation type 1, the starting and ending positions of two clusters of frequency-contiguous resource blocks are encoded and indicated with an index. The uplink resource allocation

type 1 does not support frequency hopping since the frequency diversity is achieved through the use of the clustered allocation. It is evident that indication of two clusters of resources requires additional signaling compared to the single-cluster case. However, the total number of bits used for resource allocation type 1 is identical to that of type 0 (see Figure 10.13). This is similar to the case of downlink allocation types 0 and 1. Since frequency hopping is not supported for allocation type 1, the bit, otherwise designated to the hopping flag, can be used for extending the resource allocation field. However, despite the extension of the resource allocation field by 1 bit, the number of bits is not sufficient to provide a single-resource-block resolution in the two clusters for all bandwidths. Instead, similar to downlink resource allocation type 0, groups of resource blocks are used and the starting and ending positions of the two clusters are given in terms of group numbers. The size of the group is proportional to the uplink bandwidth. The size of the group is one for small bandwidths and four for large system bandwidths [16]. Both DCI formats 0 and 4 support uplink resource allocation types 0 and 1. The number of allocated resource blocks is always equal to one or a multiple of 2, 3, or 5. It means that 1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, etc., are permissible resource allocation sizes in the uplink (Figure 10.14).

10.5.2.1 Uplink resource allocation type 0

The resource allocation information contained in uplink resource allocation type 0 for a scheduled UE contains a set of contiguously allocated VRB indices denoted by n_{VRB}. A resource allocation field in the scheduling grant consists of a resource indication value (RIV) parameter corresponding to the starting resource block $\mathrm{RB}_{\mathrm{START}}$ and the length of the allocation in terms of contiguously allocated resource blocks $L \geq 1$. The RIV is defined as $\mathrm{RIV} = N_{\mathrm{RB}}^{\mathrm{UL}}(L-1) + \mathrm{RB}_{\mathrm{START}}$ if $(L-1) \leq \lfloor N_{\mathrm{RB}}^{\mathrm{UL}}/2 \rfloor$; otherwise, the RIV is given as $\mathrm{RIV} = N_{\mathrm{RB}}^{\mathrm{UL}}(N_{\mathrm{RB}}^{\mathrm{UL}} - L + 1) + (N_{\mathrm{RB}}^{\mathrm{UL}} - 1 - \mathrm{RB}_{\mathrm{START}})$. This is similar to resource allocation type 2 in the downlink. An example of uplink resource allocation using type 0 in 1.4 MHz (6 resource blocks) is shown in Figure 10.15.

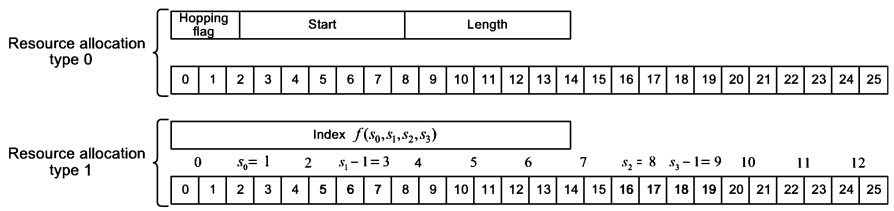


FIGURE 10.14

Illustration of uplink resource allocation types [16].

10.5.2.2 Uplink resource allocation type 1

The uplink resource allocation type 1 was introduced in LTE Rel-10 to allow allocation of non-contiguous resource-block groups (RBGs). In uplink resource allocation type 1, an indication is provided to a scheduled UE of two sets of resource blocks where each set includes one or more consecutive RBGs of size N_{RBG}, assuming a system bandwidth of $N_{\text{RB}}^{\text{UL}}$. This type of resource allocation is defined by four indices that are combined into a single combinatorial index l before being sent to the UE. A complete representation of combinatorial index l requires $\left\lceil \log_2 \left[\binom{\lceil N_{\text{RB}}^{\text{UL}}/N_{\text{RBG}} + 1 \rceil}{4} \right] \right\rceil$ bits. The bits from the resource allocation field in the scheduling grant represent index l, unless the number of bits in the resource allocation field in the scheduling grant is either smaller or larger than that required number to represent index l. If the number of bits is smaller than that required to represent l, the bits in the resource allocation field in the scheduling grant would represent the least-significant bits of l and the value of the remaining bits of l are set to zero. If the number of bits is larger than that required for representation of l, index l would be mapped to the least-significant bits of the resource allocation field in the scheduling grant. Note that resource-block allocation in this case is defined by four indices (s_0, s_1, s_2, s_3). The parameters s_0 and s_1 define the first and the last RBG within the first set of RBGs, and the parameters s_2 and s_3 define the first and the last RBG within the second set of RBGs. Therefore, the combinatorial index l corresponds to the starting and the ending index of resource-block-set 1 (s_0 and s_1), and resource-block-set 2 (s_2 and s_3), respectively, where $l = \sum_{i=0}^{M-1} \binom{N - s_i}{M - i}$, $M = 4$, and $N = \lceil N_{\text{RB}}^{\text{UL}}/N_{\text{RBG}} \rceil + 1$. If the corresponding ending RBG index is equal to the starting index, only one RBG is allocated.[2]

---

[2]The combinatorial index l is calculated based on the sorted set of M sub-band indices $\{s_i\}_{i=0,1,\ldots,M-1}$ given $1 \le s_k \le N, s_k < s_{k+1}$ as $l = \sum_{i=0}^{M-1} \binom{N - s_i}{M - i}$ where $\binom{x}{y} : x \ge y$ is the extended binomial coefficient and $l \in \left\{0, 1, \ldots, \binom{N}{M} - 1\right\}$. The uniqueness and the range of the combinatorial index are proven in the literature [162,163]. In the receiving end, the de-mapper regenerates the set of indices $\{s_i\}_{i=0,1,\ldots,M-1}$ from the received label l using the following algorithm:

$x_{\min} = 1;$
for $i = 0$ to $M - 1$
$x = x_{\min};$
$p = \binom{N - x}{M - i};$
while $p > l$
  $x = x + 1;$
  $p = \binom{N - x}{M - i};$
end
$s_i = x;$
$x_{\min} = s_i + 1;$
$l = l - p;$
end

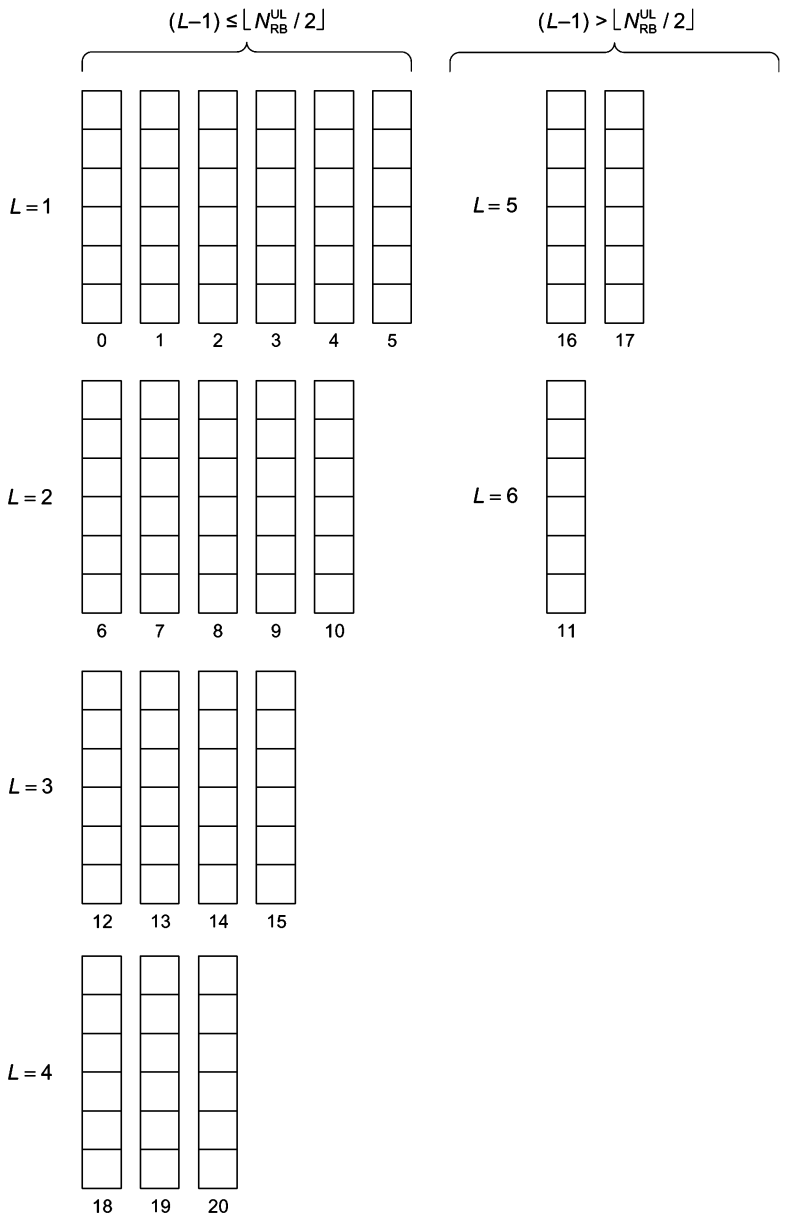


FIGURE 10.15

Illustration of an example uplink resource allocation type 0 in 1.4 MHz [14].

