ADVANCED MOBILE COMMUNICATION

Rony Kumer SAHA
KDDI Research

ADVANCED
MOBILE COMMUNICATION

Technical Note

RONY. K. SAHA

R&D
NextEVOlution
Branding Intellects

ACKNOWLEDGEMENTS

Information incorporated in this lecture note may be found fully or partly with or without any change in figures, tables, sentences, etc. in the **references**. The lecture note is developed **only for the purpose of class room use**. The lecture note developer and the reviewer fully acknowledge and grateful to these helpful contributions (mentioned references as well as the references there-in) to help make this lecture note developed.

Lecture Note Developer

R. K. SAHA
Master of Engineering (Info. & Comm. Tech.)
Asian Institute of Technology, Bangkok, Thailand

PREFACE

The field of Cellular Mobile Communications (CMC) is highly emerging. CMC Technologies have been in the move towards achieving ever increasing user needs for rich multimedia services. In the past less than three decades, enormous changes in CMC have been addressed in technology, standard and system level. This student handout is developed to get the students a fast but with considerable explanation of the evolution of CMC technologies, standards and systems. A good trade-off between depth-and-breath of any context is considered. The handout can be used for teaching and research both senior undergraduate students and graduate students.

In the first part (please refer to Cellular Mobile Communication – A Fundamental Perspective), fundamentals of CMC including channel propagation characteristics and modeling, multi-path fading, channel classifications and parameters, cellular concept, cellular system design fundamentals are addressed. Standards mainly Global System for Mobile Communications (GSM) and GSM Evolution and Code Division Multiple Access (CDMA) are explained in detail.

In this part, we start with a general understanding of high data rate communications and the means to achieve so. Advanced enabling technologies such as Multiple-input Multiple-output (MIMO), orthogonal Frequency Division Multiplexing (OFDM), Resource scheduling, Link adaptation, Hybrid Automatic Repeat Request (HARQ) are discussed. Third generation (3G) standards and beyond 3G are addressed. Standards including Wideband CDMA (WCDMA), High Speed Packet Access (HSPA), Long Term Evolution (LTE), and LTE-Advanced (LTE-Advanced) are explained in detail. Changes in architectures from one generation to another are explained, and their requirements for changes are detailed. Students will get a seamless experience throughout the handout as they read up.

There are a number of problems that are designed and collected to understand the topics discussed in the chapters. Many of these problems are solved, and others are left for the students to exercise. The purpose here is not to get the exercise solved, rather is to gain understanding how to solve. Students are therefore advised to discuss with others about the outcome of each exercise.

We consider avoiding mathematical explanation as much as possible so that students do not need much mathematical background as prerequisite on mastering the matter. However, a good understanding of linear algebra, Fourier analysis, signals and systems, digital signal processing, and fundamentals of telecommunications is highly desirable. Many of the topics are taken from industry standardization body's whitepapers, IEEE publications, and renowned authors in the field. Please refer to the Reference section for further information.

Finally, we believe that this handout will serve for understanding the fundamentals of CMC technologies, standards and systems to the reader. Should there be any further information, inquiry, or suggestion, please reach us with the followings.

Regards,

R. K. SAHA

A Contribution of R&D NEXTEVOLUTION

CONTENTS

CHAPTER 11

HIGH DATA RATES IN MOBILE COMMUNICATION

One of the most critical demands in mobile communication is high data rates. Fundamentals constraints that limit the achievable high data rates in different scenarios are discussed. Different approaches are addressed in order to increase the data rates.

11.1FUNDAMENTAL CONSTRAINTS IN ACHIEVING HIGH DATA RATES

According to Shannon, the channel capacity can be given as the following.

$$C = BW \,.\log_2\left(1+\frac{S}{N}\right)$$

Factors that limit high data rates are

- channel bandwidth
- signal-to-noise ratio /received signal power

note that the above expression assumes only AWGN, though the real case would be rather complicated such as interference, etc.

Define R is the information rate (bits/s)
S is the received signal power (Watt)
N is the noise power (Watt)
E_b is the received per information bit energy (Watt/s)
N_0 is the noise power spectral density

$$S = E_b.R$$
$$N = N_0.BW$$

Since information rate is upper limited by the channel capacity

$$R \le C = BW \,.\log_2\left(1+\frac{S}{N}\right) = BW \,.\log_2\left(\frac{E_b.R}{N_0.BW}\right)$$

If γ is the link bandwidth utilization $=\dfrac{R}{BW}$, the minimum required per bit energy

$$\frac{E_b}{N_0} \ge \min\left\{\frac{E_b}{N_0}\right\} = \frac{2^{\gamma}-1}{\gamma}$$

The minimum required received per bit energy (normalized to N_0) as a function of bandwidth utilization can be plotted and shown in Figure 11.1.

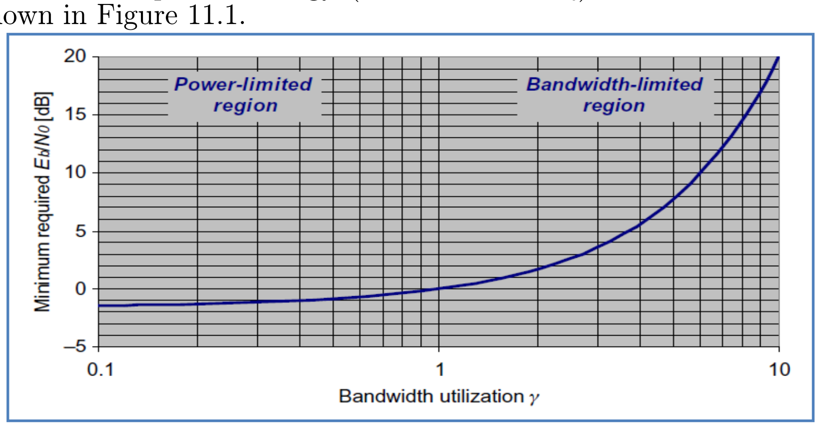


Figure 11.1: minimum required $\frac{E_b}{N_0}$ as a function of bandwidth utilization.

Hence, for low bandwidth utilization, the channel bandwidth has insignificant impact on increasing received signal strength (assuming receiver noise is flat over the channel bandwidth). The only alternative is the increase in power to increase the received signal strength and so does the name power-limited region. Conversely, any rate greater than the channel bandwidth, needs an enormous increase in received signal power, however a proportionate increase in channel bandwidth. This is why, this region is called bandwidth-limited region. If the channel bandwidth is kept no change, the only alternative to increase in data rate is increase received signal power proportionately. And any data rate (in theory) can be possible to achieve.

In power-limited region, an approximate proportional increase in received signal power is required to increase any amount of data rate. However, in high bandwidth utilization region, a very high increase in received power or a proportional increase in channel bandwidth is required to increase any amount of data rate. In general, the efficient use of received transmit power can be achieved when the channel bandwidth is equal to the data rate.

Noise-limited scenarios

Consider Noise-limited scenarios, where a channel is assumed to be impaired by noise only. In this scenario, the following approaches can be used for providing higher data rate.

Approach 1. if transmit power is kept constant, the received power can be increased by reducing the TX and Rx distance, i.e. the cell coverage. The other way this can be stated that the data rate at the Rx can be increased by reducing the cell coverage, since this provides better signal strength at the Rx, with no change in channel BW.

Approach 2. multiple antennas at the Rx with appropriate combining of the received signals or multiple antennas at the Tx with proper beam forming directing towards the Rx, it can be possible to increase the received signal strength, consequently the higher data rate. This either approach is useful for power limited scenarios rather since both techniques provide saturation in performance after certain extent.

Approach 3. Using multiple antennas both at the transmitter and at the receiver the problem with Approach 2 can be overcome. This antenna arrangement in literature is termed as Multiple-Input Multiple-Output, shortly MIMO technique. Using different MIMO techniques such as spatial multiplexing the data rate can be provided up to min–N_T, N_R″ where N_T and N_R are respectively number of transmit and receive antennas.

Interference-limited scenarios

Normally in addition to the inherent noise, interference from other signal source, Tx or Rx are also the major impairment to signal over communication radio links. The interference in general is of two types such as inter-cell interference and intra-cell interference. The former is caused between Tx and Rx in adjacent cells, whereas the later caused in the same cell. Note that the interferer could be BS or MS, depending on the scenario. Reducing the cell size and using MIMO technique is also applicable for interference –limited scenario to improve the signal strength at the RX. To be specific,

- Reducing cell size implies reducing distance coverage, consequently the number of users covered by the cell. And, hence less intra-cell interference from the neighbors.
- Appropriate MIMO technique increases SINR at the Rx and hence higher data rates.
- Beam forming using transmit antenna arrangements can increase the signal strength towards the Rx direction.
- Interference can be suppressed by proper arrangements of antennas at the receiver so that the interfering signal direction is misaligned with the desired received signal direction.

11.2 HIGHER-ORDER MODULATION

Higher data rate can also be provided with increasing the modulation order. For example, if we increase the modulation order from QAM to 16 QAM to 64 QAM, this means the number of information bits per symbol increases correspondingly from 2 to 4 to 6. Hence, in comparison to simple QAM 16 QAM and 64 QAM give respectively 2-times and 3-times data rates. Figure 1.2 shows typical QAM, 16-QAM and 64-QAM schemes.

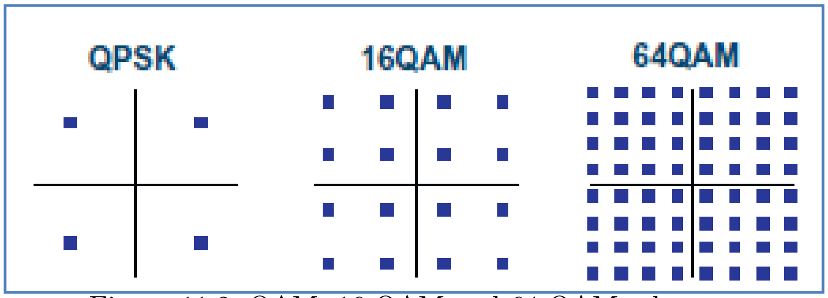


Figure 11.2: QAM, 16-QAM and 64-QAM schemes.

However, increase in Figure 11.3), modulation order increase bit error probability at the receiving end (see hence higher E_b /N_0, consequently more received power. Hence higher order modulation is good for small cell cells such as picocell.

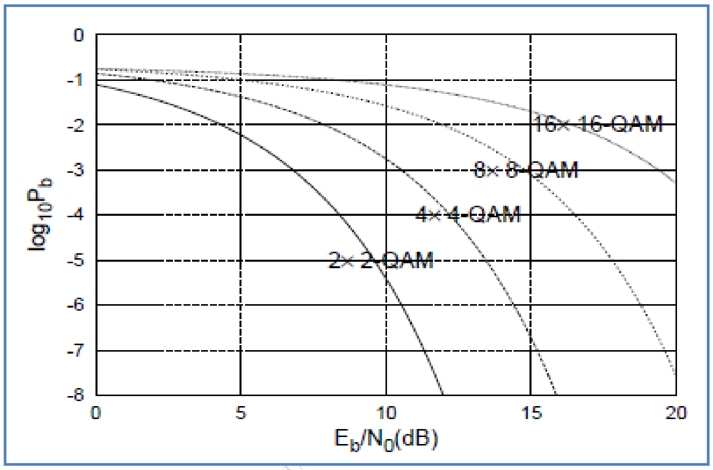


Figure 11.3: Bit error probability for different order QAM signal sets [poompats, 2011].

11.3 WIDER BANDWIDTH

It has already pointed out that increasing the signal bandwidth results in increasing the data rate. However, this approach has a number of drawbacks as follows.

- bandwidth is scarce and expensive.
- increases the complexity in equipment design. For example, increasing BW means increase in sampling rate for D/A converter at the Rx side as well as more power consumption of converters.
- multi-path fading impact is much more prominent and hence increase in received bit errors (see Figure 1.3).

This is because, wider signal such as shown in Figure 11.4 for example experiences more channel frequency selectivity property, i.e. different part of the signal attenuated differently over the channel BW. This drawback can be overcome using multi-carrier approach, i.e. splitting the wider BW into number of narrower segments and transmitting information using these narrow band carriers. One such example technique is orthogonal frequency division multiplexing (OFDM) technique as shown in Figure11.5.

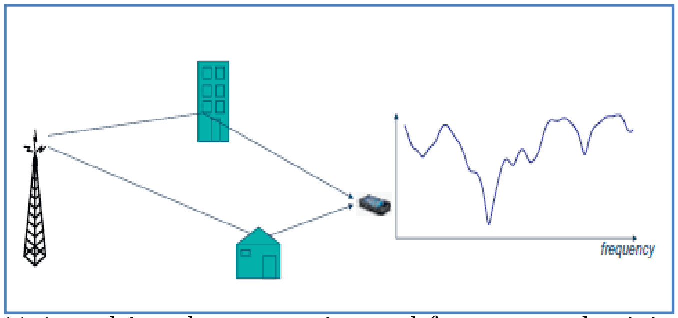


Figure 11.4: multi-path propagation and frequency selectivity effect.

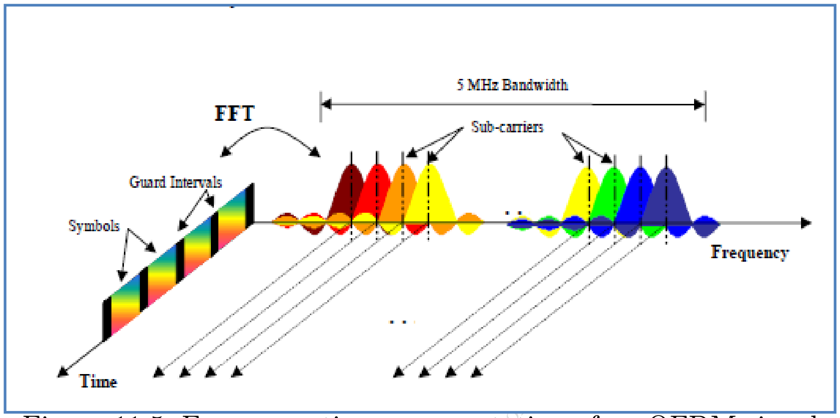


Figure 11.5: Frequency-time representation of an OFDM signal [Rohde and Schwarz, 2007].

The second approach is multiplexing a number of existing narrowband to extend it to a wider one as shown in Figure 11.6. All the narrowband signals are transmitted jointly after multiplexing over the same radio link. However, this technique suffers from inefficient spectrum multiplexing (since they normally do not allow for tightly packing) as well as transmitter power efficiency because of large variations in transmit power. Conversely, multi-carrier provides a option for evolution from the legacy system so that any legacy terminals can receive only the original narrowband and multi-carrier enabled terminal can use the full wider band.

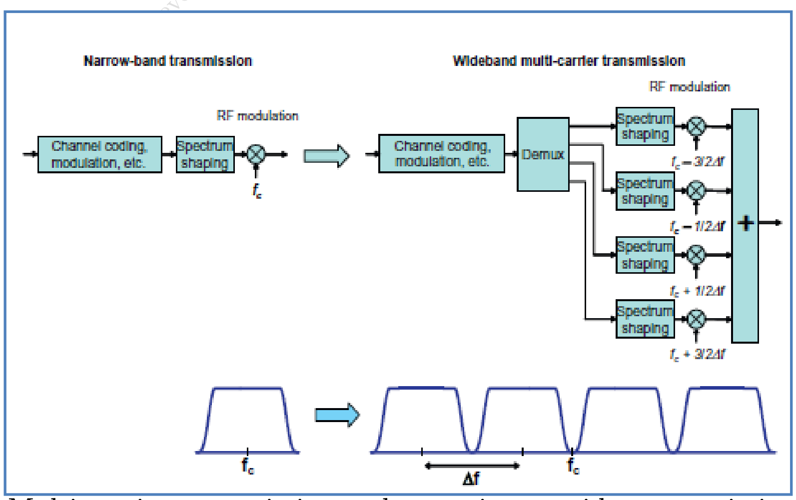


Figure 11.6: Multi-carrier transmission and extension to wider transmission bandwidth.

CHAPTER 12
MULTIPLE-INPUT MULTIPLE-OUTPUT TECHNIQUES

MIMO is a set of techniques with the common theme; rely on the use of multiple antennas at the receiver and / or the transmitter, in combination with advanced signal processing. MIMO can be used for various system improvement purposes such as improved

- system capacity (more users per cell),
- coverage (possibility for larger cells),
- service provisioning (for example, higher per-user data rates), etc.

We shall address in this section an overview of different kinds of multiple antenna techniques and advanced communication techniques in next generation mobile communication systems such as 3GPP LTE-Advanced.