Table 10.4 Subframe Delay k Between PDCCH Assignment and PUSCH Transmission for TDD Mode [5]

TDD Uplink/Downlink Configuration (DL:UL)	PDCCH/PHICH Subframe Number n									
	0	1	2	3	4	5	6	7	8	9
0 (2:3)	4	6	–	–	–	4	6	–	–	–
1 (3:2)	–	6	–	–	4	–	6	–	–	4
2 (4:1)	–	–	–	4	–	–	–	–	4	–
3 (7:3)	4	–	–	–	–	–	–	–	4	4
4 (8:2)	–	–	–	–	–	–	–	–	4	4
5 (9:1)	–	–	–	–	–	–	–	–	4	–
6 (5:5)	7	7	–	–	–	7	7	–	–	5

An important consideration concerning the uplink resource allocation is the timing relevance of PDCCH uplink assignment (or PHICH transmission) and the uplink subframe in which physical resources are reserved for the UE PUSCH transmission or retransmission. In FDD mode, the uplink resource allocation using PDCCH (or PHICH transmission) for the UE in subframe n is related to PUSCH transmission/retransmission in subframe $n + 4$, i.e., there is a delay of four subframes between the uplink resource assignment (or HARQ ACK/NACK) in the downlink and the corresponding uplink transmission. For TDD mode, the timing relevance of PDCCH assignment (or PHICH transmission) and PUSCH transmission/retransmission is more complicated and depends on the downlink/uplink subframe configuration. The delay (in number of subframes) between receiving an uplink resource assignment in PDCCH or HARQ ACK/NACK in PHICH in subframe n and PUSCH transmission in subframe $n + k$ (where $k \geq 4$ and is selected such that $n + k$ is an uplink subframe when the ACK/NACK can be transmitted from the UE on PUCCH or PUSCH) for TDD mode for various downlink/uplink configurations is given in Table 10.4 [5]. It can be noted that, due to certain limitations imposed by downlink/uplink subframe configuration, PUSCH transmission may be delayed by up to seven subframes.

The uplink/downlink subframe configuration 0 is a special case because the number of uplink subframes is greater than the number of downlink subframes. Due to an insufficient number of downlink subframes, the downlink portion of the special subframe is also used for downlink transmission.

Multiplexing is a method for grouping and sending several independent HARQ ACK/NACKs in the TDD mode in a single uplink subframe for the previously received transport blocks. This method allows independent retransmission of erroneous transport blocks. However, it also implies that multiple ACK/NACK bits need to be transmitted from the UE, which can potentially limit the uplink coverage. Alternatively, bundling of HARQ ACK/NACKs implies that the outcome of the decoding of downlink transport blocks from multiple downlink subframes can

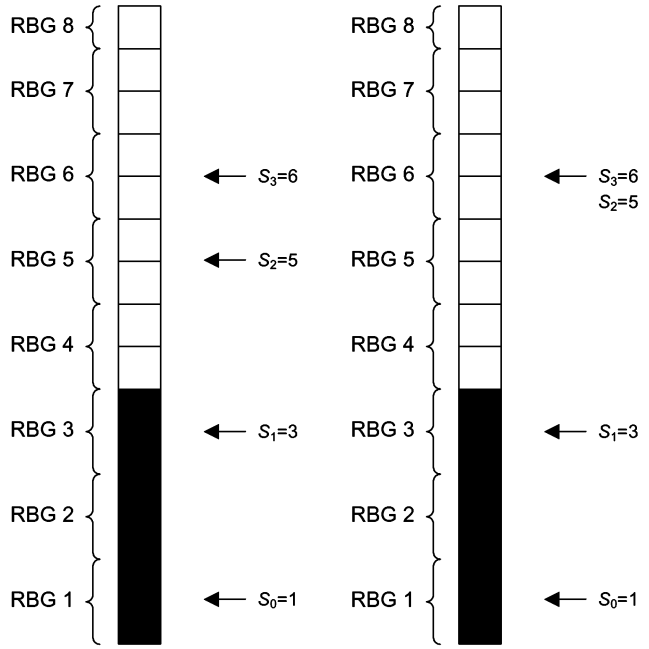


FIGURE 10.16

Example of uplink resource allocation type 1 in 3 MHz [14].

be combined into a single acknowledgement in the uplink. Combining acknowledgments related to multiple downlink transmissions into a single uplink message assumes that the terminal has not missed any of the scheduling assignments on which the ACK is based. As an example, assume that the eNB schedules the UE in two (subsequent) subframes, but the terminal misses the PDCCH transmission in the first subframe and successfully decodes the control information in the second subframe. The terminal would transmit one ACK as if it was scheduled in the second subframe only, while the eNB would interpret the ACK as if both transmissions were successfully received by the UE. To avoid such misinterpretations, the downlink assignment index (DAI) field in the scheduling assignment on the PDCCH is used. The DAI informs the terminal about the number of transmissions it should expect before generating a combined acknowledgment. If there is a mismatch between the DAI and the number of transmissions the terminal actually received, the terminal concludes that some assignments are missing. As a result, the UE does not transmit any HARQ acknowledgment [15] (Figure 10.16).

Let us consider a TDD example, downlink/uplink subframe configuration 2 includes six downlink subframes, two special subframes, and two uplink subframes. Downlink data can be transmitted within the DwPTS portion of the special subframes, which effectively results in eight downlink transmission opportunities.

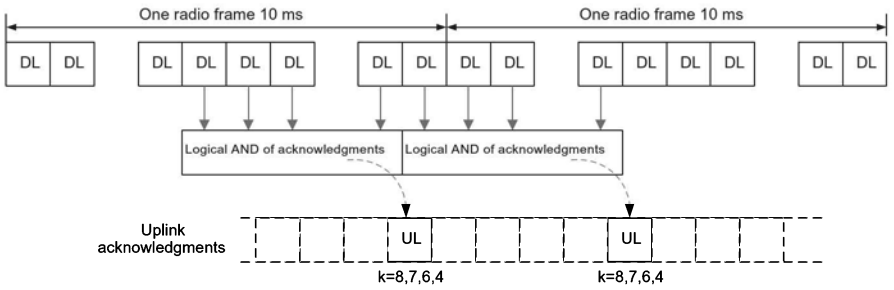


FIGURE 10.17

Illustration of HARQ acknowledgment bundling for TDD frame configuration 2 [14].

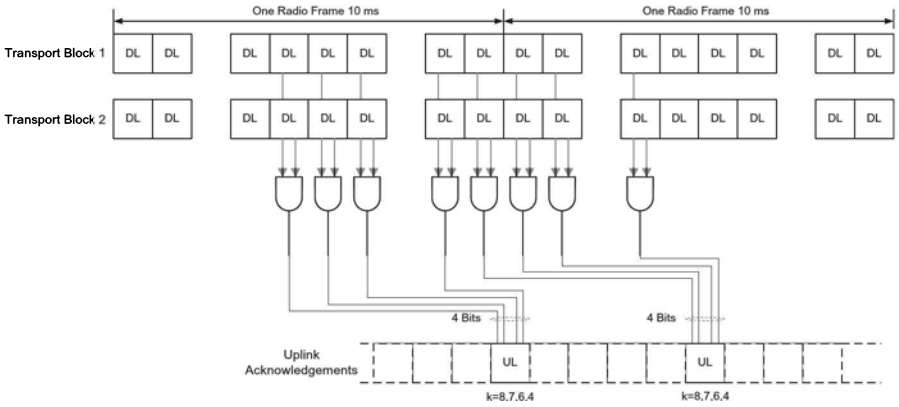


FIGURE 10.18

Illustration of HARQ ACK multiplexing for TDD frame configuration 2 [14].

If the downlink transmissions are divided into two groups and a single acknowledgment is generated from each group, each uplink subframe is required to carry a single HARQ ACK/NACK, which is called ACK/NACK bundling as depicted in Figure 10.17. On the other hand, for the same subframe configuration, 16 acknowledgments per radio frame can be reduced to eight acknowledgments using two uplink subframes. PUCCH format 1b can be used to carry four acknowledgments when channel selection is used, i.e., the PUCCH payload carries two ACKs and the selection of PUCCH resources carriers the other two ACKs, as illustrated in Figure 10.18 [14].

The use of bundled ACKs in a single uplink subframe is not limited to TDD mode. The carrier aggregation in FDD with a different number of uplink and downlink component carriers encounters the same problem. In LTE Rel-10, there is only one PUCCH that is always transmitted on the primary component carrier.

Hence, even in asymmetric carrier aggregation scenarios, the possibility to transmit more than two HARQ ACK bits in the uplink must be supported. This is performed via resource selection or by using PUCCH format 3. The ACKs in response to PUSCH transmissions are transmitted using multiple PHICHs, where the PHICH is transmitted on the same downlink component carrier as the uplink scheduling grant initiating the uplink transmission.

10.6 Uplink reference signals and channel estimation

There are two types of reference signals specified for the LTE/LTE-Advanced uplink. The uplink demodulation reference signals are used by the eNB for channel estimation and coherent demodulation of uplink physical channels (PUSCH and PUCCH). The demodulation reference signals are only transmitted in conjunction with PUSCH or PUCCH within the same bandwidth as the corresponding physical channel. Uplink demodulation reference signals are used by the eNB for channel state information (CSI) estimation, and to support uplink channel-dependent scheduling and link adaptation. The demodulation reference signals can also be used in cases when uplink transmission is needed, although there is no data to transmit. An example would be to enable the network to adjust uplink frame timing via an uplink timing alignment procedure. The uplink demodulation reference signals can also be used to estimate the downlink CSI from uplink transmission, when sufficient uplink/downlink reciprocity exists.

Since maintaining low CM value and PA efficiency for uplink transmissions is very important, the requirements for design of uplink reference signals are different compared to those in the downlink. In principle, transmitting reference signals together with other uplink transmissions from the same terminal is not efficient for the uplink. Instead, certain SC-FDMA symbols are exclusively designated for DM-RS transmission. Therefore, the reference signals are time-multiplexed with other uplink transmissions from the same terminal. The structure of the reference signal per se ensures a low CM value over the reference symbols. In the case of PUSCH transmission, a DM-RS is transmitted over the fourth SC-FDMA symbol of each uplink slot. In the case of PUCCH transmission, the number of SC-FDMA symbols used for a reference signal in a slot, as well as the exact position of these symbols, differs among different PUCCH formats [16].

In general, there is no advantage in estimating the channel outside the frequency band of the corresponding PUSCH or PUCCH transmission that is going to be coherently demodulated. The bandwidth of the reference signal corresponding to the length of the reference signal sequence is equal to the bandwidth of the corresponding PUSCH or PUCCH transmission measured in the number of subcarriers. The latter requirement suggests that, for PUSCH transmission, it should be possible to generate reference signal sequences of different length corresponding to the possible bandwidths of PUSCH transmission. However, it must be noted that the length of the reference signal sequence must be a multiple of 12

(i.e., the size of a basic PRB), since uplink resource allocations for PUSCH transmissions are always measured in terms of the number of resource blocks. The PUCCH transmission from a terminal is always carried out over a single resource block in the frequency-domain. In the case of PUCCH transmission, the length of the reference signal sequence is equal to 12. The uplink reference signals should preferably have the following properties [121]:

- Limited power variation in the frequency-domain to allow similar channel estimation quality for all frequencies, or constant amplitude in the frequency-domain for equal excitation of all the allocated sub-carriers for unbiased channel estimates.
- Low CM value, no worse than data transmissions, in the time-domain. Note that a lower CM value for reference signals relative to the data would allow power boosting of the reference signals at the cell-edge.
- Good autocorrelation property for accurate channel estimation.
- Good cross-correlation property to reduce interference among different reference signals transmitted on the same resources by neighboring cells.

We will discuss how the above requirements have been addressed in the design of the LTE/LTE-Advanced uplink reference signals.

10.6.1 Uplink demodulation reference signals

The uplink DM-RS design in LTE Rel-8 is based on extended Zadoff–Chu sequences. This choice is because of the low PAPR property of these sequences, as well as the possibility of using mechanisms for orthogonalizing multiple DM-RS sequences which are utilized in spatial multiplexing schemes. The use of different cyclic shifts of a root Zadoff–Chu sequence in the time-domain allows multiple users to transmit their (orthogonally code-division multiplexed) DM-RS to the same eNB provided that the reference signals are transmitted in the same set of resource blocks [121].

In LTE-Advanced, SU-MIMO with up to four-layer spatial multiplexing is supported in the uplink. While the uplink DM-RS corresponding to different layers can be theoretically orthogonalized via assigning different cyclic shifts to each layer, the orthogonality property of the cyclically shifted sequences may not hold in high-delay-spread scenarios. Furthermore, the orthogonality of the cyclically shifted sequences is guaranteed only when the root Zadoff–Chu sequences have identical waveform and length, which implies that the time–frequency resource allocation of MU-MIMO UEs should be identical. The uplink DM-RS in LTE-Advanced are designed to overcome these drawbacks, while maintaining backward compatibility.

In LTE/LTE-Advanced, two SC-FDMA symbols (one DM-RS per slot) are configured for DM-RS in a subframe within the allocated time–frequency resources as shown in Figure 10.19 for PUSCH transmission. Each uplink demodulation reference signal sequence $r_{uv}(k, \tau)$ associated with PUSCH transmission is defined by the time-domain cyclic shift $\tau(\lambda)$ of the base sequence $r_{uv}(k)$ according to

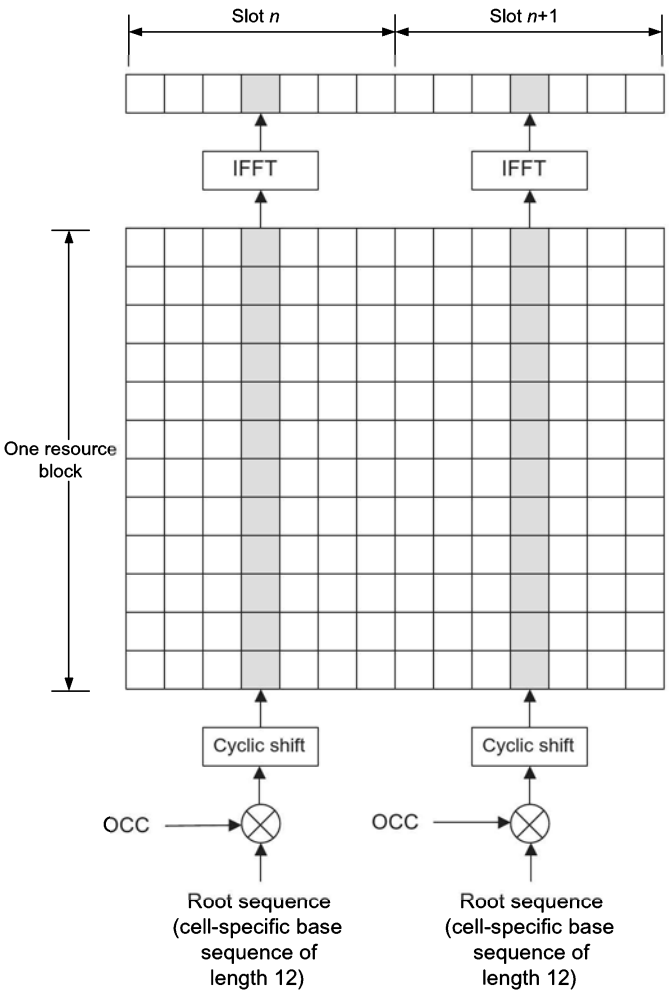


FIGURE 10.19

Uplink DM-RS generation and structure in time and frequency [121,14].