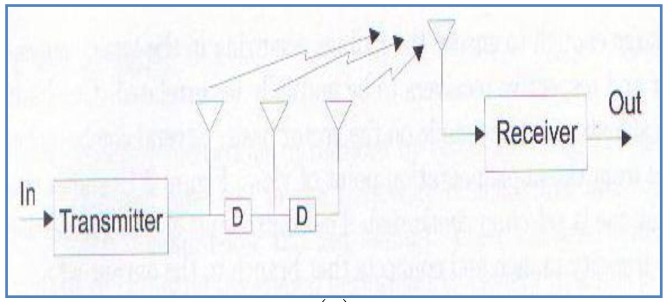
Practical Considerations

- in-between antennas distance (measures the degree of fading effect at the antenna terminals): in general longer distance relatively provides low mutual fading correlation between antennas in the antenna array.
- fading correlation (mutual correlation among the signals experiences at different antenna terminals): low or high depends on what is to be achieved by the array, for example diversity, beam-forming, or spatial multiplexing.
- radio carrier frequency or wavelength (defines antenna separation or distance in the array): typically an order of tenfold wavelength is needed for macro-cell antenna separation. And half fold wavelength for mobile station, irrespective of environments. However, for cells such as picocells, femtocells (which are alike mobile station's height), a small antenna separation distance is sufficient to provide low mutual fading correlation.
- environment (measures the effect of objects on signal distortion or fading): signal is being obstructed more in urban than in rural.
- antenna height (measures the multi-path effect on signals) : lower height antennas, for example mobile stations, experience significant multi-path reflection that causes fading since multipath effect mainly occurs in the near field zone (around mobile stations). The opposite is the case for base stations.
- angle of arrivals (defines the angles at which signals arrive at the stations): for Base stations, relatively narrower signal reception angles at different antennas results in large antenna separation distance. However, wider angles at the mobile stations results in lower antenna separation enough to provide low fading correlation.
- number of antennas in the array (measures the degree of improvement in signal reception): at least in theory, data rate increases with min(N_t , N_r) where N_t and N_r are respectively transmit antenna number and receive antenna number.

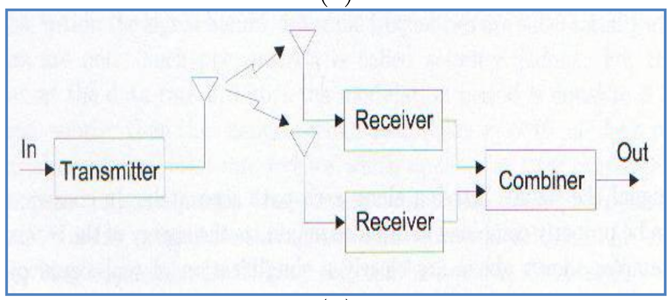
12.2 MIMO TECHNIQUES

- Diversity such as spatial diversity to reduce fading effect.
- Beamforming to provide increase in gain to the desired signal or to mitigate interference signal effect in the respective the directions.
- Spatial multiplexing to create parallel communication channels in space that can provide high data rate with a limited bandwidth.

The idea of diversity is to provide to receiver two or more such inputs that their fast fading characteristics are uncorrelated. By combining these inputs we can smooth out the fast fading and reduce the variations in the received signal to the level of slow fading due to the shadowing on the path from BS to MS [5]. The below Figure 12.1 illustrates a typical space diversity (receiver) technique.



(a)

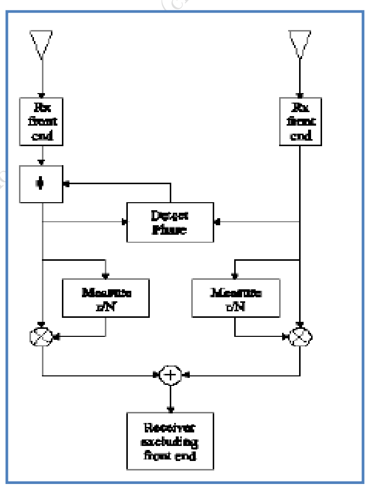


(b)

Figure 12.1: (a) transmitter and (b) receiver (space) diversity [5].

Combining Techniques

In maximal-ratio combining the signal of each branch is weighted in proportion to its own signal envelope-to-noise power ratio before summation. In selective combining technique, by selecting the strongest baseband signal among M diversity branches, we can reduce fast fading considerably Figures 12.2 and 12:3 represent respectively maximal-ration combining and selective combining techniques.



Figures 12.2: maximal –ratio combining.

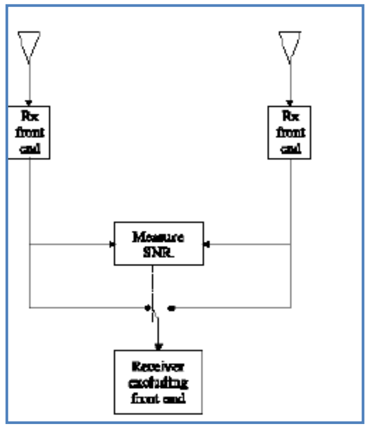


Figure 12.3: Selective combining.

Multiple Receive Antennas: antenna combining techniques

Multiple antennas are used generally used to provide receive diversity against channel fading effect at receiver antennas. Shown in Figure 12.4 is a typical linear combination of received signals at different receive (RX) antennas where $r_1, r_2, ..., r_{N_R}$ are the received signals at 1, 2, ..., N_R antennas $w_1^*, w_2^*, ..., w_{N_R}^*$ are the corresponding complex weight vectors to the channel responses $h_1, h_2, ..., h_{N_R}$. The received signal is multiplied by the weight vectors before being linearly combined. The linear receive antenna combination is expressed as follows.

$$\gamma \le \log_2\left(1 + \gamma \frac{E_b}{N_0}\right)$$

If the channel as assumed non-frequency selective, i.e. impaired by only white noises, the linear receive antenna combination can be shown as in Figure 12.4. The maximum E_b/N_0 can be achieved if the following condition is fulfilled.

$$\overline{w} = \overline{h}\ .$$

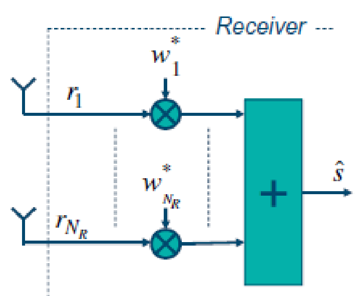


Figure 12.4: Linear receive antenna combination.

188

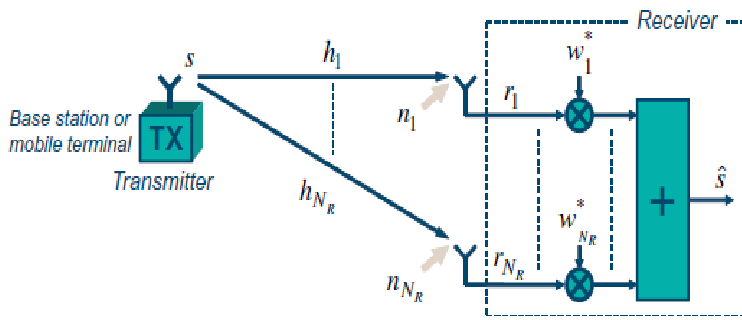


Figure 12.5: Linear receive antenna combination considering channel gain.

Equation 12.2 is called Maximum-Ratio Combination (MRC), which ensures signals are phase aligned and weights signals proportionate to their respective channel gains. good strategy when channel is impaired by noise only or when the number of users in the cell is large (so that the combined interference impact seems noise alike). However, for a single dominating interferer, Rather than MRC, it is good to interference rejection strategy. Interference rejection combination (IRC)is such a type, which uses forming beams that provides attenuation to the direction of interference signal propagation, consequently, the suppression of that interferer (Figures 12.6 and 12.7).

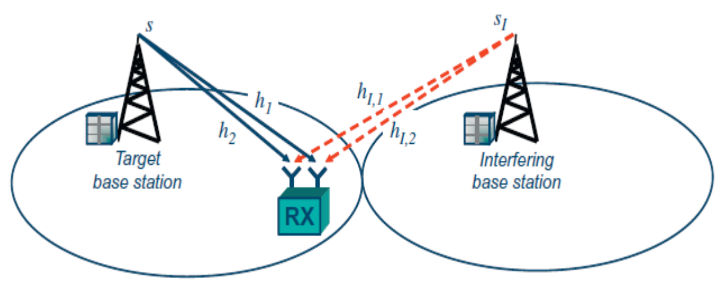


Figure 12.5: Downlink with a single dominating interferer.

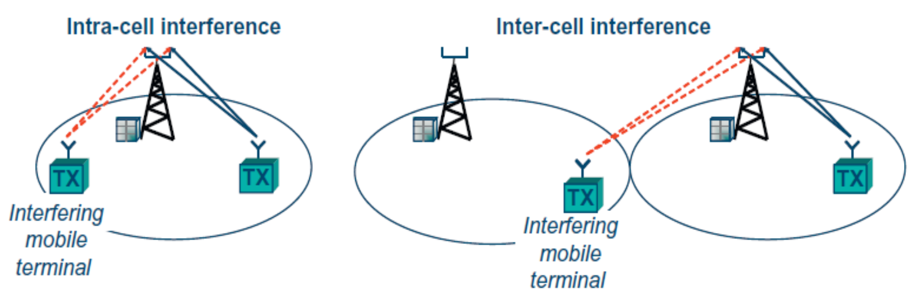


Figure 12.7: Receiver with one strong interfering terminal
(a) Intra-cell interference (b) Inter-cell interference.

For single interferer case, the received signal can be expressed as follows.

$$\bar{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_{N_R} \end{pmatrix} = \begin{pmatrix} h_1 \\ \vdots \\ h_{N_R} \end{pmatrix} \cdot s + \begin{pmatrix} h_{I,1} \\ \vdots \\ h_{I,N_R} \end{pmatrix} \cdot s_I + \begin{pmatrix} n_1 \\ \vdots \\ n_{N_R} \end{pmatrix} = \bar{h} \cdot s + \bar{h}_I \cdot s_I + \bar{n}. \quad (12.3)$$

the interference signals can be fully suppressed up to N_R -1 if

$$\overline{w}^H.\overline{h} = 0 \tag{12.4}$$

since large number of interference suppression causes increase in noise level, it may, however, good to choice weight vectors that minimizes the following expression called as Minimum Mean Square Error (MMSE) combination.

$$\varepsilon = E\left\{\left|\hat{s} - s\right|^2\right\} \tag{12.5}$$

Suppression of uplink intra-cell interference employing IRC is referred to as Spatial-Division Multiple Access (SDMA).If the channel is frequency selective (time dispersive), generally two dimensional space-time linear processing as given in Figure 12.8 is employed. Spatial-domain processing addresses the multi-antenna benefits and time domain processing addresses the frequency selective property of the channel. However for cyclic prefix insertion, space-frequency domain processing is employed (Figure 12.9), where the weights are represented in frequency domain and jointly combined to address interference, noise and signal corruption because of channel selective property. Note that for OFDM, weights can be selected taking only interference and noise only in account because of its immunity to channel frequency selectivity. Note that we consider the discussion per sub-carrier basis.

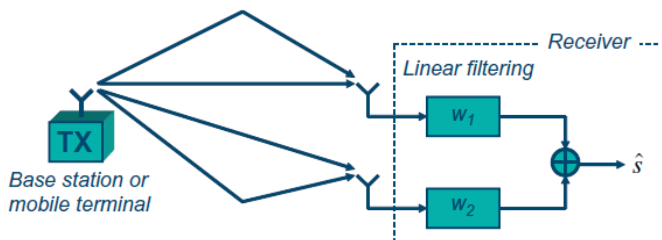


Figure 12.8: Two-dimensional space-time linear processing.

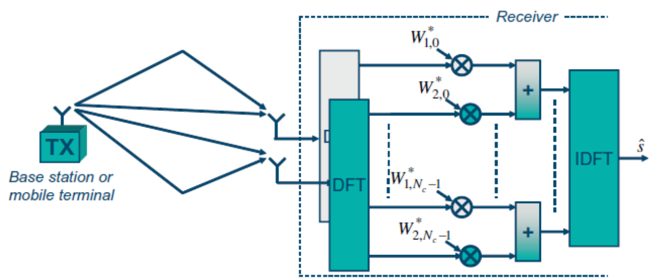


Figure 12.9: Two-dimensional space-frequency linear processing.

Multiple Transmit Antennas

Multiple transmit antennas are normally used for diversity and beamforming. The mutual fading correlation between antennas should be low and consequently, large separation distance between antennas. Because of why multiple transmit antennas are applied at the base stations and hence in the downlink.If channel knowledge is unknown, beamforming cannot be provided. However, diversity of several types can still be possible. Follows are few approaches to address diversity using multiple antennas at the transmitter.

Delay Diversity

A frequency selective channel shows time dispersion (different delays along signal propagations towards the receiver) or in other word multipath fading effect (where different signal frequency component experiences different delays and attenuations). This property of the channel to frequency is the subject that matters for diversity, i.e. multiple channels realization. If, however, the channel is not frequency selective, an artificial frequency selective channel can be realized transmitting signals with different delays from different transmit antenna. Figure 12.10 shows a typical two-antenna delay diversity technique.

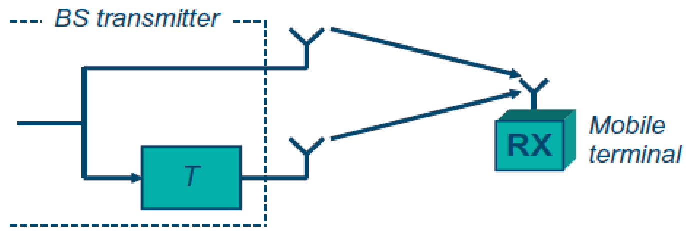


Figure 12.10: Two-antenna delay diversity.

Cyclic-Delay Diversity

Similar to delay diversity except it uses transmission in block wise and applies cyclic shifts (Figure 12.11). Applicable for block based transmission schemes such as OFDM where a cyclic shift of the time-domain signal that corresponds to a frequency-dependent phase shift is performed before modulation.

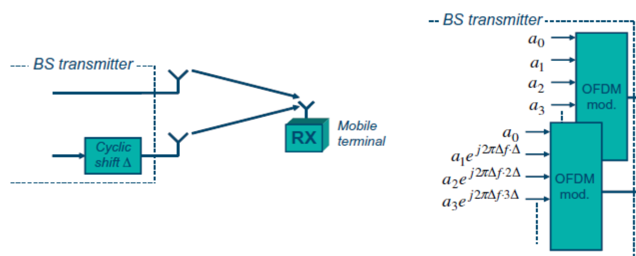


Figure 12.11: Two-antenna Cyclic-Delay Diversity (CDD).

Space-Time Transmit Diversity

Diversity is provided by mapping modulation symbols both in time and space domain. Figure 12.12 shows a typical two-antenna space time transmit diversity scheme that operates on pairs of modulation symbols. Symbols are transmitted directly to the first antenna, however a complex-conjugated and sign-reversed symbols are transmitted to the second antenna

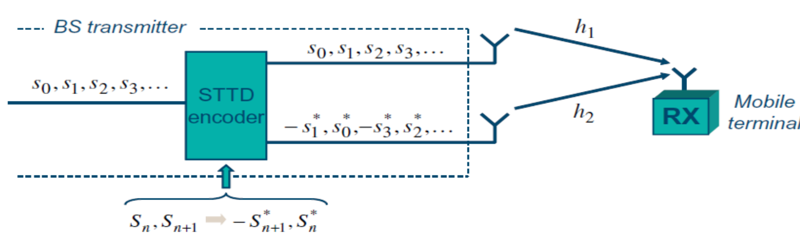


Figure 12.12: Space–Time Transmit Diversity (STTD).

The transmission scheme can be expressed as follows.

$$\overline{r} = \begin{pmatrix} r_{2n} \\ r^{*}_{2n+1} \end{pmatrix} = \begin{pmatrix} h_1 & -h_2 \\ h^{*}_2 & h^{*}_1 \end{pmatrix} \cdot \begin{pmatrix} s_{2n} \\ s^{*}_{2n+1} \end{pmatrix} = H \cdot \overline{s} \qquad (12.6)$$

where r_{2n} and r_{2n+1} are the received symbols intervals 2n and 2n+1 respectively. The transmitted signals can be recovered using

$$W = H^{-1}$$

to the received vector given in Equation 12.6 .

For two-antenna scheme, the coding rate as well as bandwidth utilization can be achieved to unity. However, complex-valued modulation such as QPSK, 16 or 64 QAM with coding rate unity and without inter-symbol interference can be achieved only for two-antenna case. For more antennas than two, costs coding rate as well as bandwidth utilization less than unity, without inter-symbol interference.

Space-Frequency Transmit Diversity

Similar to STTD, except encoding is performed in space-frequency domain (Figure 12.13). Applicable to frequency domain transmission schemes such as OFDM where inputs to the first antenna is carried out directly using block of symbols a_0 , a_1, ... and their complex-conjugated as well as sign-reversed forms are input to the second antenna. Similar to STTD, a straightforward extension more than two-antenna is not possible without reduction in coding rate.

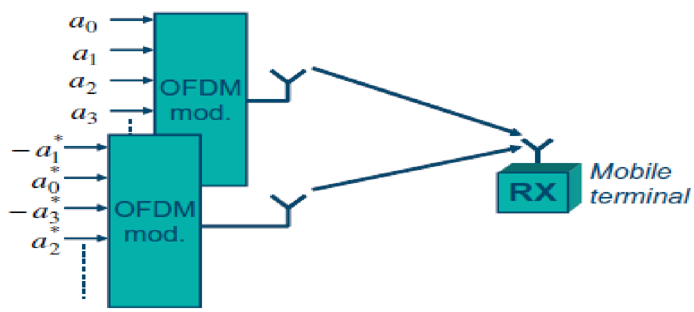


Figure 12.13: Space-Frequency Transmit Diversity for two transmit antennas.

Transmitter-Side Beam-Forming

Beamforming is the technique to provide the overall beam in the direction in interest, for example, desired receiver. It needs the channel knowledge so that the phase and the gain (if frequency-selective channel is considered) of the beam can be adjusted according to the requirements (target MS movement, channel variation, etc.). If the channel is frequency non-selective, then only change in phases of signals at different antennas is needed assuming constant gain all signals. With proper beamforming, received signal strength of up to N_T can be achieved in the direction of interest. Fading correlation among antennas play an important role in beamforming.