$r_{uv}(mL_{RS}+k,\tau) = w_m(\lambda)e^{j\tau(\lambda)k}\, r_{uv}(k) | 0 \le k < L_{RS}-1, m = 0,1,$ where $L_{RS} = 12m'$ (i.e., the length of the reference signal is an integer multiple of the basic resource-block size in frequency) is the length of the reference signal sequence, $1 \le m' \le \max(N_{RB}^{UL})$, and $\lambda \in \{0,1,\ldots,N_l-1\}$ denotes the spatial layer index. Using the latter expression, multiple reference signal sequences can be generated from a single base sequence through different values of cyclic shift $\tau(\lambda)$ [3].

The base sequences $r_{uv}(k)$ are divided into groups where the index $u \in \{0,1,\ldots,29\}$ represents the group number and v is the base sequence number within the group, such that each group contains one base sequence $v = 0$ for each

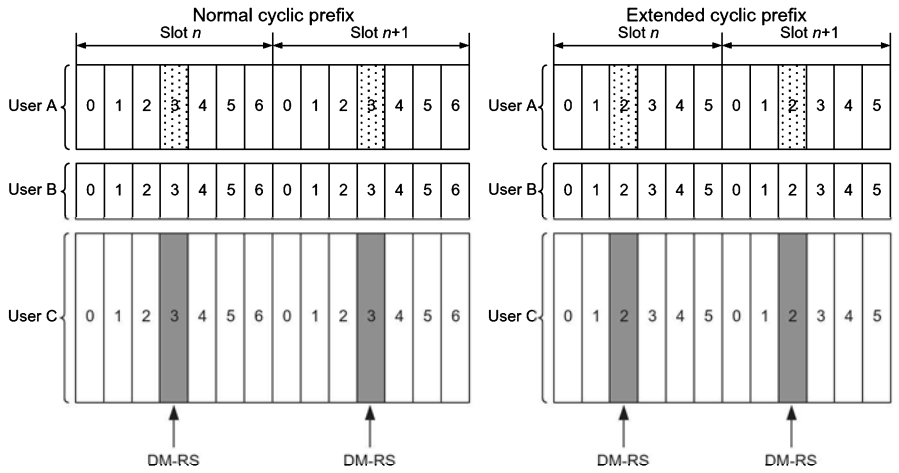


FIGURE 10.20

PUSCH DM-RS location in time and frequency [155].

length $L_{RS} = 12m' | 1 \leq m' \leq 5$ and two base sequences denoted by $v \in \{0, 1\}$ of each length $L_{RS} = 12m' | 6 \leq m' \leq \max(N_{RB}^{UL})$. The sequence-group number u and the index v within the group may vary across time. The definition of the base sequence $\{r_{uv}(0), r_{uv}(1), \ldots, r_{uv}(L_{RS} - 1)\}$ depends on the sequence length L_{RS} (Figure 10.21).

The time-domain cyclic shift (post IFFT in the OFDM modulation) can be interpreted as a phase rotation in the frequency-domain (pre-IFFT in the OFDM modulation). For frequency non-selective channels over 12 sub-carriers constituting a resource block, it is possible to achieve orthogonality among DM-RSs generated from the same base sequence, if $\tau(\lambda) = \frac{k\pi}{6}, k = 0, 1, \ldots, 11$, assuming the DM-RSs are synchronized in time. The orthogonality can be exploited to transmit DM-RS at the same time, using the same frequency resources without interference. In general, DM-RS generated from different base sequences will not be orthogonal; however, they will present low cross-correlation properties. To maximize the number of available Zadoff–Chu sequences, a prime length sequence is needed. The minimum sequence length in the uplink is 12 (i.e., the number of sub-carriers in a resource block), which is not prime number. Therefore, Zadoff–Chu sequences are not suitable by themselves; thus, there are effectively two types of base reference sequences: (1) those with a sequence length larger than 36 sub-carriers (spanning three or more resource blocks), which use a cyclic extension of Zadoff–Chu sequences; and (2) those with a sequence length less than 36 sub-carriers (spanning one or two resource blocks where the latter use a special QPSK sequence ($r_{uv}(k) = e^{j\phi(k)\pi/4}, \quad 0 \leq k \leq L_{RS} - 1$) [3].

For sequences of length three resource blocks and larger, the base sequence is a repetition, with a cyclic offset, of a Zadoff–Chu sequence of length N_{ZC}^{RS}, where

N_{ZC}^{RS} is the largest prime number that satisfies the condition $N_{ZC}^{RS} < L_{RS}$. Therefore, the base sequence will contain one complete length N_{ZC}^{RS} Zadoff–Chu sequence plus a fractional repetition appended on the end. At the receiver, the appropriate de-repetition can be done and the zero autocorrelation property will hold across the length N_{ZC}^{RS} vector. For sequences shorter than three resource blocks, the sequences are a composition of unity modulus complex numbers derived from computer simulation, and are defined in the LTE specifications [3].

There are a total of 30 sequence groups, each containing one sequence with length shorter than 60 sub-carriers. This corresponds to transmission bandwidths of 1, 2, 3, 4, and 5 resource blocks. Additionally, there are two sequences (one for $v = 0$ or 1) with length greater than 72 sub-carriers corresponding to transmission bandwidths of six resource blocks or more. Note that not all values of m' are allowed, where m' is the number of resource blocks used for transmission of the PUSCH DM-RS. Only values which are the product of powers of 2, 3, and 5 are valid, i.e., $m' = 2^{\alpha_1}3^{\alpha_2}5^{\alpha_3}$ and $\alpha_1, \alpha_2,$ and α_3 are non-negative integers. This is due to the fact that the DFT sizes of the SC-FDMA precoding operation are limited to values which are the product of powers of 2, 3, and 5. The DFT operation can span more than one resource block. Since each resource block has 12 sub-carriers, the total number of sub-carriers fed to the DFT will be equal to $12m'$. Given that the value of $12m'$ has to be a product of powers of 2, 3, and 5, this implies that the number of resource blocks must themselves be the product of powers of 2, 3, and 5. Therefore, values of m' such as 7, 11, 14, or 19 are not valid [16].

For a given time slot, the uplink reference signal sequences to use within a cell are taken from a specific sequence group. If the same group is used for all slots, then this is known as fixed assignment. On the other hand, if the group number u varies for all slots within a cell, this is known as group hopping. The cell system information determines which method to use, and the group assignment depends on the type of physical channel (PUSCH or PUCCH). The PUSCH demodulation reference signal is mapped to the fourth SC-FDMA symbol of the slot for a normal cyclic prefix and to every third SC-FDMA slot for an extended cyclic prefix, as shown in Figure 10.20.

The LTE Rel-10 uses codebook-based precoding for uplink SU-MIMO data transmissions, and obviously the same precoding mechanism is applied to the uplink DM-RS. As such, a receiver can directly estimate the spatially precoded channel. Note that the uplink DM-RSs are distributed in the same manner as data when clustered SC-FDMA is used. An eNB can choose root Zadoff–Chu sequences for the reference signal by considering the autocorrelations and cross-correlations of the sequences that are in use among neighboring eNBs. However, such coordination is not always practically feasible for cellular operators. Hence, a root Zadoff–Chu sequence randomization mechanism, referred to as sequence hopping and sequence-group hopping, is available depending on each operator's choice, where the root Zadoff–Chu sequence varies in every slot, if sequence-group hopping is enabled. The spreading factor $w(\lambda_i) = \pm 1 | i = 0, 1, 2, 3$ is an

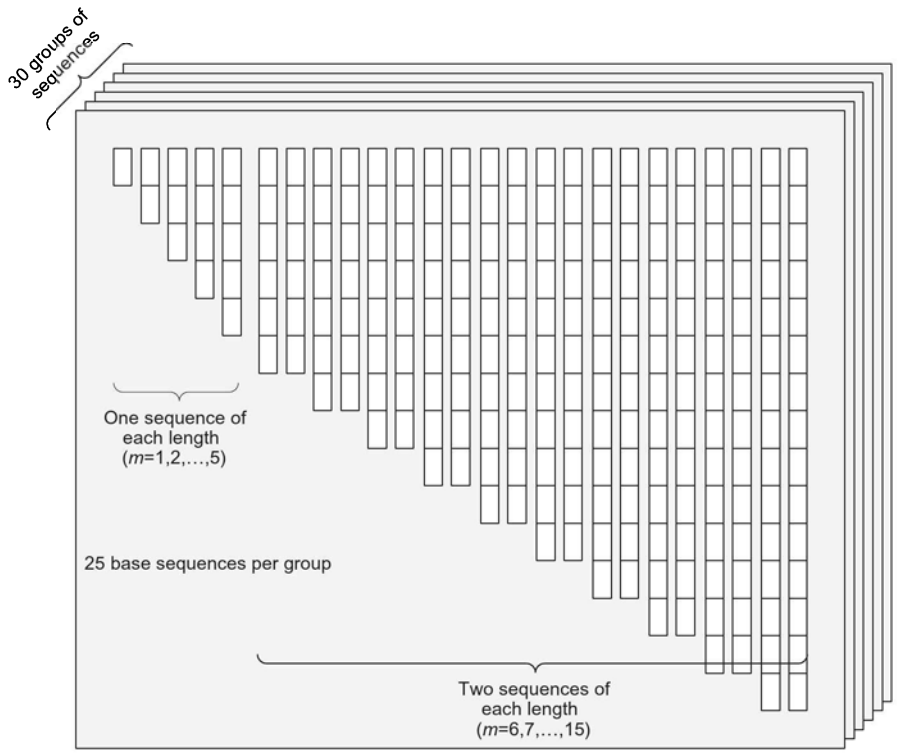


FIGURE 10.21

Illustration of base sequence grouping used for PUCCH/PUSCH DM-RS in 3 MHz [14].

orthogonal cover code (OCC), which is used to spread the reference signal sequence over two SC-FDMA symbols. The introduction of OCC was motivated in order to efficiently support SU-MIMO and MU-MIMO operation in the uplink (see Figure 10.19).

For $L_{RS} \geq 36$, the base sequence $\{r_{uv}(0), r_{uv}(1), \ldots, r_{uv}(L_{RS}-1)\}$ is given by $r_{uv}(k) = x_l(k \bmod N_{ZC}^{RS}) \mid \quad 0 \leq k < L_{RS} - 1$, where the lth-root Zadoff–Chu sequence is defined by the following expression [3]:

$$x_l(m) = -\exp\left(j\frac{\pi l m(m+1)}{N_{ZC}^{RS}}\right), \quad 0 \leq m \leq N_{ZC}^{RS} - 1, \quad l = \lfloor l' + 0.5 \rfloor + (-1)^{\lfloor 2l' \rfloor} v, \quad l' = N_{ZC}^{RS}(u+1)/31 \tag{10.3}$$

The length N_{ZC}^{RS} of the Zadoff–Chu sequence is given by the largest prime number such that $N_{ZC}^{RS} < L_{RS}$. The sequence-group number u in slot n_s is defined by a group hopping pattern $f_{group\text{-}hopping}(n_s)$ and a sequence-shift pattern $f_{shift\text{-}pattern}$ according to $u = [f_{group\text{-}hopping}(n_s) + f_{shift\text{-}pattern}] \bmod 30$. There are 17 different

hopping patterns and 30 different sequence-shift patterns. Sequence-group hopping can be enabled or disabled through a cell-specific parameter signaled by higher layers (see Figure 10.20) [3].

The DM-RS group hopping pattern is defined as $f_{\text{group-hopping}}(n_s) = \left[\sum_{i=0}^{7} c(8n_s + i)2^i\right] \bmod 30$ when group hopping is enabled and is set to zero, otherwise, which is the same for PUSCH and PUCCH. The pseudo-random sequence $c(i)$ was defined in Chapter 9 and is initialized with $c_{\text{init}} = \lfloor N_{\text{ID}}^{\text{RS}}/30 \rfloor$ at the beginning of each radio frame.[3] Note that the definition of the sequence-shift pattern $f_{\text{shift-pattern}}$ is different for PUCCH and PUSCH. For PUCCH, the sequence-shift pattern is given by $f_{\text{shift-pattern}}^{\text{PUCCH}} = N_{\text{ID}}^{\text{RS}} \bmod 30$. For PUSCH, the sequence-shift pattern is given by $f_{\text{shift-pattern}}^{\text{PUSCH}} = (N_{\text{ID}}^{\text{cell}} + \Delta_{\text{ss}}) \bmod 30 \quad |\Delta_{\text{ss}} \in \{0, 1, \ldots, 29\}$ and is configured via higher-layer signaling [3], if no value for $n_{\text{ID}}^{\text{PUSCH}}$ is provided by the higher layers, or if the PUSCH transmission corresponds to a *Random Access Response* or a retransmission of the same transport block as part of the contention-based random access procedure; otherwise, it is defined as $f_{\text{ss}}^{\text{PUSCH}} = N_{\text{ID}}^{\text{RS}} \bmod 30$. The parameter Δ_{ss} is cell-specific and is broadcast in SIB2 which defines an offset between 0 and 29 that can be used to select different sequence groups for cells with the same physical cell identifier $N_{\text{ID}}^{\text{cell}}$. The concept of virtual cell identity was introduced in LTE Rel-11 to overcome the limitation of earlier releases of LTE technology in new deployment scenarios related to CoMP, HetNet, etc., where several low-power nodes or subsets of a macro-cell may share the same physical layer cell identity with the macro-cell. Sharing of the same cell identifier by the neighboring nodes would increase intra-cell interference and degrade the channel estimation performance and coherent detection of data and control channels, as well as limiting the use of advanced closed-loop SU-MIMO/MU-MIMO beamforming techniques. Also, the use of the same identifier by subordinate nodes would prevent transmission/reception node identification and switching. The use of different scrambling sequences initialized with unique cell identifiers would achieve interference randomization and help the UE to separate data and control channels transmitted by different transmission nodes. Since the reference signals (e.g., CRS, DM-RS, CSI-RS) are also scrambled with cell-specific sequences to allow the UE to conduct various channel estimations and mobility measurements on different transmission nodes, the use of unique scrambling sequences would facilitate these functions as well as node selection and switching in homogeneous and heterogeneous networks.

[3]The definition of virtual cell identity $N_{\text{ID}}^{\text{RS}}$ (a new concept in LTE Rel-11) depends on the type of uplink physical channel [3]. For transmissions on PUSCH, $N_{\text{ID}}^{\text{RS}} = N_{\text{ID}}^{\text{cell}}$, if no value for $N_{\text{ID}}^{\text{PUSCH}}$ is configured by the higher layers or if the temporary C-RNTI was used to transmit the most recent uplink-related DCI for the transport block associated with the corresponding PUSCH transmission; otherwise $N_{\text{ID}}^{\text{RS}} = N_{\text{ID}}^{\text{PUSCH}}$. For transmissions on PUCCH, the virtual cell identity $N_{\text{ID}}^{\text{RS}} = N_{\text{ID}}^{\text{cell}}$, if no value for $N_{\text{ID}}^{\text{PUCCH}}$ is configured by the higher layers; otherwise $N_{\text{ID}}^{\text{RS}} = N_{\text{ID}}^{\text{PUCCH}}$. For sounding reference signal transmission, the virtual cell identity is defined as $N_{\text{ID}}^{\text{RS}} = N_{\text{ID}}^{\text{cell}}$ [3,5].

The sequence hopping is only applied to reference signals whose length $N_{RS} \geq 72$ and the base sequence number defined as $v = c(n_s)$ within the base sequence group in slot n_s, if group hopping is disabled but sequence hopping is enabled. The pseudo-random sequence $c(i)$ was defined earlier and is initialized with $c_{init} = \lfloor N_{ID}^{RS}/30 \rfloor 2^5 + f_{shift\text{-}pattern}^{PUSCH}$ at the beginning of each radio frame. For reference signals of length $N_{RS} < 72$, the base sequence number v within the base sequence group is given by $v = 0$ [3].