A high correlation means small separation distance between antennas and the channels experienced at all antennas in the TX side and RX side is almost the same. Hence, varying the phase shifts in transmitted signal at different transmit antennas, the direction of the beam can be varied (Figure 12.14). This kind of highly correlated beam forming is referred to as classical beamforming. A low correlation among antennas means large separation in distance among antennas and the channels experienced at individual antenna is almost independent to another (Figure 12.15). Weight vectors are complex valued (changes in both amplitude and phase) in contrast to classical beamforming (only change in phase). Complex weights (or can be termed as precoding vectors denoted as) are multiplied to the transmitted signals as given in the following Equation 12.7.

$$\bar{s} = \begin{pmatrix} s_1 \\ \vdots \\ s_{N_T} \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_{N_T} \end{pmatrix} \cdot \boldsymbol{s} = \bar{v} \cdot s. \tag{12.7}$$

Precoding vector is the complex conjugate of the corresponding channel coefficient h_i and with a normalization (to ensure a fixed overall transmit power). Precoding vector should be selected according to the Equation 12.8 to maximize the received signal strength.

$$v_i = \frac{h_i^*}{\sqrt{\sum_{k=1}^{N_T} |h_k|^2}}. \tag{12.8}$$

Precoding ensures the followings.

- received signals are phase aligned
- adjust power to different antennas according to instantaneous channel conditions (High h_i allocated by Tx high power)
- a overall constant transmit power.

low correlation requires more channel knowledge and fast change in precoding vectors to track the instantaneous channel variations. low correlation can provide diversity as well in contrast to classical beamforming. Channel can be estimated by the MS and feedback to the BS for FDD (frequency division duplex), if uplink and downlink operates on different bands. Or the MS can select a precoding vector from the precoder codebook and feedback to the BS.

For TDD, since both uplink and downlink operates on the same frequency, the BS can estimate the downlink channel from the uplink. for OFDM, precoding vector should selected from Equation 12.8.

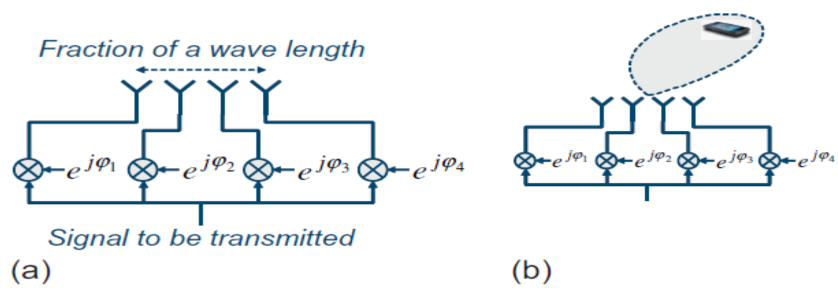


Figure 12.14: Classical beam-forming with high mutual antennas correlation: (a) antenna configuration; (b) beamstructure.

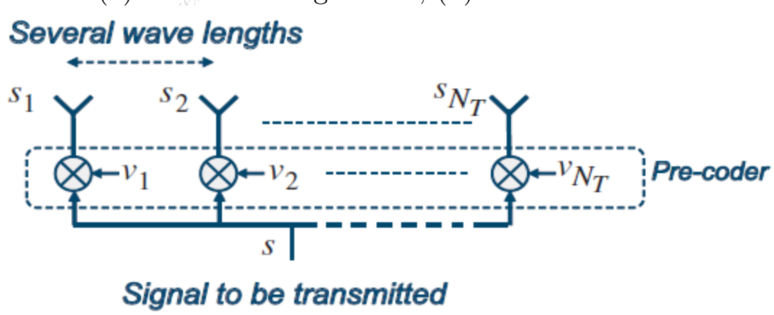
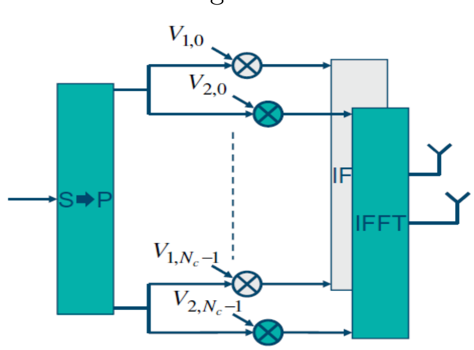


Figure 12.15: Pre-coder-based beam-forming in the case of low mutual antenna correlation.



193

Figure 12.16: Per-subcarrier precoding in the case of OFDM (two transmit antennas).

Spatial Multiplexing

Spatial multiplexing is the technique that provides creating multiple channels in space, the number of what depends on (at least in theory) the minimum value of the number of antennas at the transmitter and receiver.

- Uses multiple antennas both at the transmitter and receiver, because of why termed as Multiple-Input Multiple-Output (MIMO)
- Provides high data rates with improving received SNR.

In spatial multiplexing, the received SNR is N_L times lower since the power is split into N_L channels in the space. Hence for N_L parallel channels, the channel capacity per unit bandwidth can be expressed as follows.

$$\frac{C}{BW} = N_{\mathrm{L}} \cdot \log_2\left(1 + \frac{N_{\mathrm{R}}}{N_{\mathrm{L}}} \cdot \frac{S}{N}\right)$$
$$= \min\{N_{\mathrm{T}}, N_{\mathrm{R}}\} \cdot \log_2\left(1 + \frac{N_{\mathrm{R}}}{\min\{N_{\mathrm{T}}, N_{\mathrm{R}}\}} \cdot \frac{S}{N}\right) \tag{12.9}$$

Since for low values of x, $\log_2(x)$ is approximately equals to x, under certain conditions, the capacity can be linearly increased with the increase of the number of antennas. And, hence, the term spatial multiplexing. Figure 12.17 shows a 2 X 2 antenna configuration. Consider a frequency non-selective channel (i.e. no time dispersion).

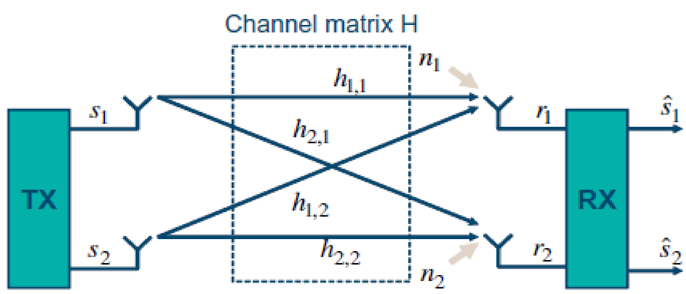


Figure 12.17: 2 X 2 antenna configuration.

The received signal can be expressed as follows.

$$\overline{r} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{pmatrix} \cdot \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} = \boldsymbol{H} \cdot \overline{s} + \overline{n} \tag{12.10}$$

If we consider no noise and no residual interference, the transmitted signals (Figure 12.18) can be recovered applying the following operation on the received signal expressed in Equation 12.11.

$$\begin{pmatrix} \hat{s}_1 \\ \hat{s}_2 \end{pmatrix} = \boldsymbol{W} \cdot \overline{r} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} + \boldsymbol{H}^{-1} \cdot \overline{n}. \tag{12.11}$$

where

$$W = H^{-1}$$

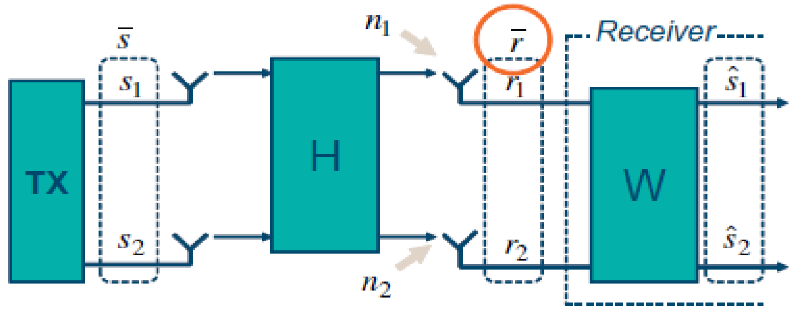


Figure 12.18: Linear reception/demodulation of spatially multiplexed signals.

if the receiver processing is not linear (joint demodulation), optimal receiver approach such as Maximum-Likelihood (ML) detection can be used. However, because of it complexity, other nonlinear approach to demodulation such as Successive Interference Cancellation (SIC) can be used. SIC uses the assumption that spatially multiplexed signals are coded separately for being spatially multiplexed and is referred to as Multi-Codeword transmission (Figure 12.19 (b)) in contrast to Single-Codeword transmission (Figure 12.19 (a)): multiplexed signals are jointly coded data generated even from the same source are first demultiplexed into different signals before coding that are to be spatially multiplexed for multi-codeword transmission.

The receiver first demodulates followed by decoding any one of the multiplexed signals and after proper decoding re-encodes the same decoded signal , which is then subtracted the received spatially multiplexed signal to avoid interference with the next to decode signal. The second signal is then taken its turn and follows the same process as described for the first signal.

This iterative process continues until the last signal in the multiplexed signal(Figure 12.20).This turns out that the first decoded signal is more susceptible to interference than later decoded signals. To address this issue, different modulation schemes and coding rates should be applied to different signals, typically, lower order modulation as well as lower coding rate for the first signal. This is referred to as Per-Antenna Rate Control (PARC).

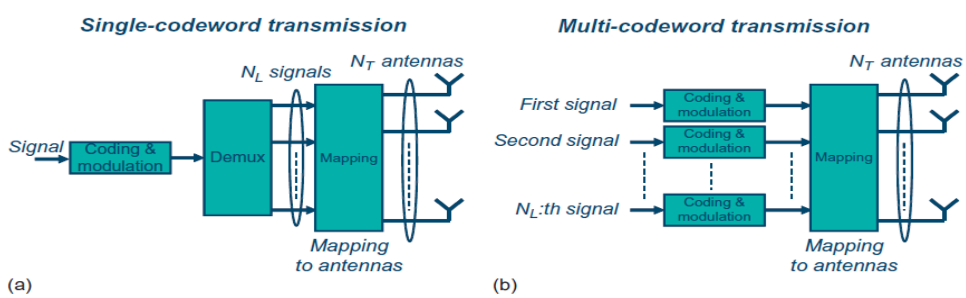


Figure 12.19: Single-codeword transmission (a) vs. multi-codeword transmission (b)

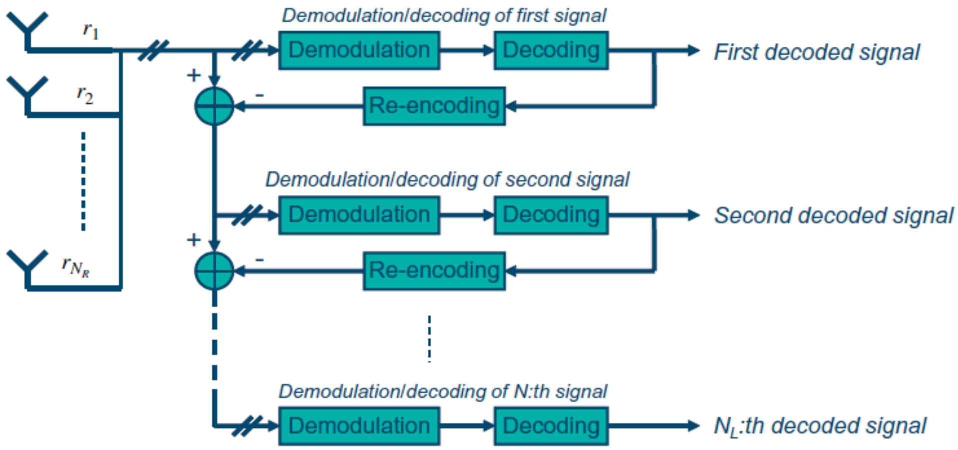


Figure 12.20: Demodulation/decoding of spatially multiplexed signals based on Successive Interference Cancellation.

12.3 ADAPTIVE MIMO SYSTEMS

Adaptive MIMO transmission is in general a technique that provides switching in time, frequency as well as spatial signalling techniques in response to changing conditions of the propagation channel. Typically, the signal measured at the receiver consists of multiple copies of the same transmit signal produced by different paths in the propagation environment. The multipaths cause different wavefronts to impinge, with uncorrelated phases, on the receive antenna. The wavefronts add up constructively or destructively, yielding, over time (time selectivity), fluctuations (i.e., peaks or fades) in the signal strength].

Figure 12.21 (a) shows the power angle vs. delay profile of typical outdoor channel environments. With different delays, the wavefronts impinge on the antenna array from different angles of arrival (ranging between −60fl and 60fl with respect to the broadside direction of the uniform linear array). The impinging rays are clustered around a few angles and delays. Each cluster identifies the energy coming from one specific scattering object in the propagation environment. The higher the number and angle spread of the clusters, the lower the channel spatial correlation. Figure 12.21 (b) shows the effect of the spatial correlation on the signal power measured at two different antennas of the array. Low spatial correlation (i.e., large antenna spacing) causes the received signals h11 and h21 fading independently, resulting higher spatial diversity.

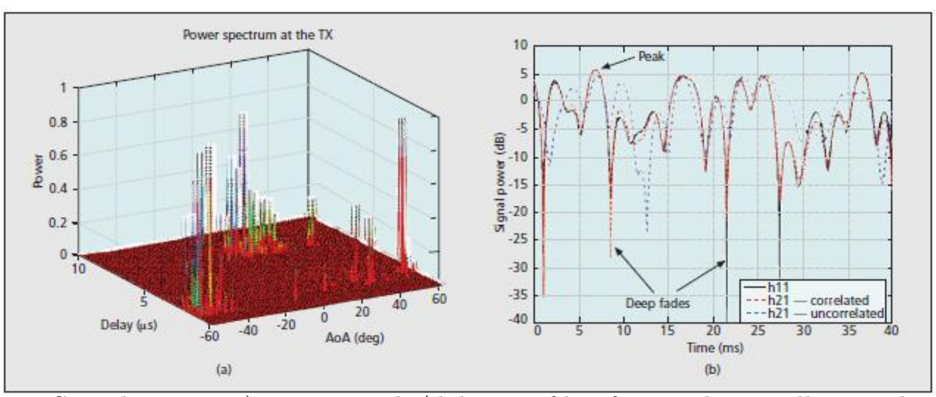


Figure 12.21: Signal power: a) power angle/delay profile of typical spatially correlated channel environments; b) signal power measured at two different antennas of a MIMO array, with and without channel spatial correlation. The temporal channel fading is due to Doppler effects.

A framework for adaptive MIMO transmission is depicted in Figure 12.22.

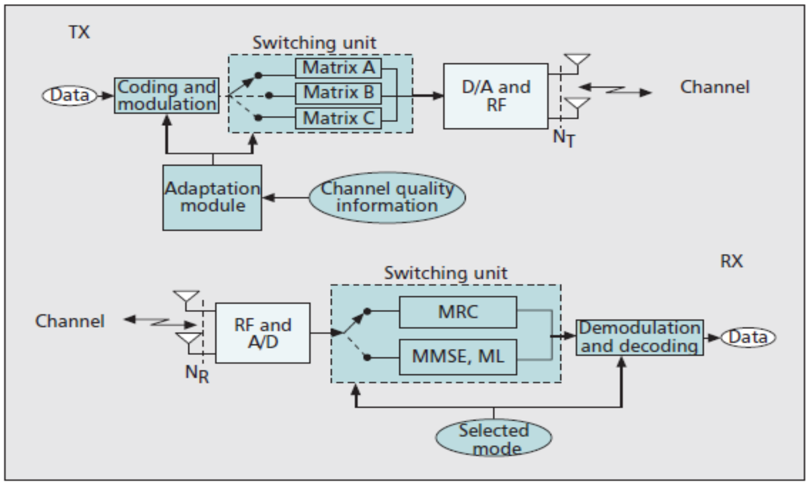


Figure 12.22: Block diagram of adaptive MIMO communication systems. Note that Matrix A is diversity, Matrix B is spatial multiplexing plus diversity, Mtrix C is spatial multiplexing methods.

The receiver estimates the channel quality information (CQI) and sends it back to the transmitter. The transmitter processes the CQI and selects the best transmission mode (i.e., combination of MCS and MIMO method: diversity, spatial multiplexing, or beamforming). The receiver is informed of the new selected mode via a low-rate control channel and adaptively switches between different receivers, depending on the selected mode. A key design challenges of adaptive MIMO architectures is to define efficient adaptation modules that use a minimal amount of feed-back information. Different adaptation criteria can be defined to exploit the time, frequency, and spatial selectivity of the wireless channel.

12.4 MIMO-OFDM SYSTEMS

Figure 12.23 shows a typical MIMO-OFDM system that takes the benefits of both MIMO and OFDM techniques.

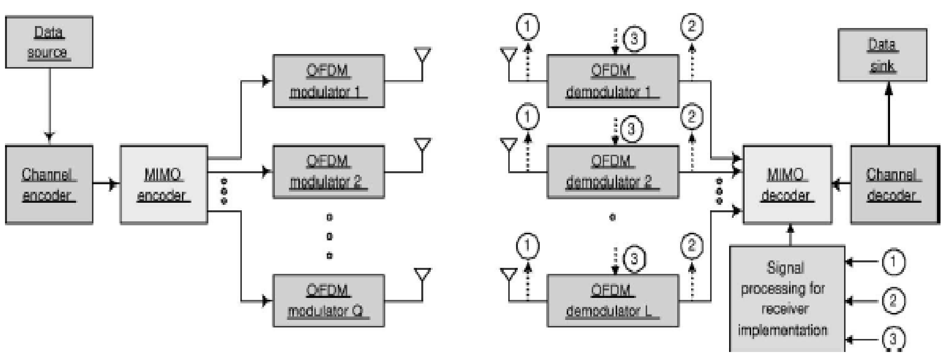


Figure 12.23: Q X L MIMO-OFDM system, where Q and L are the numbers of inputs and outputs, respectively .