The cyclic shift $\tau(\lambda)$ in slot n_s is defined as $\tau(\lambda) = \pi n_{CS}(\lambda)/6$, where $n_{CS}(\lambda) = [n_{DM\text{-}RS}^{(1)} + n_{DM\text{-}RS}^{(2)}(\lambda) + n_{PN}(n_s)] \bmod 12$. The values of parameters $n_{DM\text{-}RS}^{(1)}$ and $n_{DM\text{-}RS}^{(2)}(\lambda)$ are determined based on the value of the cyclic shift for the DM-RS field in the latest uplink grant within DCI format 0 or 4, and are given in Table 10.5. The first row of Table 10.5 can be used to obtain $n_{DM\text{-}RS}^{(2)}(0)$ and $w_m(\lambda)$, if there is no uplink-related DCI for the same transport block associated with the corresponding PUSCH transmission, and if the initial PUSCH for the same transport block is semi-persistently scheduled, or if the initial PUSCH for the same transport block is scheduled by the random-access response grant. The parameter $n_{PN}(n_s)$ is given by $n_{PN}(n_s) = \sum_{i=0}^{7} c(8N_s^{UL}n_s + i)2^i$, where the cell-specific pseudo-random sequence $c(i)$ was defined in Chapter 9. The pseudo-random sequence generator in this case is initialized with $c_{init} = \lfloor N_{ID}^{cell}/30 \rfloor 2^5 + f_{shift\text{-}pattern}^{PUSCH}$ at the beginning of each radio frame, if no value for $N_{ID}^{PUSCH\text{-}DM\text{-}RS}$ is configured by the higher layers, or if the temporary C-RNTI was used to transmit the most recent uplink-related DCI for the transport block associated with PUSCH transmission; otherwise, it is defined as $c_{init} = \left\lfloor \frac{N_{ID}^{PUSCH\text{-}DM\text{-}RS}}{30} \right\rfloor 2^5 + (N_{ID}^{PUSCH\text{-}DM\text{-}RS} \bmod 30)$ [3].

The vector of the UE-specific DM-RSs is precoded according to the following equation:

$$[\tilde{r}_{PUSCH}^{(0)}, \tilde{r}_{PUSCH}^{(1)}, \ldots, \tilde{r}_{PUSCH}^{(N_P-1)}]^T = \mathbf{W}[r_{PUSCH}^{(0)}, r_{PUSCH}^{(1)}, \ldots, r_{PUSCH}^{(N_l-1)}]^T \quad (10.4)$$

where N_P denotes the number of antenna ports used for PUSCH transmission and $\mathbf{W}$ is the precoding matrix. For PUSCH transmission using a single antenna port $N_P = 1, W = 1$, and $N_l = 1$. For spatial multiplexing $N_P = 2$ or $N_P = 4$ and the precoding matrix $\mathbf{W}$ is identical to the precoding matrix used for precoding of the PUSCH in the same subframe [3].

In uplink SU-MIMO where a UE transmits multiple spatial layers simultaneously, multiple DM-RS can be orthogonally multiplexed by assigning a different cyclic shift value for each layer. However, as mentioned earlier, the orthogonality among the sequences may not be preserved when the delay spread is high due to increased correlation among the spatial layers. To overcome this limitation, the OCCs were introduced in LTE-Advanced. As shown in Figure 10.19 and further emphasized in Table 10.5, different OCCs are applied for higher-numbered layers (i.e., $\lambda \geq 3$) and lower-numbered layers (i.e., $\lambda \geq 2$), so reference signals associated with the higher rank transmissions (i.e., $N_l \geq 3$)

Table 10.5 OCC and Cyclic Shift Parameters for Uplink DM-RS [3]

Cyclic Shift Value Carried in Uplink Grant	Layer-Dependent Cyclic Shift					Layer-Dependent Orthogonal Cover Code $(w_0(\lambda), w_1(\lambda))$			
	$n^{(1)}_{\text{DM-RS}}$	$n^{(2)}_{\text{DM-RS}}(\lambda)$							
		$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 0$	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$
000	0	0	6	3	9	[1 1]	[1 1]	[1 −1]	[1 −1]
001	2	6	0	9	3	[1 −1]	[1 −1]	[1 1]	[1 1]
010	3	3	9	6	0	[1 −1]	[1 −1]	[1 1]	[1 1]
011	4	4	10	7	1	[1 1]	[1 1]	[1 1]	[1 1]
100	6	2	8	5	11	[1 1]	[1 1]	[1 1]	[1 1]
101	8	8	2	11	5	[1 −1]	[1 −1]	[1 −1]	[1 −1]
110	9	10	4	1	7	[1 −1]	[1 −1]	[1 −1]	[1 −1]
111	10	9	3	0	6	[1 1]	[1 1]	[1 −1]	[1 −1]

are still sufficiently orthogonal, even though cyclic shift separation is small. Although the orthogonality provided by the OCCs may also fall apart due to high Doppler shift, the effect is minimal because higher rank transmission is mainly used for stationary or nomadic UEs [121].

The generation of LTE Rel-8/9 DM-RS sequences (no OCC is applied) for PUSCH in each uplink slot can be summarized as follows:

1. Determine whether group hopping and sequence hopping are enabled from higher layer signaling.
2. Obtain Δ_{ss} and the cyclic shift from higher layer signaling.
3. Calculate $c_{init} = \lfloor N_{ID}^{cell}/30 \rfloor$ thus find the scrambling sequence $c(i)$.
4. Obtain $n_{DM\text{-}RS}^{(1)}, n_{DM\text{-}RS}^{(2)}(\lambda)$ from Table 10.5 given the cyclic shift value is known.
5. Calculate $f_{group\text{-}hopping}(n_s) = \left[\sum_{i=0}^{7} c(8n_s + i)2^i\right] \bmod 30$ and $v = c(n_s)$ or 0 given n_s is known.
6. Calculate $f_{shift\text{-}pattern}^{PUSCH} = (N_{ID}^{cell} + \Delta_{ss}) \bmod 30$.
7. Calculate $u = \left[f_{group\text{-}hopping}(n_s) + f_{shift\text{-}pattern}\right] \bmod 30$.
8. Calculate $n_{PN}(n_s) = \sum_{i=0}^{7} c(8N_s^{UL}n_s + i)2^i$.
9. Calculate $n_{CS}(\lambda) = \left[n_{DM\text{-}RS}^{(1)} + n_{DM\text{-}RS}^{(2)}(\lambda) + n_{PN}(n_s)\right] \bmod 12$.
10. Calculate $\tau(\lambda) = \pi n_{CS}(\lambda)/6$.
11. Calculate lth-root Zadoff–Chu sequence $x_l(m) = -\exp\left(j\frac{\pi lm(m+1)}{N_{ZC}^{RS}}\right)$ given that $0 \le m \le N_{ZC}^{RS} - 1$, $l = \lfloor l' + 0.5 \rfloor + (-1)^{\lfloor 2l' \rfloor} v, l' = N_{ZC}^{RS}(u+1)/31$ (*Note that N_{ZC}^{RS} is the largest prime number less than the size of PUSCH allocation or L_{RS}*).
12. Calculate $r_{uv}(mL_{RS} + k, \tau) = e^{j\tau(\lambda)k} r_{uv}(k) | 0 \le k < L_{RS} - 1, m = 0, 1$.

The MU-MIMO schemes can be categorized into two classes from a DM-RS orthogonality perspective, i.e., identical-band MU-MIMO and non-identical-band MU-MIMO [121]. Identical-band MU-MIMO is a scheme in which the frequency band allocations for MU-MIMO-capable UEs are identical and the DM-RS orthogonality can be achieved by assigning different cyclic shifts. Although this scheme is LTE Rel-8/9 compatible and can be used for multiplexing LTE Rel-8/9 and Rel-10 UEs, the scheduling restriction may adversely impact the uplink system throughput. Non-identical-band MU-MIMO is a scheme in which the frequency band allocations for MU-MIMO UEs are not identical. Although this improves frequency scheduling flexibility and may improve the uplink system throughput, it implies that the DM-RS root Zadoff–Chu sequences for the two slots of a subframe should be the same in order to utilize the OCCs. Therefore, the OCCs can be used to realize non-identical-band MU-MIMO [121].

10.6.2 Sounding reference signals

The SRS are physical signals transmitted in the uplink to enable the eNB to estimate the CSI over a range of frequencies. The estimation of the CSI assists the

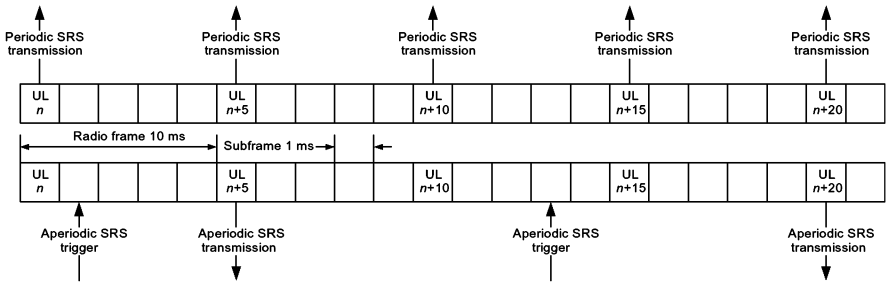


FIGURE 10.22

Illustration of periodic and aperiodic SRS transmission [121].

eNB scheduler to properly allocate radio resources to the UE, and to select different transmission parameters such as the instantaneous data rate and uplink multi-antenna transmission parameters for optimized uplink transmission (i.e., uplink channel-dependent scheduling). The SRS transmission can be further used for uplink timing estimation as well as to estimate downlink channel conditions assuming downlink/uplink channel reciprocity. When uplink/downlink channel reciprocity is assumed, SRS measurements can also be used to support downlink transmissions such as angle-of-arrival (AoA) measurements on SRS to support downlink beamforming [14]. Therefore, SRS is not necessarily transmitted together with an uplink physical channel and when transmitted in conjunction with PUSCH, the SRS may cover a different and typically larger frequency region. There are two types of SRS specified for uplink transmission, i.e., periodic and aperiodic SRS (Figure 10.22).

The subframe in which SRS is transmitted by any UE within the cell is signaled via cell-specific broadcast signaling. A 4-bit cell-specific parameter indicates 15 possible sets of subframes in which SRS may be transmitted within each radio frame. This configurability provides sufficient flexibility in adjusting the SRS overhead depending on the deployment scenario. The 16th configuration switches the SRS off in the cell, which may be appropriate for a cell serving primarily high-speed UEs. The SRS transmissions are always in the last SC-FDMA symbol in the configured subframes. The SRS and DM-RS are located in different SC-FDMA symbols. The PUSCH data transmission is not permitted on the SC-FDMA symbol designated for SRS, resulting in a worst-case sounding signal overhead of around 7% with the SRS symbol in every subframe.

The eNB may either request an individual SRS transmission from a UE, or configure a UE to transmit SRS periodically until it is no longer necessary. A single-bit UE-specific signaling parameter indicates whether the requested SRS transmission is periodic or aperiodic. If periodic SRS transmissions are configured for a UE, the periodicity may be 2, 5, 10, 20, 40, 80, 160, or 320 ms. The SRS periodicity and SRS subframe offset within the period in which the UE should transmit its SRS are configured via a 10-bit UE-specific dedicated signaling

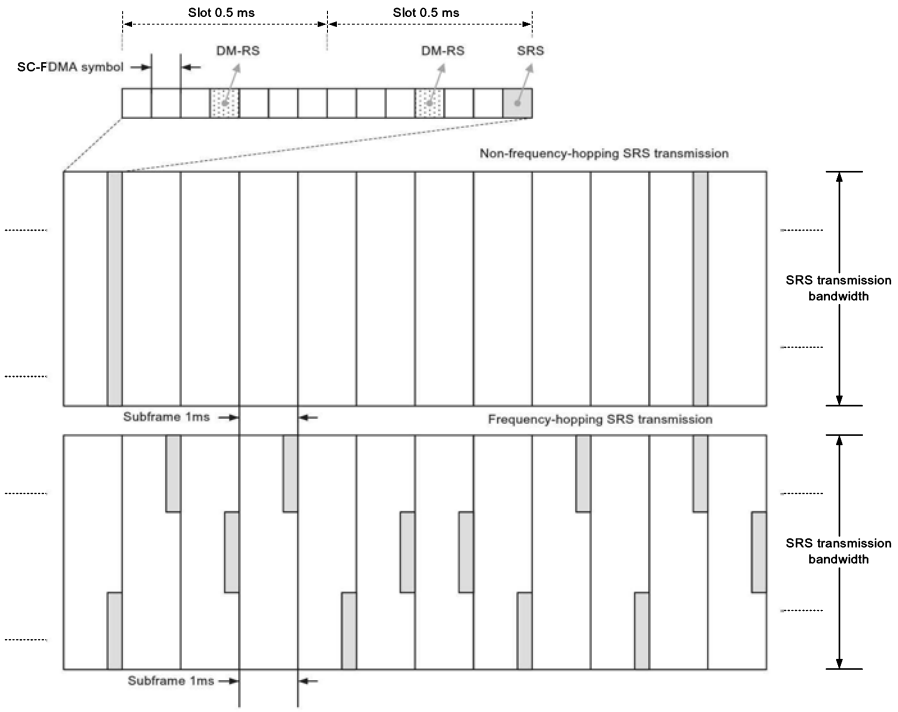


FIGURE 10.23

Non-frequency-hopping (wideband) versus frequency-hopping (sub-band) SRS [16].

parameter. There are differences between cell-specific SRS (signaled via cell-specific broadcast signaling) and UE-specific SRS (signaled via UE-specific dedicated signaling) configuration parameters, and their method of signaling to the UE noted in the forthcoming discussions.

The periodic SRS is based on the extended Zadoff–Chu sequence and is periodically transmitted in the last SC-FDMA symbol of an uplink subframe, as shown in Figure 10.23. As the example in Figure 10.23 shows, the SRS transmitted by the UEs are multiplexed in the time- and frequency-domain through configuring SRS periodicity T_{SRS}, frequency comb pattern, and SRS bandwidth. Similar to DM-RS, up to eight orthogonal SRS transmissions can be code-division multiplexed over a frequency region using different cyclic shifts of the root Zadoff–Chu sequence. The periodic nature of the LTE Rel-8 SRS transmission was not sufficiently flexible to support sounding from an increased number of UE antennas and/or an increased number of UEs, due to use of semi-static RRC signaling for UE configuration. As a result, aperiodic SRS transmission was introduced in LTE Rel-10 to complement periodic SRS transmission, where the eNB dynamically schedules a UE for SRS transmission on demand

(see Figure 10.22). Therefore, the SRS resources are not tied to a single UE until RRC (re)configuration by the eNB. This mechanism allows efficient management of a fixed set of time/frequency/code SRS resources for a larger number of UEs.

Periodic and aperiodic SRS transmission share a common set of cell-specific SRS resources (subframe configuration period, subframe offset, and maximum SRS bandwidth). Different sets of UE-specific sounding parameters are independently allocated for periodic and aperiodic SRS transmission including transmission bandwidth, periodicity, frequency comb pattern, and cyclic shift. The UE is configured with a fixed set of subframes in which it may be scheduled for aperiodic SRS transmission. This reduces the eNB scheduling complexity, because in any subframe only a subset of UEs are scheduled for aperiodic SRS transmission. An example of periodic and aperiodic SRS transmission timing is shown in Figure 10.22 for comparison, where periodic and aperiodic SRS have the same subframe offset and $T_{\mathrm{SRS_{Periodic}}} = 5$ ms. An aperiodic SRS transmission occurs in subframe $n + k$ corresponding to a trigger in subframe n, where $k \geq 4$, corresponding to a 4 ms delay constraint which is meant to maintain the minimum timing response similar to a normal downlink assignment or uplink grant.

To minimize the downlink signaling overhead, a semi-static configuration of aperiodic SRS parameters similar to that of periodic SRS was adopted in LTE Rel-8. Furthermore, the aperiodic SRS-related signaling bits are piggybacked in either a downlink assignment (scheduling downlink data transmission) or an uplink grant (scheduling uplink data transmission), in order to reduce signaling overhead. Since the aperiodic SRS is primarily used to support multi-antenna sounding, further configuration flexibility can be achieved by forming three independent sets of parameters that can be selected via a 2-bit SRS field in the uplink grant, scheduling uplink SU-MIMO data transmission. For an uplink single-antenna transmission, a single-bit aperiodic SRS field in the uplink grant is used [14,16].