12.5 CATEGORIZATION OF MULTIPLE-ANTENNA TECHNIQUES

Spatial diversity, Spatial multiplexing, and Beamforming techniques: based on the benefits

- spatial diversity to provide smaller error rates
- spatial multiplexing to provide higher bit rates,
- beamforming to suppress interference, higher bit rate or smaller error rates

SIMO, MISO, and MIMO techniques: according to the number of transmit and receive antennas used

- SIMO: techniques that only utilize multiple receive antennas
- MISO: techniques that utilize multiple transmit antennas only
- MIMO: techniques that require multiple antennas at both ends of the wireless link

Narrowband and Broadband techniques: based on channel characteristics

- Narrowband: Transmission techniques that are designed for frequency flat fading channels are called narrowband techniques
- Wideband: transmission techniques that are suitable for frequency-selective fading channels are referred to as wideband or broadband techniques.

Open-loop, Closed-loop, and Non-coherent techniques: based on channel knowledge requirement

- Open-loop: Transmission techniques that require no channel knowledge at the transmitter side are referred to as open-loop techniques (spatial-multiplexing) Closed-loop: transmission techniques that require full or partial channel knowledge at the transmitter (beamforming).

Co-located and Distributed MIMO systems: based on antenna array physical placement

- Co-located MIMO: the antennas at transmitter and receiver are part of some sort of antenna array.
- Distributed MIMO: antennas can also be spatially distributed on a large scale (virtual MIMO system)

Single-user and Multi-user MIMO techniques: based user scenarios

- Single-user MIMO: for broadcast scenarios
- Multi-User MIMO: for multiple-access scenarios.

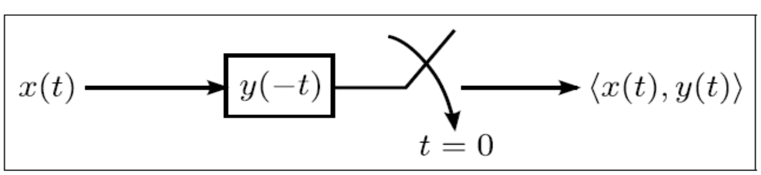
CHAPTER 13
ORTHOGONAL FREQUENCY DIVISION MULTIPLEXING

13.1 FUNDAMENTALS OF ORTHOGONALITY

Inner product between (real) signals x(t) and y(t) can be written as [1]

$$\langle x(t), y(t)\rangle = \int x(t)y(t)dt$$

The inner product can be computed as follows.



Signals x(t) and y(t) are said to be orthogonal if the inner product between them is zero, i.e.

$$\langle x(t), y(t)\rangle = 0$$

Orthogonal Multiplexing

Consider two data symbols a and b. And two orthogonal signals x(t) and y(t). Assume

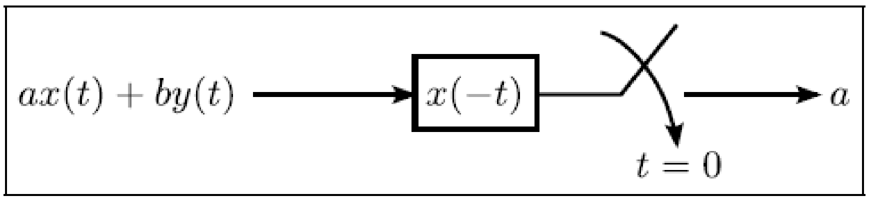
$$\langle x(t), x(t)\rangle = \langle y(t), y(t)\rangle = 1$$

if the transmitted signal is ax(t) + by(t), we can recover the symbol a as follows.

$$\langle ax(t) + by(t), x(t)\rangle = a\langle x(t), x(t)\rangle + b\langle y(t), x(t)\rangle$$

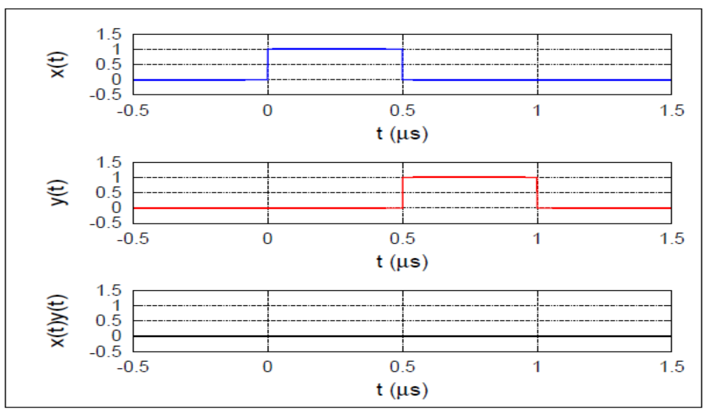
$$= a + 0 = a$$

The block diagram representation is given below for symbol recovery at the receiver.



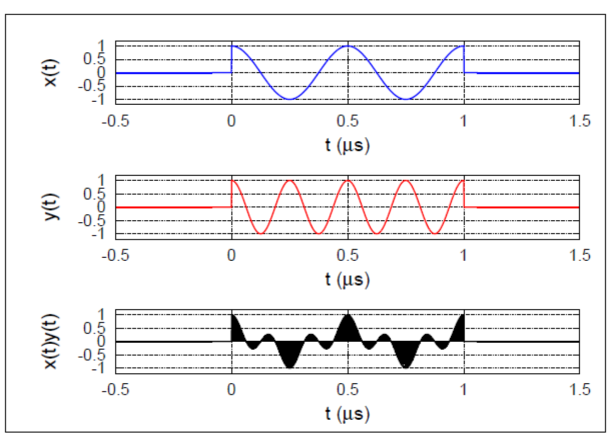
Orthogonality in Time-domain

Time division multiplexing (TDM) [1]

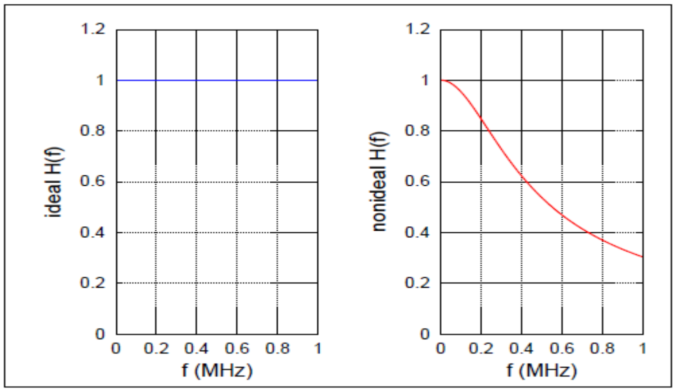


Orthogonality in Frequency-domain

Frequency division multiplexing (FDM)

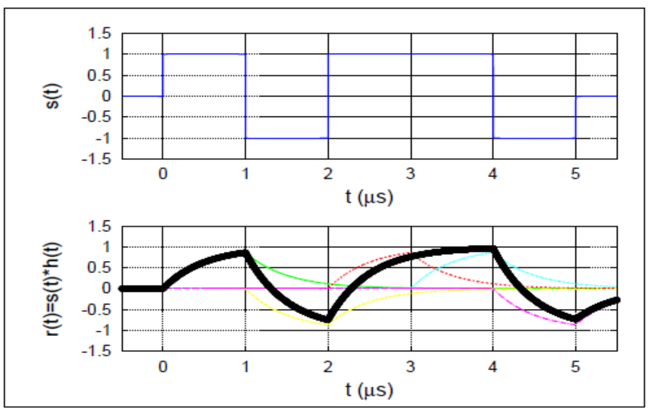


Ideal (flat for all frequency) channel response.
Non-ideal (different frequency components are attenuated differently) channel response.



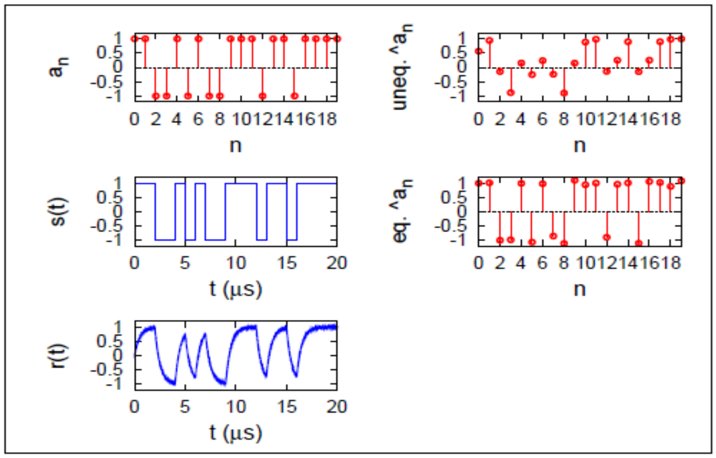
Pulse Spreading

Pulse spreading due to channel frequency selectivity (time dispersive) that causes inter symbol interference (ISI).



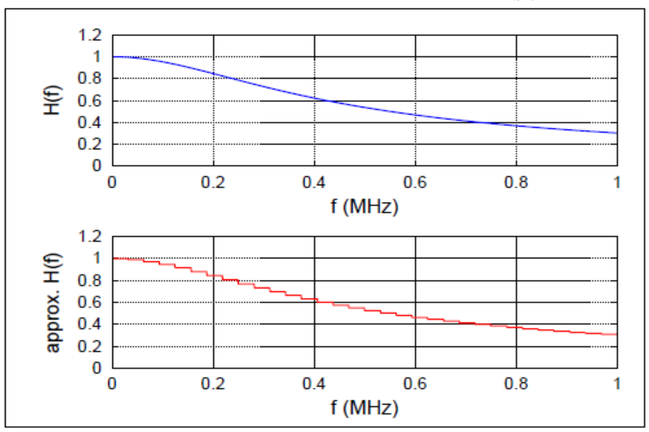
Equalization

Equalization is needed for proper recovery of signals at the receiver and hence to reduce bit errors. Equalization needs the knowledge of the channel response H (f) in transmission band [1].



13.3 MULTI-CARRIER TRANSMISSION

In contrast to single carrier, in multi-carrier transmission channel bandwidth is divided into a number of narrower band (called as sub-carrier in OFDM) so that the channel response remains constant over the narrowband. Consequently immune to ISI.



13.4 INTRODUCTION TO OFDM

- multi-carrier transmission scheme where a wideband is split into a large number (typically several hundreds) of narrowband carriers called subcarriers and transmitted over the same radio link to the same receiver.
- tight frequency domain packing of OFDM subcarriers.
- two contiguous OFDM subcarriers are mutually orthogonal to one another over the OFDM symbol interval.
- OFDM transmission is block based, i.e. a number of modulated symbols are transmitted in parallel.
- orthogonality between symbols is specific frequency domain structure based [2] .

Per subcarrier spectrum (sinc-square shaped) in frequency domain that corresponds to a rectangular pulse shape in time domain are given in Figures 13.1 and 13.2. Figure 13.3 represents OFDM subcarriers spacing where the spacing T_u is symbol time interval. In addition, time-frequency resource grid where symbols are allocated for transmission is given in Figure 13.4.

An illustrative example of OFDM modulator is given in Figure 3.5.

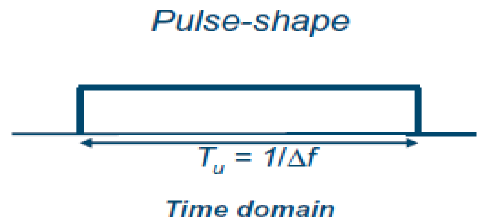


Figure 13.1: A subcarrier in time domain.

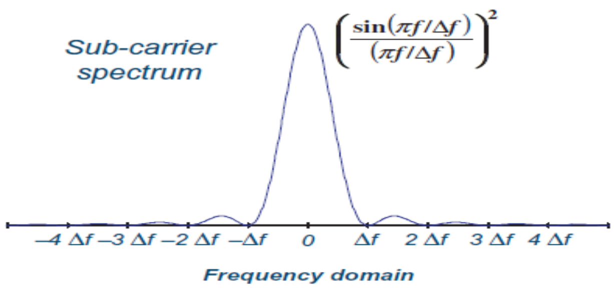


Figure 13.2: Corresponding subcarrier spectrum in frequency domain .

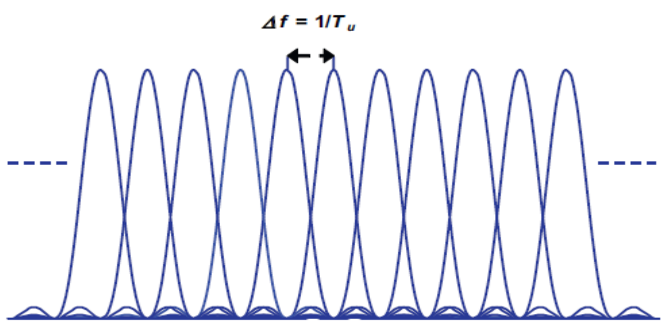


Figure 13.3: OFDM subcarriers and spacing .

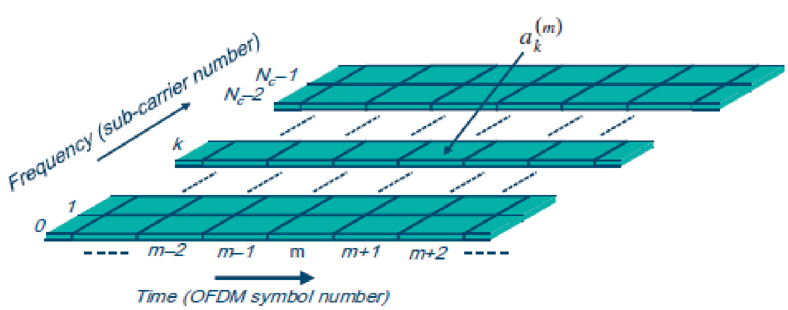


Figure 13.4: OFDM time-frequency grid.

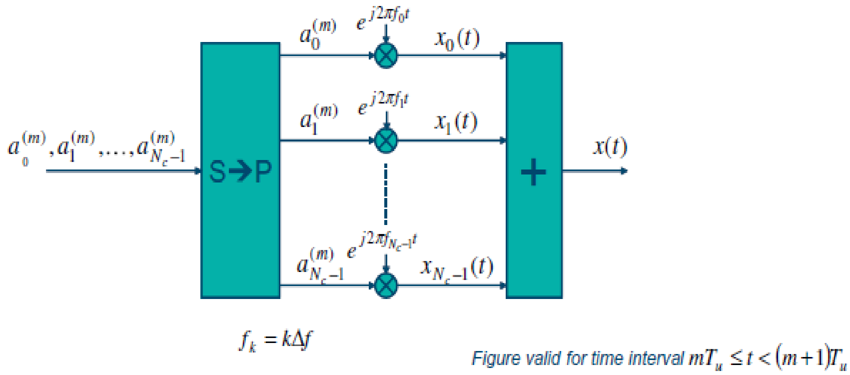


Figure 13.5: Basic OFDM modulator.

OFDM Modulator

Consider in any symbol interval $mT_u \le t < (m+1)T_u$, x(t) denotes an OFDM signal. x(t) can expressed as follows where k denotes the subcarrier number, and m denotes an arbitrary OFDM symbol interval number. Also denote $a_k^{(m)}$ as the modulated subcarrier and $x_k(t)$ is the modulation symbol applied to the subcarrier. Then OFDM signal x(t) can be expressed as

$$x(t) = \sum_{k=0}^{N_c-1} x_k(t) = \sum_{k=0}^{N_c-1} a_k^{(m)} e^{j2\Pi\Delta ft} \qquad (13.1)$$

OFDM Demodulator

The demodulation of the OFDM signal x(t) can be performed following the reverse modulation process as shown in Figure 13.6.

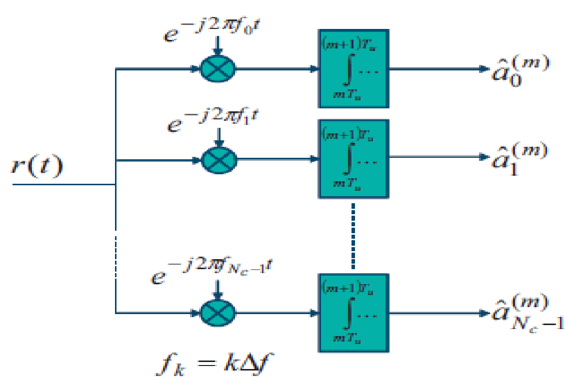


Figure 13.6: OFDM Demodulator.

Cyclic-Prefix

OFDM demodulator consists of a bank of correlators , each corresponds to a subcarrier. If no significant change in channel response is considered, there should not have any inter subcarrier interference because of subcarrier's orthogonality feature, i.e.

$$\int_{mT_u}^{(m+1)T_u} x_{k_1}(t)x_{k_2}^*(t)dt = 0 \quad \text{for } k_1 \neq k_2$$

However, any variation in channel response can be overcome by employing cyclic-prefix insertion, shortly described as follows.

Shown in Figure 13.7 is a typical time dispersive (frequency selective) channel and the received time dispersive signal (reflected path signal) response. Channel time dispersive characteristics causes overlap between successive symbols (in demodulator correlator interval) , which may arrive the receiver through different paths. Consequently, inter –symbol as well as inter-subcarrier interference occur.

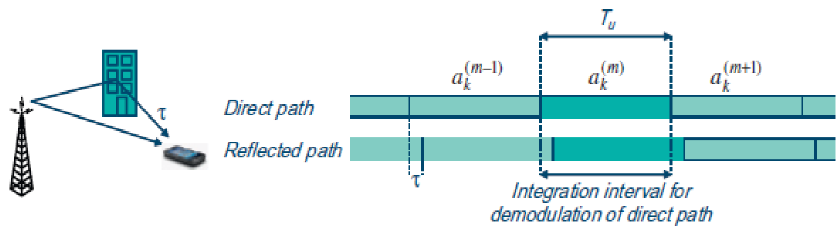


Figure 13.7: channel time dispersion and the corresponding received signal.

Cyclic-prefix insertion can overcome this interference in OFDM transmission. Cyclic-prefix insertion implies copying last part (length of what depends on the channel characteristics) of the OFDM symbol and inserting the same at the beginning of the same OFDM symbol (Figure 13.8). As long as the span of channel time dispersion is less than cyclic-prefix length (T_{CP}), Orthogonality property still upholds even if the correlation performs over the symbol time interval.

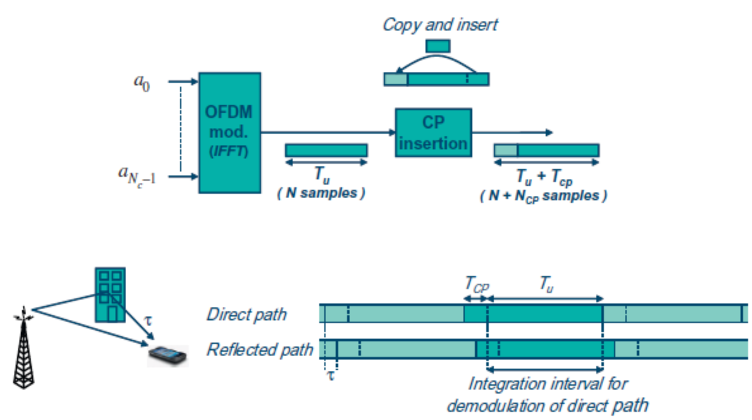


Figure 13.8: Cyclic-prefix insertion and avoidance of interference in OFDM signal.

However, Cyclic-prefix causes additional ($T_{CP} / (T_u + T_{CP})$) power loss at the demodulator as well as loss in bandwidth use because of additional Cyclic-prefix bits.

13.5 OFDM TRANSMISSION-RECEPTION MODEL: FREQUENCY DOMAIN

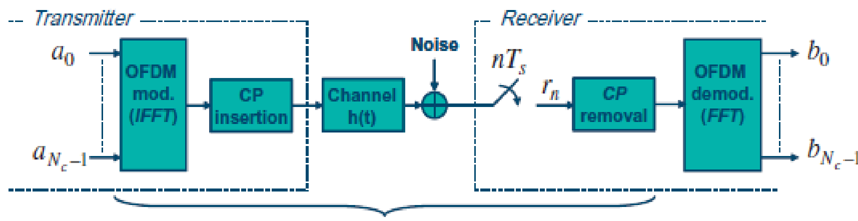
Denote OFDM symbols as $a_0 \dots a_{N_c-1}$, frequency domain channel taps $H_0 \dots H_{N_c-1}$ (derived from channel impulse response), and noise as $n_0 \dots n_{N_c-1}$

The received signal can be represented as b=H . a + n, i.e.

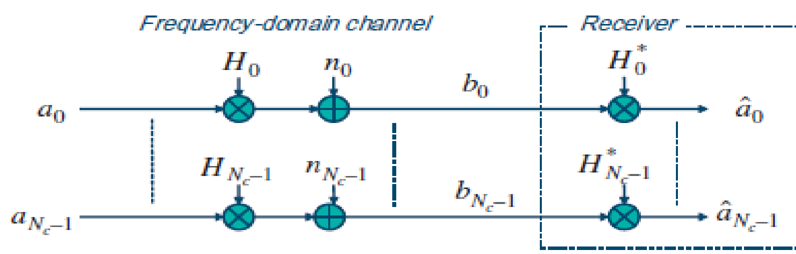
$$b_i = H_i a_i + n_i$$

where i=0,…,$N_c - 1$. The received signal is first demodulated and then scaled with complex conjugate of the channel taps (multiplication of b_i with H_i^*) for proper reception of the transmitted OFDM symbols. In literature, this is often termed as one-tap equalizer (that is being applied to each subcarrier). The model is illustrated in Figure 13.9.

The channel response (frequency-domain channel taps H) can be estimated directly by sending reference (some used to say pilot) symbols along with information symbols in certain regular intervals in the time-frequency grid, around which channel response is estimated (Figure 13.10).



(a)



(b)

Figure 13.9: (a) OFDM transmission-reception model (frequency domain) ; (b) corresponding one-tap equalization at the receiving end.

13.6 OFDMA

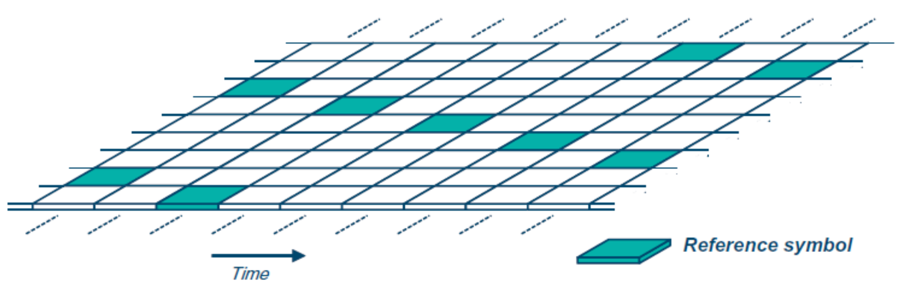
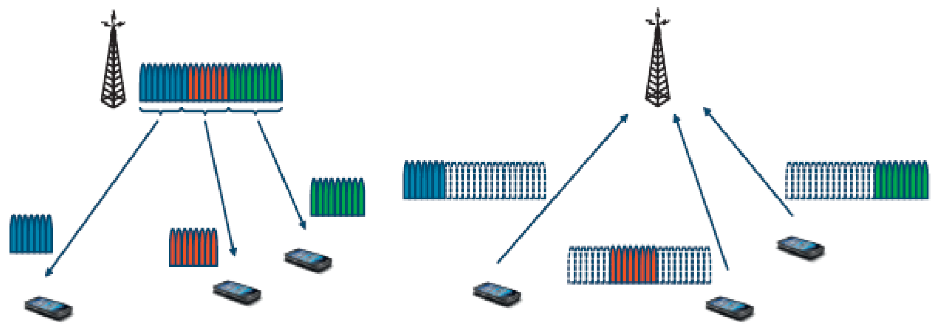


Figure 3.10: Reference symbols in OFDM time-frequency grid.

OFDM can be used as user multiplexing or multiple access scheme termed as OFDMA (orthogonal frequency division multiple access) . In OFDMA, orthogonal frequencies are assigned simultaneously among several users for downlink. Similarly, for uplink, several users are transmitting simultaneously each user employing a subset of all subcarriers in the system bandwidth to the base station (BS). the distribution of subcarriers, both for uplink and downlink, can be consecutive or distributed over the subcarriers. The principle is illustrated in Figures 13.11 (consecutive) and 13.12 (distributed).



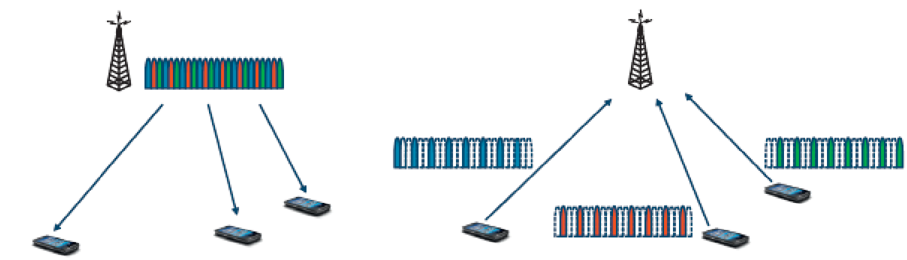


Figure 13.11: (a) OFDMA (consecutive subcarrier distribution);
(b) OFDMA (distributed subcarrier distribution).

13.7 ADVANTAGE AND DISADVANTAGE OF OFDM

In comparison with single-carrier modulation, Multi-carrier OFDM is

Advantage

1. Immune to inter symbol interference (ISI)
2. Relatively insensitive to time synchronization
4. Simple single-tap equalization
3. Efficient IFFT/FFT based implementation

Disadvantage
1. Overhead due to cyclic prefix
2. Highly sensitive to frequency synchronization
3. High signal peak-to-average power ratio (PAPR)

13.8 OFDM APPLICATIONS IN PRACTICAL SYSTEMS

1. Asymmetric Digital Subscriber Line (ADSL)
2. IEEE 802.11 (Wi-Fi: Wireless Fidelity)
3. IEEE 802.16 (WiMAX: Worldwide Interoperability for Microwave Access)
4. Advanced 3G and 4G cellular systems
5. Digital TV broadcasting
6. IEEE 802.15 for personal area networks (PANs)
7. ITU-T G.hn for powerline communications (PLC)

CHAPTER 14
RESOURCE SCHEDULING, LINK ADAPTATION AND HYBRID AUTOMATIC REPEAT REQUEST

14.1 INTRODUCTION

Mobile radio communication suffers mainly from fading and distance dependent path loss and normally the channel varies significantly, to some extent random in nature. Channel dependent scheduling is one such approach that can address the frequently varied channel by efficiently allocating resources among users and consequently the system per-cell capacity (more users can be served at a time). Link adaptation is another approach that complements the scheduling decision process by adapting the radio-link parameters according to the channel variation. Since perfect adaptation to channel variation is almost impossible, there is a need for repetition mechanism termed Hybrid automatic repeat request (HARQ) that deals with requesting erroneously received signal to be retransmitted so that the received bit error is reduced.

14.2 LINK ADAPTATION

Link adaptation can be performed mainly in two ways such as controlling transmit power and controlling data rate. Transmit power control responds to the channel variation inversely as shown Figure 14.1 where transmit power is increased if the channel condition degrades and vice versa. This dynamic power control provides the receiver with a constant data rate and hence useful for constant rate connection such as voice communication.

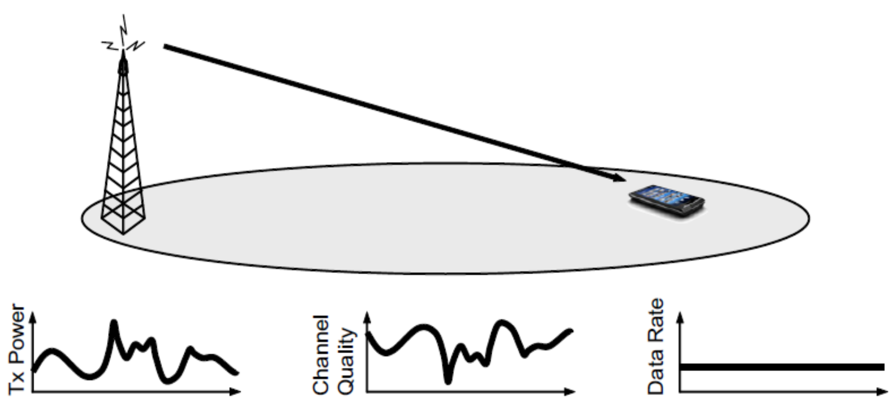


Figure 14.1: Link adaption by power control.

Rate control is mainly applied when the requirement for instantaneous constant data is not important, instead the average data rate. The transmit power can be kept constant while the data rate is varied by dynamically changing modulation order and coding scheme. Figure 14.2 illustrates an example rate control mechanism where the data rate varies according to the channel quality.

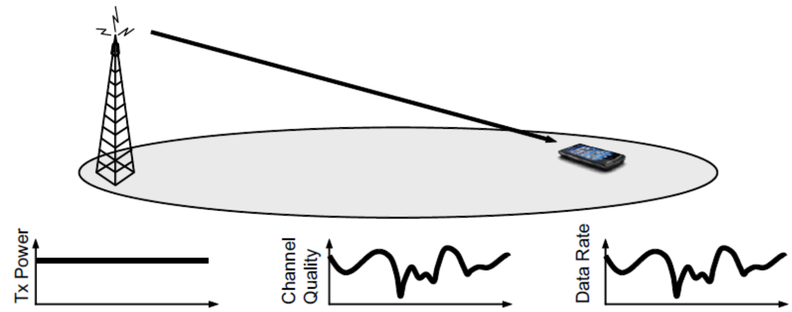


Figure 14.2: Link adaptation by rate control.

If the channel condition is good, higher-order modulation such as 16QAM or 64QAM and high-rate coding can be used, whereas for poor channel condition lower-order modulation such as QPSK and low-rate coding can be used. Note that here implicitly we consider that the channel bandwidth is fixed. This technique of changing modulation scheme and channel coding rate to adapt with the channel link variation is called adaptive modulation and coding (AMC).

14.3 CHANNEL DEPENDENT SCHEDULING

Channel dependent scheduling relies on the instantaneous channel conditions and assigns resources (time, frequency, space, etc.) accordingly. Scheduling resources is an implementation perspectives and is not normally specified in standards. However, channel quality measurement and signalling mechanism is included.

In general, scheduling defines how the system resources are to be shared among users in a cell and differs one strategy from another based on the radio interface in consideration (uplink or downlink) as well as whether simultaneous multi-user transmissions are orthogonal or not .

14.4 DOWNLINK SCHEDULING

in general transmission to users from the BS in a cell is orthogonal and consequently no intra-cell interference. The scheduling strategy can be time division multiplexing, (TDM), frequency division multiplexing (FDM) or code division multiplexing (CDM) based in association with space.

Generic scheduling principles

There are basically three well known basic schedulers such as round robin(RR), proportional fair (PF) and max-SNR/max-rate.

Max-SNR Scheduling

The **Max-SNR** (MSNR) or Max Rate scheduler takes into consideration the current or instantaneous channel condition in terms of SINR for scheduling users in each time slot. Mathematically, the scheduling decision metric can be expressed as follows.

$$k_{i,j}(\tau) = \arg \max_{k} \left(\gamma_{k,j,i}(\tau)\right)$$

where $k_{i,j}(\tau)$ is the selected user in the i^{th} CC at the j^{th} RB in time ; τ is the SINR of the user k in the CC at the RB in time .

Proportional Fair Scheduling

The Proportional Fair (PF) scheduler provides an optimal throughput demand while ensuring optimal

fairness performance. The PF scheduler ensures it by taking into account the instantaneous channel state information from the user equipment (UE) through uplink by a control signaling channel. The PF scheduling algorithm is based on maximizing the scheduling metric, i.e. for any RB j in any TTI , the PF scheduler schedules only the user with the maximum performance metric (max PM).
Mathematically, it can be expressed as follows [3]:

$$k_{i,j}(\tau) = \arg\max_{k} \left\{ \frac{R_{k,j,i}(\tau)}{\tilde{R}_{k,i}(\tau)} \right\}$$

where $k_{i,j}(\tau)$ is the selected user in the i^{th} CC at the j^{th} RB in time; $R_{k,j,i}(\tau)$ is the estimated throughput of the user k . $\tilde{R}_{k,i}(\tau)$ is the average throughput in the past of the user k .

The throughput of the user is updated according to the following condition.

$$\tilde{R}_{k,i}(\tau+1) = \begin{cases} \tilde{R}_{k,i}(\tau)(1-\frac{1}{t_c}) + \frac{1}{t_c} R_{k,j,i}(\tau), & k = k_{i,j}(\tau) \\ \tilde{R}_{k,i}(\tau)(1-\frac{1}{t_c}), & k \neq k_{i,j}(\tau) \end{cases}$$

t_c is the adjusted time constant or fairness factor. The time constant should be sufficiently large so that it can capture both short term and long term channel variation and typically it is in the order of second, 211

Round Robin Scheduling

The Round Robin (RR) scheduler is considered for user admission control in time domain. The RR scheduler can be used for layer-3 carrier load balancing since it provides the best fairness criterion of all schedulers. The RR scheduler works on distributing the load evenly in all CCs. Hence, when there is an arrival, the RR scheduler simply looks for the CC which has the least number of users.

Once it finds the CC with the least number of allocated users, it assigns the new arrival (user) to that CC (Figure 14.3). This ensures the full load and the uniform user distribution in all CCs. Figure 14.4 illustrates channel dependent scheduling and Examples of three different scheduling behaviors for two users with different average channel quality is shown in Figure 14.5.

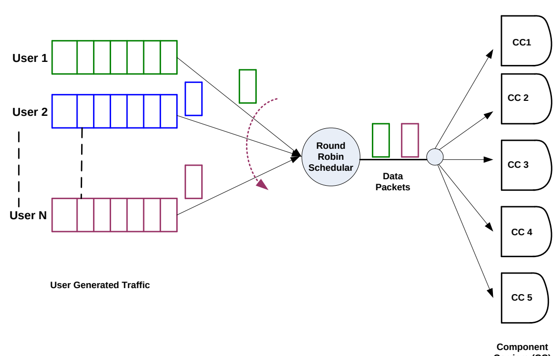


Figure 14.3: A typical RR scheduler operation.

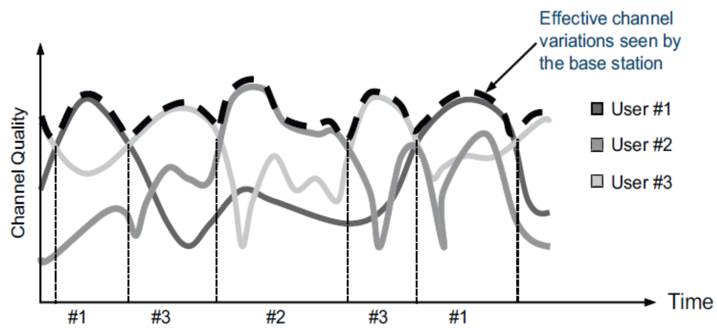


Figure 14.4:Channel-dependent scheduling.