There are only a few subframes in which the eNB can send an aperiodic SRS trigger, in order to satisfy the timing constraint ($k \geq 4$) for a specific aperiodic SRS transmission opportunity. The latter scheduling constraint has some impacts on some TDD frame configurations, where only a few uplink subframes are available in a radio frame. Therefore, to increase the scheduling opportunities for aperiodic SRS, LTE Rel-10 specifies a single-bit aperiodic SRS control field in some downlink assignments. Note that the uplink sounding can be used to support downlink beamforming by exploiting channel reciprocity in TDD systems.

To ensure that there is no timing mismatch between CSI estimates from multiple antennas, the UE is configured to simultaneously transmit SRS from all configured antenna ports. The orthogonality of the cyclic shifts falls apart as the frequency selectivity of the channel increases. This degradation is more pronounced as the number of antenna ports increases. Therefore, for both periodic and aperiodic SRS, a combination of the frequency comb (i.e., a multiplexing/interlacing pattern in frequency-domain) and cyclic shift can be used to multiplex the SRS sequences corresponding to four antenna ports in large delay-spread channels, while for low delay-spread channels or when the UE is equipped with

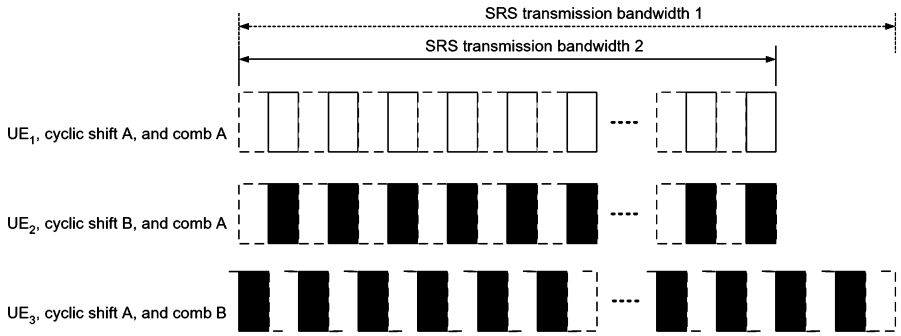


FIGURE 10.24

Multiplexing SRS transmission from different UEs [16].

two antenna ports, only different cyclic shifts are exploited to multiplex the SRS (as shown in Figure 10.24). The cyclic shifts and comb values for all antenna ports are implicitly derived from the values in each of the semi-statically configured parameter sets.

A UE transmits SRS in the serving-cell SRS resources using one of the following (UE-specific) trigger types [5]:

- Trigger type 0 via higher layer signaling.
- Trigger type 1 via DCI format 0/4/1A for FDD and TDD, and DCI format 2B/2C for TDD.

A UE may be configured with SRS parameters for trigger types 0 and 1 in each serving cell. The following SRS parameters are serving-cell specific and semi-statically configurable by higher layers for trigger type 0 and for trigger type 1 [5]:

- Transmission comb for trigger type 0 and each configuration of trigger type 1.
- Starting PRB assignment for trigger type 0 and each configuration of trigger type 1.
- Duration for trigger type 0, i.e., single or indefinite (until disabled).
- SRS configuration index I_{SRS}, SRS periodicity T_{SRS}, and SRS subframe offset $T_{\text{SRS-offset}}$ for trigger type 0 and trigger type 1.
- SRS bandwidth for trigger type 0 and each configuration of trigger type 1.
- Frequency hopping bandwidth for trigger type 0.
- Cyclic shift for trigger type 0 and each configuration of trigger type 1.
- Number of antenna ports for trigger type 0 and each configuration of trigger type 1.

A UE may be configured to transmit SRS on N_p antenna ports of a serving cell where N_p may be configured via higher layer signaling. For PUSCH transmission mode 1, $N_p \in \{0, 1, 2, 4\}$ and for PUSCH transmission mode 2, $N_p \in \{0, 1, 2\}$

when two antenna ports are configured for PUSCH and $N_p \in \{0, 1, 4\}$ when four antenna ports configured for PUSCH.

A UE may be configured for SRS transmission on multiple antenna ports of a serving cell within one SC-FDMA symbol of the same subframe. The SRS transmission bandwidth and starting PRB assignment are the same for all the configured antenna ports of a given serving cell [5]. The UE-specific parameters for SRS trigger type 0 in a serving cell, i.e., periodicity T_{SRS} and subframe offset $T_{\text{SRS-offset}}$ are defined in Table 10.6 for FDD and TDD modes. An example of SRS parameter configuration for an FDD system is shown in Figure 10.25 for better understanding of the effect of each SRS parameter on the SRS transmission. The periodicity of SRS transmission, in number of subframes, is serving-cell specific and is selected from the set $T_{\text{SRS}} \in \{2, 5, 10, 20, 40, 80, 160, 320\}$. For the SRS periodicity value of $T_{\text{SRS}} = 2$ ms in TDD, two SRS resources are configured in half-frame containing uplink subframes. The UE-specific parameters for SRS trigger type 1 in a serving cell, i.e., periodicity T_{SRS} and subframe offset $T_{\text{SRS-offset}}$, are given in Table 10.6 for FDD and TDD modes. The periodicity of the SRS transmission, in number of subframes, is serving-cell specific and is selected from the set $T_{\text{SRS}} \in \{2, 5, 10\}$. For the SRS periodicity value of $T_{\text{SRS}} = 2$ ms in TDD, two SRS resources are configured in a half-frame containing UL subframes in a serving cell [5].

The cell-specific SRS subframe configuration period T_{SFC} and the cell-specific subframe offset Δ_{SFC} for the transmission of SRS are listed in Table 10.7, for frame structure types 1 and 2, where the SRS subframe configuration parameter is provided by the higher layers. The sounding reference signal is transmitted in subframes whose numbers satisfy $\lfloor n_s/2 \rfloor \bmod T_{\text{SFC}} \in \{\Delta_{\text{SFC}}\}$. For frame structure type 2, SRS is transmitted only in configured uplink subframes or UpPTS portion of the special subframe. Note that the cell-specific SRS parameters $(T_{\text{SFC}}, \Delta_{\text{SFC}})$ and UE-specific SRS parameters $(T_{\text{SRS}}, T_{\text{SRS-offset}})$ are not identical.

The SRS sequence is defined by $r_{\text{SRS}}(n, \bar{p}) = r_{\text{uv}}(n, \tau_{\bar{p}})$, where $r_{\text{uv}}(n, \tau_{\bar{p}})$ is a cyclically shifted Zadoff–Chu sequence defined in the previous section, u is the PUCCH sequence-group number, and v is the base sequence number. The antenna-dependent cyclic shift $\tau_{\bar{p}}$ of the SRS is given as $\tau_{\bar{p}} = \frac{\pi}{4}(n_{\text{SRS}} + \frac{8\bar{p}}{N_p}) \bmod 8|_{\bar{p} \in \{0, 1, \ldots, N_p - 1\}}$, where $n_{\text{SRS}} = \{0, 1, 2, 3, 4, 5, 6, 7\}$, is configured separately for periodic and each configuration of aperiodic sounding by the higher-layer signaling for each UE, $\bar{p}$ is sequential numbering of the antenna ports, and N_p is the number of antenna ports used for SRS transmission.

The SRS sequence is scaled with scaling factor β_{SRS} in order to conform to the transmit power P_{SRS} and is then mapped sequentially starting with $r_{\text{SRS}}(0, \bar{p})$ to resource elements (k, l) on antenna port $\bar{p}$ such that $a_{[2k'+k_0(\bar{p})]l} = \frac{1}{\sqrt{N_p}} \beta_{\text{SRS}} r_{\text{SRS}}(k', \bar{p})|_{k'=0,1,\ldots,N_{\text{SRS}}-1}$, where N_p is the number of antenna ports used for SRS transmission and the relation between the index $\bar{p}$ and the antenna port p is described in 3GPP TS 36.211 [3]. The set of antenna ports used for SRS transmission is configured independently for periodic and each configuration of aperiodic sounding. The quantity $k_0(p)$ is the frequency-domain starting position