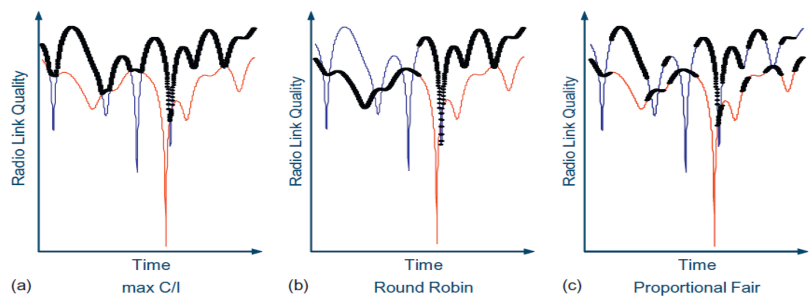


Figure 14.5: Examples of three different scheduling behaviors for two users with different average channel quality: (a) max-C/I; (b) round robin; (c) proportional fair. The selected user is shown with bold lines.

14.5 TIME-FREQUENCY GRID

In LTE-Advanced system, OFDM is used both in the uplink and downlink. We shall consider LTE-Advanced for the discussion of basic schedulers operations. OFDM can be used as user multiplexing or multiple access scheme termed as OFDMA (orthogonal frequency division multiple access). In OFDMA, orthogonal frequencies are assigned simultaneously among several users for downlink. Similarly, for uplink, several users are transmitting simultaneously each user employing a subset of all subcarriers in the system bandwidth to the base station (BS) (Figure 14.6).)

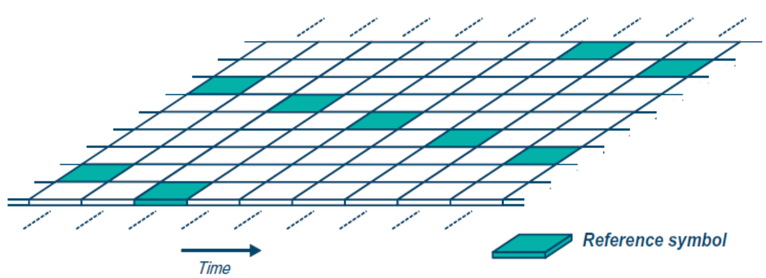


Figure 14.6: Reference symbols in OFDM time-frequency grid.

14.6 EXAMPLE MODEL OF SCHEDULERS

The main purpose of the scheduler is to assign user packets to RBs in the CC. The generated traffic from user equipments (UEs) waits in their respective buffers (queues). The scheduler makes the decision on traffics to assign to what RB or based on a number of aspects such as priority, channel condition, packet size, traffic type, etc. Typical joint and disjoint schedulers [4] are illustrated in Figures 14.7 and 14.8 respectively.

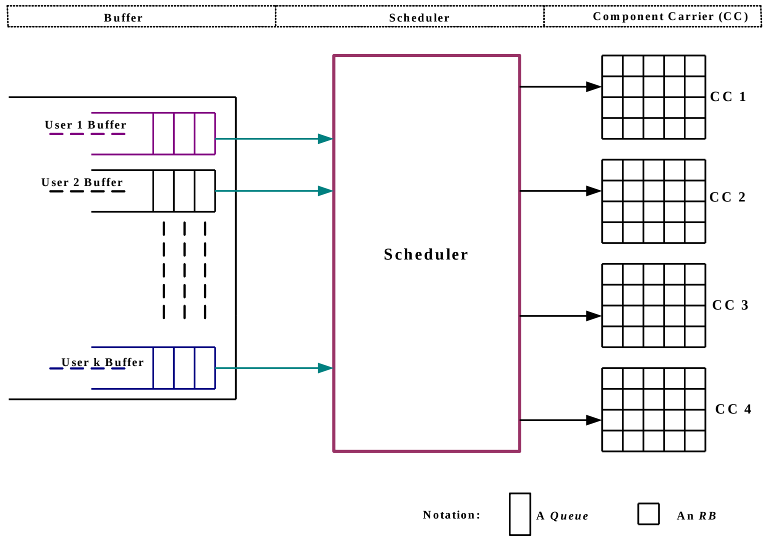


Figure 14.7: Typical model of a joint scheduler (LTE-Advanced system .

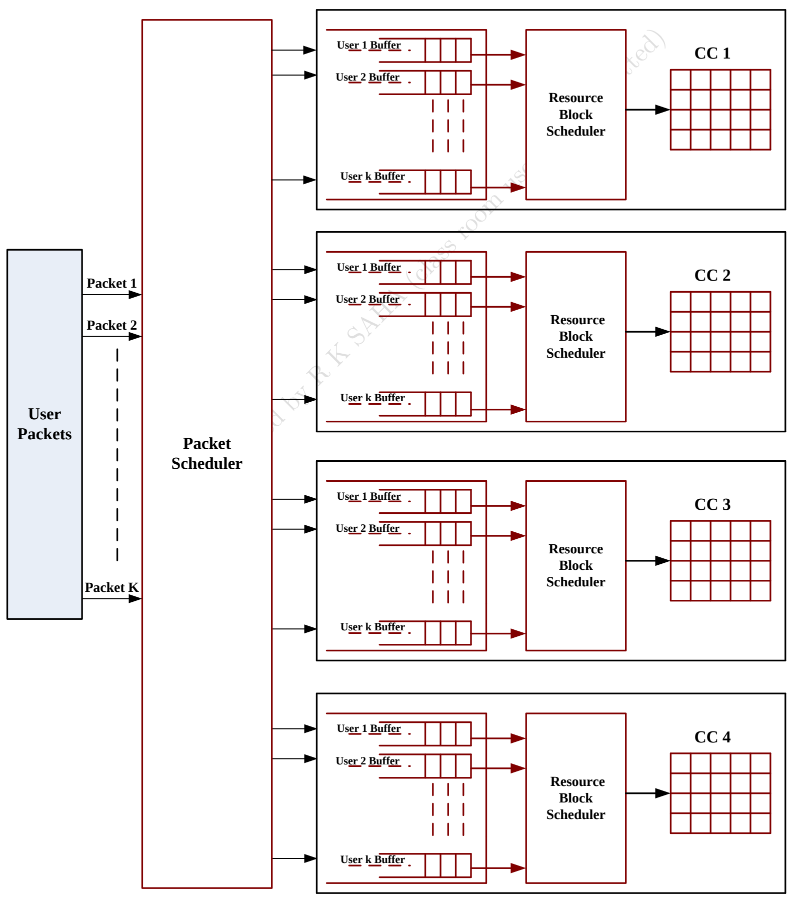


Figure 14.8: Typical model of a disjoint scheduler (LTE-Advanced system).

Implementation of Schedulers

CQI feedbacks from UEs play a vital role in proper scheduling operation, both in time and frequency domain. Since the packet scheduler considered in this study is the PF scheduler, the performance metric of the PF scheduler is calculated based on the instantaneous channel quality. A typical CQI reporting and radio resource management in LTE systems is shown in Figure 14.9.

The function of the Link Adaption (LA) is to adapt with the variable channel, both in time and frequency, by adapting the modulation and coding scheme (MCS). The CQI report is used for this purpose by consulting with the CQI manager in the physical layer. The CQI manager stores CQI codeword, which are transmitted by UEs. LA comprises of two blocks: Inner Loop LA (ILLA) and Outer Loop LA (OLLA). The OLLA handles the BLER, and provides an offset margin (based on NACK or ACK received), which is subtracted from the estimated SINR in the CQI manager before the SINR is being used by the ILLA. The ILLA estimates the instantaneous supported throughput and MCS based on the CQI report and the offset provided by the OLLA.

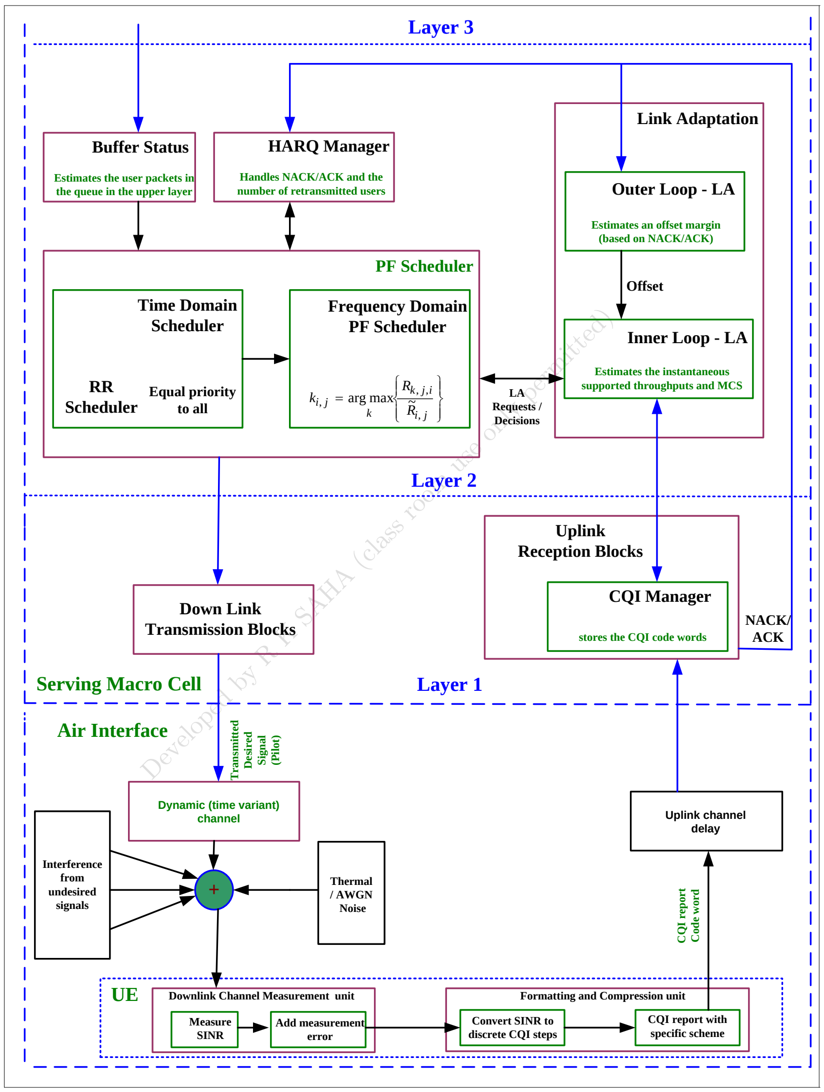


Figure 14.9: A typical CQI reporting and radio resource management in LTE systems.

14.7 CHANNEL FEEDBACK MECHANISM FOR CHANNEL DEPENDENT SCHEDULING

Feedback Mechanism

As already motioned, CQI is used at BS to configure the MCS and to estimate the performance metric in a scheduler. In the CQI simulation environment, there are four major steps that are followed to measure and report the CQI to BS at the UE as follows.
Step 1: Measure SINR
Step 2: Introduce measurement error to SINR

Step 3: Convert SINR values to discrete CQI steps
Step 4: Report CQI with a specific scheme

The SINR is calculated for each RB (j) from the pilot signal and interference signals received at the UE every measurement period.
The measured SINR value in addition to measurement error for each RB is then converted into decibels and is given by the following equation.

$$SINR_{dB}(j) = 10 \log_{10}\left(SINR(j)\right) + error\ (dB)$$

where the measurement error is a Gaussian distributed error with zero mean and parameter specified variance.

The SINR values are then converted to discrete CQI values using quantization steps as follows.

$$CQI_{dB}(j) = Q_{step}(dB) \times floor\left(\frac{SINR_{dB}(j)}{Q_{step}(dB)} + 0.5\right)$$

The measured CQI values are then sent with a certain delay and using a CQI reporting scheme based on the number of CQI reports per TTI from the highest granularity scheme such as Full CQI reporting where every CQI values (Quantized) are reported at the achievement of highest accuracy but at the expense of highest uplink signal overhead, optimal granularity where certain portion of the CQI values reported such as Best-M CQI reporting where compression is based on M number of RBs with the highest CQI values, Threshold based CQI reporting where RBS with CQI values in a threshold window set from the highest CQI value are averaged and the CQI average value is add to the codeword (explained in detail below), or Discrete Cosine Transform (DCT-Significant M) based CQI reporting scheme where compression is based on the M DCT output coefficient with the highest absolute values, and to the lowest granularity scheme such as Wideband CQI where an average value of CQI over all RBS are calculated (no frequency selective gain) to the BS.

General Feedback Schemes

A brief detail of the wide band CQI reporting scheme, Full CQI reporting scheme, and the optimal Threshold reporting scheme with a high level of schematic representation are described below.

Assumptions

Measured CQI is post detection SINR.
The total spectrum =20 MHz per CC
Periodic SINR measurement for B RBs.
Each RBs 12 subcarriers with each 15 KHz bandwidth
The number of measurement intervals (quantization steps) is b = B = 100 as measurement is done at each sub-band
Average measurement time is 2 ms

Gaussian distributed CQI measurement error with zero mean and 1 dB standard deviation
The number of encoded bits for each quantized value $L_q = 7$, because to obtain Quantization steps corresponding to CQI error std. and as the dynamic range for LTE MCS is from QPSK with excessive coding to 64 QAM with marginal coding is approximately 25 dB, $L_q = 7$ bits is enough.

Wideband CQI Reporting Scheme

In wideband CQI reporting scheme (Figure 5.10), the average value of all the CQI value corresponding to all RBs is measured and the average CQI codeword corresponding to the average value is reported to the BS. Hence it does not take into account of the frequency selective gain and the CQI error is high as the actual RBs that are allocated to any UE may not be the same as the reported recommended RBs by CQI. This is because it sends the average CQI value where may most of them can have optimal channel feedback but some UEs may be allocated to RBs which has worst channel condition.

Full CQI Reporting

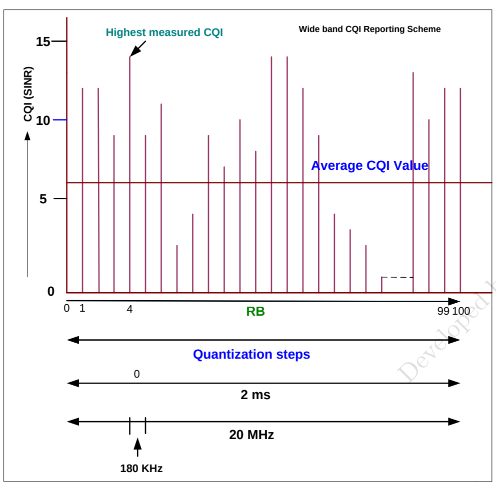
In Full CQI reporting scheme (Figure 14.11), every RB CQI is reported in each measurement time to the BS from every users. This is totally the reverse phenomenon with respect to wide band CQI reporting. Hence this the best CQI reporting scheme in terms of accuracy in CQI reports but it comes at the expense of highest signal overhead (2500% compared to Wideband).

Total CQI report bits in one TTI (1 ms) = B. L_q= 100*7 bits =700 bits
Total CQI report Bit rate = 700 kbits/s (which account a large uplink signal overhead).

Threshold-based CQI Reporting

As explained above the highest number of CQI index values are identified and a threshold window (5 dB is optimal) is set from the highest CQI index value. All CQI value within the threshold window is taken for average CQI measurement and this average CQI value is reported as part of the CQI codeword to the BS. So the Total CQI report bits (codeword) = L_q + B = 100+7 =107 bits
The total CQI report Bit rate = 107 kbits/s in one TTI (1 ms).

The codeword (as an example) of threshold scheme shown in Figure 14.12 is the first 100 bits can help notify the CQI manager what sub-bands is included in the threshold window by bit "1" and by "0", the group of sub-bands that are not included in the window. This is termed as unclaimed RBs or RBs and their channel quality are taken as the CQI value, some constant decibels lower than the reported CQI value in the code word .



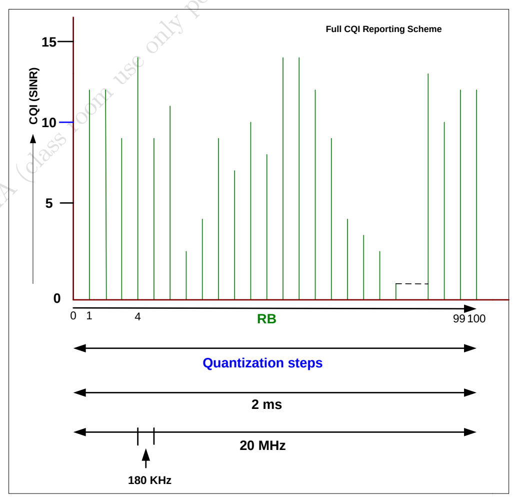


Figure 14.10: Wideband CQI Reporting scheme. Figure 14.11: Full CQI Reporting scheme.

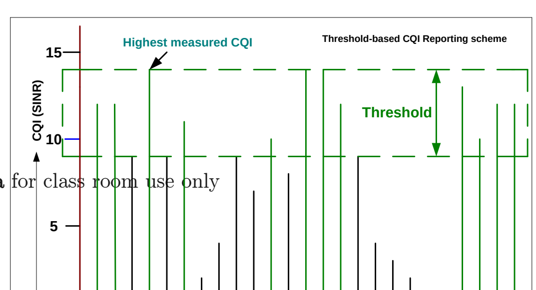


Figure 14.12: Threshold CQI reporting scheme.

14.8 UPLINK SCHEDULING

In the uplink the same basic scheduling strategies are applicable as in the downlink. Fundamentally in the uplink transmitting power is distributed among user equipments (UEs) and hence much lower than the BS transmit power. Channel dependent scheduling such as PF can be used in the uplink if the channel information as described is known. Otherwise, RR scheduling can be used. If non-orthogonal multiple access is used such as CDMA, power control mechanism as described should be used for proper operation to address intra-cell interference as well as minimum error in E_b/N_0 . However for orthogonal multiple access such as OFDM, no intra-cell interference exist and not necessary for power control.

In addition, near-far distance effect should also be taken into account since MS far from the BS normally operates in power limited region and those close to the BS in the bandwidth limited region because of limited uplink Tx power. Hence it is wise to allocate more BW to the near users rather than the those MSs who are far apart from the BS.

14.9 TRAFFIC BEHAVIOR AND SCHEDULING

As of now, we implicitly consider full buffer traffic scenario where the buffer is always is always full, i.e. is never get empty. However, for traffic that are bursty in nature (block of data transmit at a time and discontinuous) such as web page, email, etc. the performance of the scheduler aforementioned are different. Generally max-SNR can provide even some degree of fairness when the traffic is bursty rather than full buffer case. However, if the traffic is considered as full buffer, then any cell-edge user may not schedule for a long time by a max-SNR and very poor fairness performance since it almost allocates all resources to the MSs in the cell center. However, since PF scheduler takes short-term channel variations into account in addition to the long-term average throughput, in either of these traffic scenarios, PF scheduler provides almost similar fairness performance in resource allocations (Figure 14.13).

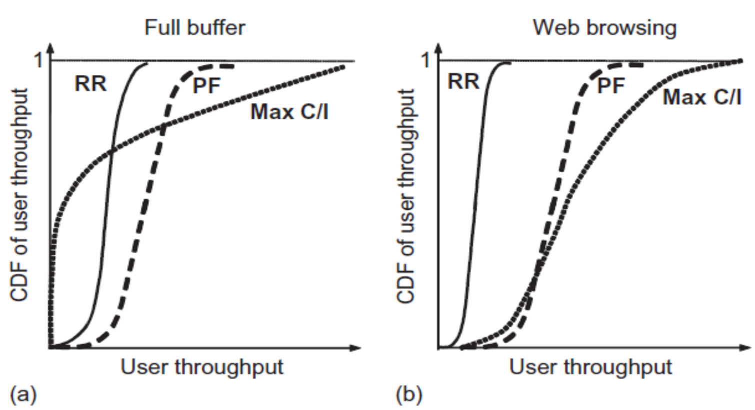


Figure 14.13: Illustration of the principle behavior of different scheduling strategies: (a) for full buffers; (b) for web-browsing traffic model.

Since wireless channel is susceptible to unpredictable variations in response to signal transmitted over the channel, received signals is not error free. In order to address this errors in the received signals, one way is to employ forward error correction (FEC) that is basically adding some redundant bits to the information bits based on the coding structure used such as convolution code, Turbo code, etc. Another approach is to use error detection technique that employs the principle of requesting error detected received data to the transmitter to retransmit. Typically, Cyclic redundancy check (CRC) code is used for error detection purpose.
The combination of these two: error correction as well as error detection is called Hybrid automatic repeat request (HARQ). Most modern communication systems such as LTE use HARQ.
In HARQ, the erroneous bits are discarded although it contains information. This shortcoming is addressed by HARQ with soft combining where erroneously received data are stored instead and the retransmitted data are then combined and forward to the decoder. If decoding fails a retransmission is normally requested. The retransmission data could be the same as the original.

HARQ with Soft Combining

Transmission or different, base on what there are two categories of soft combining. One is the Chase combining where the same bits are retransmitted as the original transmission (Figure 14.14). Another approach is the incremental redundancy (IR) where each retransmission does not necessarily has to be identical to the original transmission. Rather, multiple sets of coded bits are generated, each representing the same set of information. A retransmission is normally different from the previous transmission. the original transmission and the retransmissions are then combined (Figure14.15).

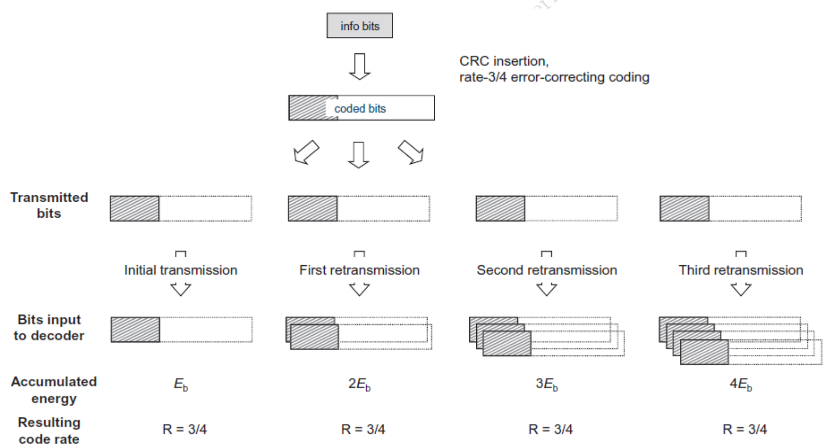


Figure 14.14: Example of Chase combining.

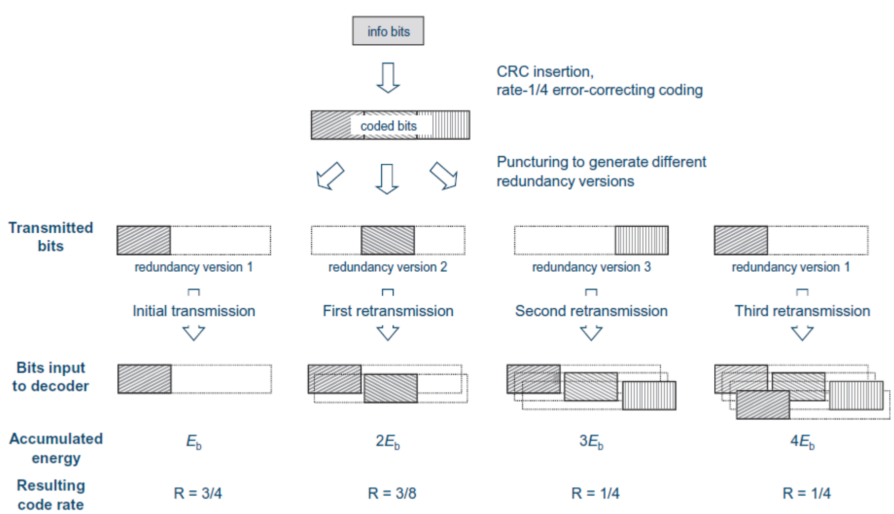


Figure 14.15: Example of incremental redundancy.

CHAPTER 15
WIDEBAND CODE DIVISION MULTIPLE ACCESS

15.1 PREFACE

The third generation (3G) mobile communication systems have been aimed to enable high bit rate in order to support high quality multimedia communications and hence, to cover the limitation of the second generation (2G) systems. The 3G systems are also referred to as Universal Mobile Telecommunication System (UMTS), and Wideband Code Division Multiple Access (WCDMA) is the major 3G air interface in the world. WCDMA, which is developed by the third-generation partnership project (3GPP), is referred to as UMTS terrestrial radio access (UTRA).

3G and WCDMA: An Introduction

Third generation (3G) is mainly designed for multimedia communication, namely as follows.

- person-to-person communication with high quality images and video
- access to information and services on public and private networks with high data rates

- new flexible communication capabilities

In the standardisation forums, WCDMA technology has emerged as the most widely adopted 3G air interface which specification has been created in 3GPP (the 3rd Generation Partnership Project: the joint standardisation project of the standardisation bodies from Europe, Japan, Korea, the USA and China). Within 3GPP, WCDMA is called UTRA (Universal Terrestrial Radio Access) FDD (Frequency Division Duplex) and TDD (Time Division Duplex), however, the name WCDMA being used to cover both FDD and TDD operation. In this course we shall focus on WCDMA FDD technology.

15.2 AIR INTERFACES AND SPECTRUM ALLOCATIONS FOR 3G

The World Administrative Radio Conference (WARC) of the ITU International Telecommunications Union), at its 1992 meeting, identified the frequencies around 2 GHz (Table 15.1) to develop third generation mobile systems. Note that, within the ITU these 3G systems are called International Mobile Telephony 2000 (IMT-2000).
In addition to WCDMA, the other air interfaces that can be used to provide third generation services are EDGE and cdma2000. The expected frequency bands and geographical areas where these different air interfaces are likely to be applied are shown in Figure 15.1. However, within each region there are local exceptions in places where multiple technologies are already being deployed.

Table 15.1: Existing frequency allocations around 2 GHz

	Uplink	Downlink	Total
GSM1800	1710–1785	1805–1880	2×75 MHz
UMTS-FDD	1920–1980	2110–2170	2×60 MHz
UMTS-TDD	1900–1920 and	2010–2025	$20 + 15$ MHz
Americas PCS	1850–1910	1930–1990	2×60 MHz

At the ITU-R WRC-2000 in May 2000 the following frequency bands were also identified for IMT-2000:

- 1710–1885 MHz
- 2500–2690 MHz and
- 806–960 MHz.

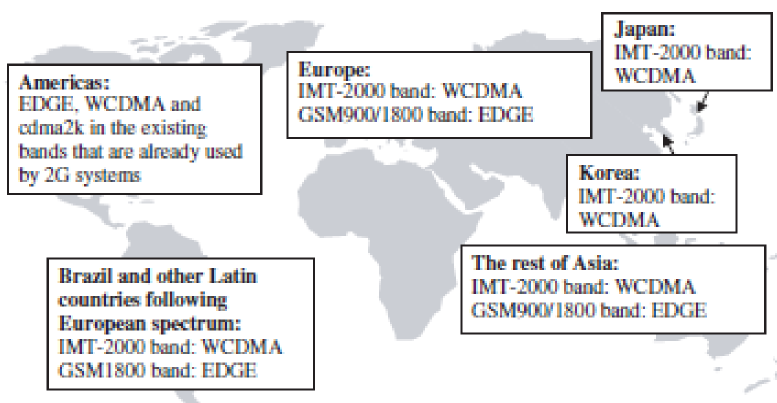


Figure 15.1: Expected air interfaces and spectrums for providing third generation services.

Requirements for 3G systems

The new requirements for 3G systems are as follows.

- Bit rates up to 2 Mbps;
- Variable bit rate to offer bandwidth on demand;
- Multiplexing of services with different quality requirements on a single connection, e.g. speech, video and packet data;

- Delay requirements from delay-sensitive real time traffic to flexible best-effort packet data;
- Quality requirements from 10 % frame error rate to 106 bit error rate;
- Co-existence of second and third generation systems and inter-system handovers for coverage enhancements and load balancing;
- Support of asymmetric uplink and downlink traffic, e.g. web browsing causes more loading to downlink than to uplink;
- High spectrum efficiency;
- Co-existence of FDD and TDD modes.

Specifications of WCDMA

3GPP wideband CDMA specifications are listed below.

Channel bandwidth	5 MHz
Duplex mode	FDD and TDD
Downlink RF channel structure	Direct spread
Chip rate	3.84 Mbps
Frame length	10 ms
Spreading modulation	Balanced QPSK (downlink) Dual-channel QPSK(uplink) Complex spreading circuit
Data modulation	QPSK (downlink) BPSK (uplink)
Channel coding	Convolutional and turbo codes
Coherent detection	User dedicated time multiplexed pilot (downlink and uplink), common pilot in the downlink
Channel multiplexing in downlink	Data and control channels time multiplexed
Channel multiplexing in uplink	Control and pilot channel time multiplexed I&Q multiplexing for data and control channel
Multirate	Variable spreading and multicode
Spreading factors	4–256 (uplink), 4–512 (uplink)
Power control	Open and fast closed loop (1.6 kHz)
Spreading (downlink)	OVSF sequences for channel separation Gold sequences 2^{18}-1 for cell and user separation (truncated cycle 10 ms)
Spreading (uplink)	OVSF sequences, Gold sequence 2^{41} for user separation (different time shifts in I and Q channel, truncated cycle 10 ms)
Handover	Soft handover Interfrequency handover

222

WCDMA and GSM

Tables 15.2 and 15.3 respectively list the main differences between WCDMA and GSM, and between WCDMA and IS-95. note that only the air interface is considered for comparison.

Table 15.2: Main differences between WCDMA and GSM air interfaces.

	WCDMA	GSM
Carrier spacing	5 MHz	200 kHz
Frequency reuse factor	1	1–18
Power control frequency	1500 Hz	2 Hz or lower
Quality control	Radio resource management algorithms	Network planning (frequency planning)
Frequency diversity	5 MHz bandwidth gives multipath diversity with Rake receiver	Frequency hopping
Packet data	Load-based packet scheduling	Time slot based scheduling with GPRS
Downlink transmit diversity	Supported for improving downlink capacity	Not supported by the standard, but can be applied

WCDMA and IS-95

Table 15.3: Main differences between WCDMA and IS-95 air interfaces.

	WCDMA	IS-95
Carrier spacing	5 MHz	1.25 MHz
Chip rate	3.84 Mcps	1.2288 Mcps
Power control frequency	1500 Hz, both uplink and downlink	Uplink: 800 Hz, downlink: slow power control
Base station synchronisation	Not needed	Yes, typically obtained via GPS
Inter-frequency handovers	Yes, measurements with slotted mode	Possible, but measurement method not specified
Efficient radio resource management algorithms	Yes, provides required quality of service	Not needed for speech only networks
Packet data	Load-based packet scheduling	Packet data transmitted as short circuit switched calls
Downlink transmit diversity	Supported for improving downlink capacity	Not supported by the standard

Core Networks and 3G Air Interface Alternatives

There are three basic solutions for the core network to which WCDMA radio access networks can be connected. The basis of the second generation has been either the GSM core network or one based on IS-41. Both will naturally be important options in third generation systems. An emerging alternative is GPRS with an all-IP-based core network. The most typical connections between the core networks and the air interfaces are illustrated in Figure 15.3.

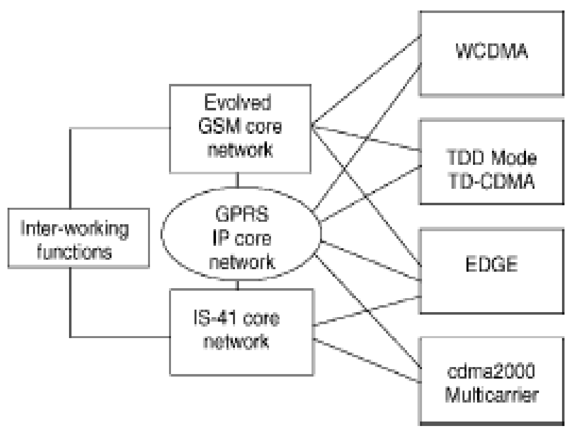


Figure 15.3: Core Networks and 3G air interface alternatives

15.3 FEATURES OF WCDMA

- WCDMA is a wideband Direct-Sequence Code Division Multiple Access (DS-CDMA) system, i.e. user information bits are spread over a wide bandwidth (by multiplying the user data with quasi-random bits (called chips) derived from CDMA spreading codes).
- WCDMA support data rate of up to 2 Mbps employing chip rate of 3.84 Mcps on 5 MHz (approx.) bandwidth. Carrier spacing can be selected on a 200 kHz grid (between approximately 4.4 and 5 MHz), depending on interference between the carriers.
- WCDMA supports highly variable user data rates: Bandwidth on Demand (BoD) on (fixed) 10 ms time frame.
- WCDMA supports both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) modes of operation. In the FDD mode, separate 5 MHz carrier frequencies are used for the uplink and downlink respectively, whereas in TDD only one 5 MHz is timeshared between the uplink and downlink.

- WCDMA supports the operation of asynchronous base stations, and hence deployment of indoor and micro base stations is easier when no GPS signal needs to be received.
- WCDMA employs coherent detection on uplink and downlink based on the use of pilot symbols or common pilot. The use of coherent detection on the uplink results in an overall increase of coverage and capacity on the uplink.
- WCDMA air interface is flexible with the provision of deploying advanced CDMA receiver (multi-user detection and smart adaptive antennas) to increase capacity and/or coverage.
- WCDMA is backward compatible with GSM and, therefore handovers between GSM and WCDMA are supported (in order to be able to leverage the GSM coverage for the introduction of WCDMA).

15.4 WCDMA: BANDWIDTH ALLOCATION

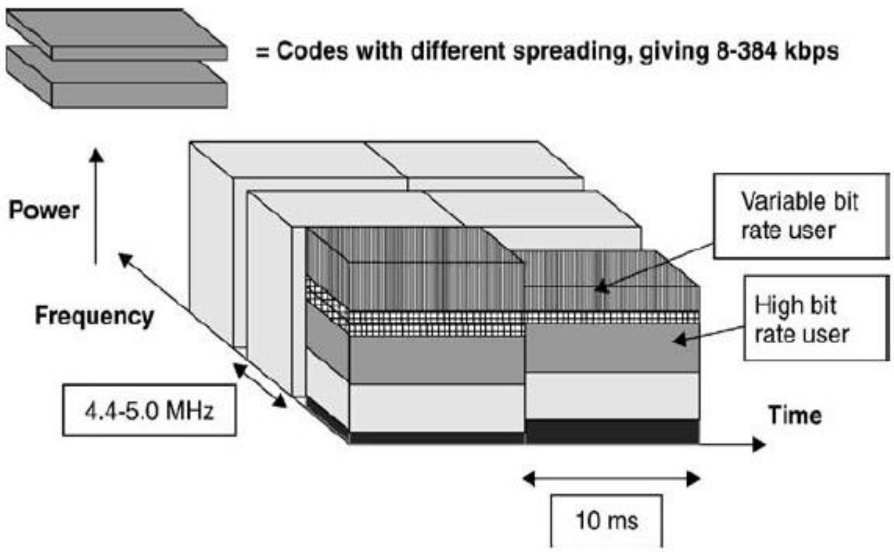


Figure 15.4: Allocation of Bandwidth in WCDMA (in time-frequency-code space).

15.5WCDMA: RADIO ACCESS NETWORK

Preface

This part named WCDMA RAN Part II gives a brief overview of the UMTS system architecture, including logical network elements and interfaces. Radio access network (RAN) as well as core network are described. General protocol model of the UMTS Terrestrial RAN (UTRAN) is shortly defined and their functionality is introduced. At the end of this part, students are expected to get an overall idea on WCDMA RAN from the system architecture view point. For Further detail, please consult with Ref. 1.

System Architecture: UMTS

The UMTS system consists of a number of logical network elements that each has a defined functionality. Functionally, the network elements are grouped into the Radio Access Network (RAN, UMTS Terrestrial RAN: UTRAN) that handles all radio-related functionality, and the Core Network, which is responsible for switching and routing calls and data connections to external networks. To complete the system, the User Equipment (UE) that interfaces with the user and the radio interface is defined. The high-level system architecture is shown in Figure PII 1. The corresponding network elements in each sub network are given in Figure PII 2. Following is a brief detail on all these elements. The functions of network elements are briefly noted in the following.

The UE consists of two parts.

- The Mobile Equipment (ME) is the radio terminal used for radio communication over the Uu interface.

- The UMTS Subscriber Identity Module (USIM) is a smartcard that holds the subscriber identity, performs authentication algorithms, and stores authentication and encryption keys, and some subscription information that is needed at the terminal.

Figure 15.5: UMTS high-level system architecture.

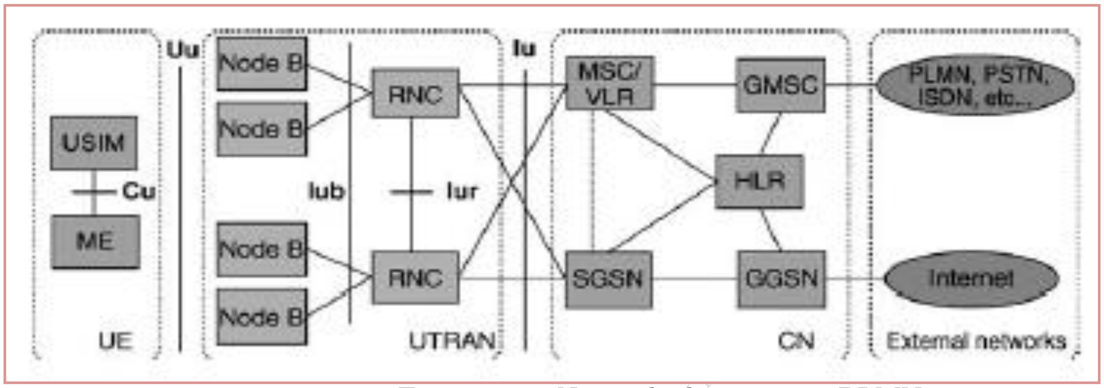


Figure 15.6: Network element in a PLMN.

UTRAN also consists of two distinct elements.

- The Node B converts the data flow between the Iub and Uu interfaces. It also participates in radio resource management.
- The Radio Network Controller (RNC) owns and controls the radio resources in its domain (the Node Bs connected to it). RNC is the service access point for all services UTRAN provides the CN, for example, management of connections to the UE.

The main elements of the core network (CN) and their functionalities are as follows.

HLR (Home Location Register) is a database located in the user's home system that stores the master copy of the user's service profile. The service profile consists of, for example, information on allowed services, forbidden roaming areas, and supplementary service information such as status of call forwarding and the call forwarding number. It is created when a new user subscribes to the system, and remains stored as long as the subscription is active.

MSC/VLR (Mobile Services Switching Centre/Visitor Location Register) is the switch (MSC) and database (VLR) that serves the UE in its current location for Circuit Switched (CS) services. The MSC function is used to switch the CS transactions, and the VLR function holds a copy of the visiting user's service profile, as well as more precise information on the UE's location within the serving system. The part of the network that is accessed via the MSC/VLR is often referred to as the CS domain.

GMSC (Gateway MSC) is the switch at the point where UMTS PLMN (Public Land Mobile Network: a sub-network with unique identities, typically one PLMN per operator) is connected to external CS networks. All incoming and outgoing CS connections go through GMSC.

SGSN (Serving GPRS (General Packet Radio Service) Support Node) functionality is similar to that of MSC/VLR but is typically used for Packet Switched (PS) services. The part of the network that is accessed via the SGSN is often referred to as the PS domain.

GGSN (Gateway GPRS Support Node) functionality is close to that of GMSC but is in relation to PS services.

The external networks can be divided into two groups.

CS networks: These provide circuit-switched connections, like the existing telephony service. PSTN is an example of CS networks (where two endpoints have dedicated connection during the period of communication).

PS networks: These provide connections for packet data services. The Internet is one example of a PS network (where data is divided to be transmitted into packets transmitted through the network independently).

The main UMTS open interfaces are as follows.

Cu interface: This is the electrical interface between the USIM smartcard and the ME. The interface follows a standard format for smartcards.
Uu interface: Uu is the interface through which the UE accesses the fixed part of the system, and is therefore probably the most important open interface in UMTS.

Iu interface: This connects UTRAN to the CN and gives UMTS operators the possibility of acquiring UTRAN and CN from different manufacturers.

Iur interface: The open Iur interface allows soft handover between RNCs from different Manufacturers.

Iub interface: The Iub connects a Node B and an RNC. UMTS is the first commercial mobile telephony system where the controller-base station interface is standardised as a fully open interface.

UTRAN Architecture

Figure 15.7 illustrates the UTRAN architecture.

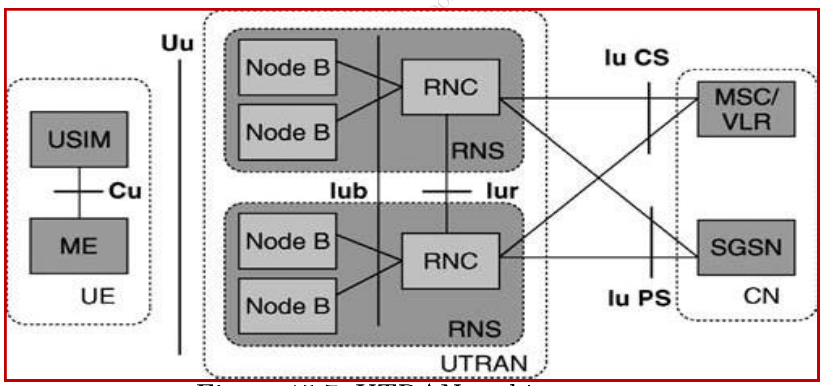


Figure 15.7: UTRAN architecture.

UTRAN consists of one or more Radio Network Sub-systems (RNS). An RNS is a sub-network within UTRAN and consists of one Radio Network Controller (RNC) and one or more Node Bs. RNCs may be connected to each other via an Iur interface. RNCs and NodeBs are connected with an Iub interface. The main characteristics of UTRAN which are also the main requirements for the design of the UTRAN architecture, functions and protocols are summarised in the followings.

- Support of UTRA and all the related functionality. In particular, the requirement to support soft handover (one terminal connected to the network via two or more active cells) and the WCDMA-specific Radio Resource Management algorithms.
- Maximisation of the commonalities in the handling of packet-switched and circuit switched data, with a unique air interface protocol stack and with the use of the same interface for the connection from UTRAN to both the PS and CS domains of the core network.
- Maximisation of the commonalities with GSM, when possible.
- Use of the ATM transport as the main transport mechanism in UTRAN.

- Use of the IP-based transport as the alternative transport mechanism in UTRAN from Release 5 and onwards.

A brief description of the UTRAN network elements are given below.

The Radio Network Controller

The RNC (Radio Network Controller) is the network element responsible for the control of the radio resources of UTRAN. It interfaces the CN (normally to one MSC and one SGSN) and also terminates the RRC (Radio Resource Control) protocol that defines the messages and procedures between the mobile and UTRAN.

The Node B (Base Station)
The main function of the Node B is to perform the air interface L1 processing (channel coding and interleaving, rate adaptation, spreading, etc.). It also performs some basic Radio Resource Management operations such as the inner loop power control

15.6 GENERAL PROTOCOL MODEL FOR UTRAN TERRESTRIAL INTERFACES

Principle

UTRAN terrestrial interfaces model is shown in Figure PII 4. The structure is based on the principle that the layers and planes are logically independent of each other and, if needed, parts of the protocol structure may be changed in the future while other parts remain intact.

Layers

The protocol structure consists of two main layers, the Radio Network Layer and the Transport Network Layer. All UTRAN-related issues are visible only in the Radio Network Layer, and the Transport Network Layer represents standard transport technology that is selected to be used for UTRAN but without any UTRAN-specific changes. 227

Planes

The Control Plane is used for all UMTS-specific control signalling. It includes the Application Protocol (i.e. RANAP in Iu, RNSAP in Iur and NBAP in Iub), and the Signalling Bearer for transporting the Application Protocol messages.

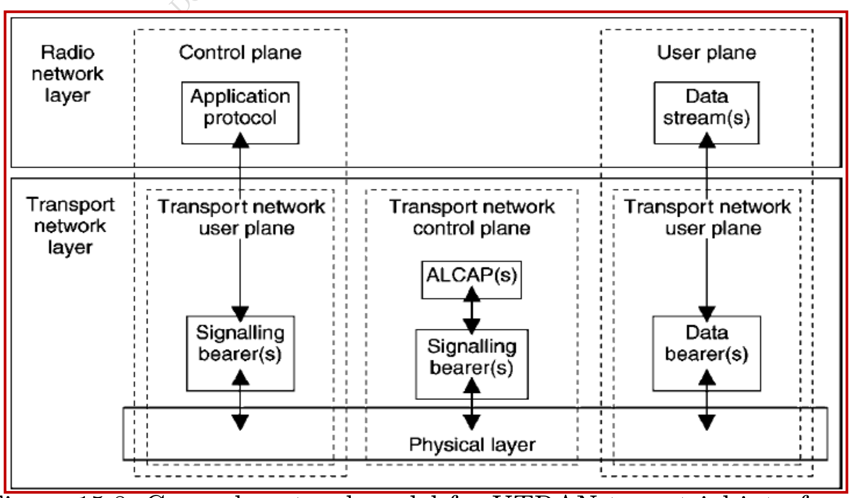


Figure 15.8: General protocol model for UTRAN terrestrial interfaces.

The Application Protocol is used, among other things, for setting up bearers to the UE (i.e. the Radio Access Bearer in Iu and subsequently the Radio Link in Iur and Iub). The Signalling Bearer for the Application Protocol may or may not be the same type as the Signalling Bearer for the ALCAP. It is always set up by O&M actions.

The User Plane includes the Data Stream(s), and the Data Bearer(s) for the Data Stream(s). Each Data Stream is characterised by one or more frame protocols specified for that interface. All information sent and received by the user, such as the coded voice in a voice call or the packets in an Internet connection, are transported via the User Plane.

The Transport Network Control Plane is used for all control signalling within the Transport Layer. It does not include any Radio Network Layer information. It includes the ALCAP protocol that is needed to set up the transport bearers (Data Bearer) for the User Plane. It also includes the Signalling Bearer needed for the ALCAP.

Transport Network Control Plane acts between the Control Plane and the User Plane. The introduction of the Transport Network Control Plane makes it possible for the Application Protocol in the Radio Network Control Plane to be completely independent of the technology selected for the Data Bearer in the User Plane. When the Transport Network Control Plane is used, the transport bearers for the Data Bearer in the User Plane are set up as accordingly: First there is a signalling transaction by the Application Protocol in the Control Plane, which triggers the set-up of the Data Bearer by the ALCAP protocol that is specific for the User Plane technology.
The Data Bearer(s) in the User Plane, and the Signalling Bearer(s) for the Application Protocol, also belong to the Transport Network User Plane. The Data Bearers in the Transport Network User Plane are directly controlled by the Transport Network Control Plane during real-time operation, but the control actions required for setting up the Signalling Bearer(s) for the Application Protocol are considered O&M actions.

15.7 UMTS CORE NETWORK ARCHITECTURE AND EVOLUTION

The UMTS core network has not been major changed in the 3GPP Release '99 specification. Basically, Release '99 structure is inherited from the GSM core network.

Release '99 Core Network Elements

The Release '99 core network has two domains: Circuit Switched (CS) domain and Packet Switched (PS) domain, to cover the need for different traffic types. The division comes from the different requirements for the data, depending on whether it is real time (circuit switched) or non-real time (packet data). Figure PII 5 illustrates the Release '99 core network structure. Registers, as well as the Service Control Point (SCP) are incorporated to indicate the link for providing a particular service to the end user.

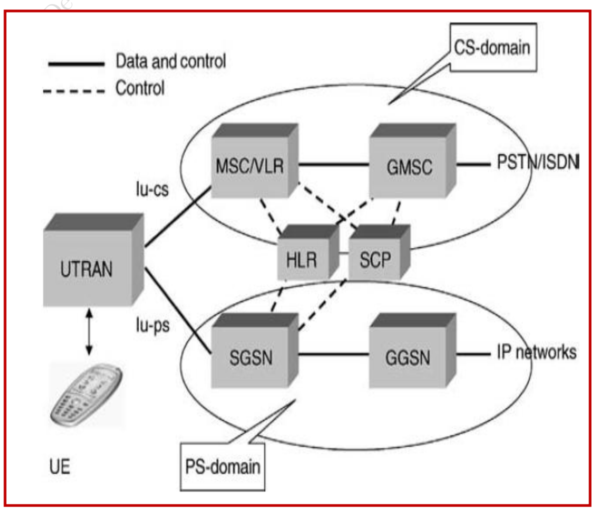


Figure 15.9: Release '99 UMTS core network structure.

The CS domain has the following elements.

- Mobile Switching Centre (MSC), including Visitor Location Register (VLR)
- Gateway MSC (GMSC).

The PS domain has the following elements.

- Serving GPRS Support Node (SGSN), which covers similar functions as the MSC for the packet data, including VLR type functionality.
- Gateway GPRS Support Node (GGSN) connects PS core network to other networks, for example to the Internet.
- In addition, the various registers are needed for the network proper operation such as Home Location Register (HLR); Equipment Identity Register (EIR): that contains the information related to the terminal equipment and can be used to, e.g., prevent a specific terminal from accessing the network.

15.8 WCDMA: PHYSICAL LAYER

Preface

This part named WCDMA RAN Part III describes (in brief) the major aspects of the WCDMA (UTRA FDD) physical layer. the transport channels together with their mapping to different physical channels, Uplink ands downlink multiplexing, transmitter characteristics, signalling and physical layer procedures are covered. For Further detail, please consult with Ref. 1.

Transport Channels and their Mapping to the Physical Channels

In UTRA the data generated at higher layers is carried over the air with transport channels, which are mapped in the physical layer to different physical channels. The physical layer is required to support variable bit rate transport channels to offer bandwidth-on-demand services, and to be able to multiplex several services to one connection.

Each transport channel is accompanied by the Transport Format Indicator (TFI) at each time event at which data is expected to arrive for the specific transport channel from the higher layers. The physical layer combines the TFI information from different transport channels to the Transport Format Combination Indicator (TFCI). The TFCI is transmitted in the physical control channel to inform the receiver which transport channels are active for the current frame; The TFCI is decoded appropriately in the receiver and the resulting TFI is given to higher layers for each of the transport channels that can be active for the connection. Figure PIII 1 shows mapping of two transport channels to a single physical channel, and also error indication is provided for each transport block.

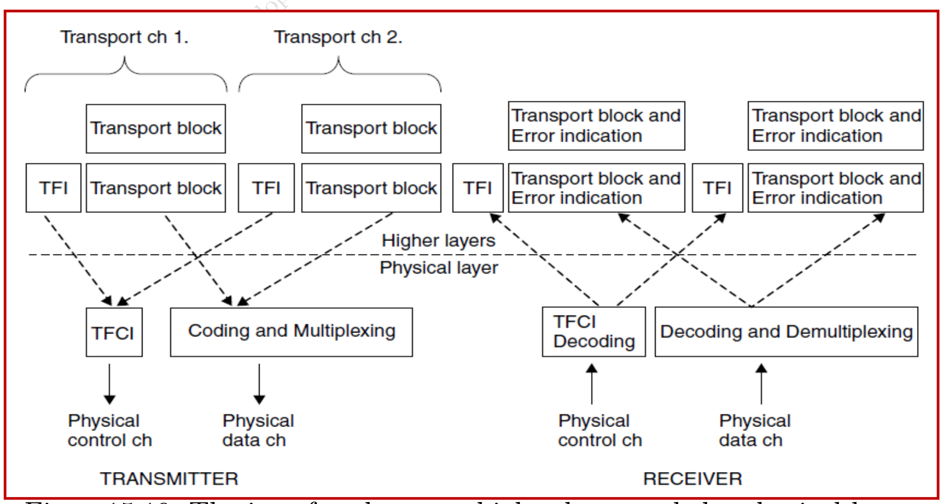


Figure15.10: The interface between higher layer and the physical layer.

One physical control channel and one or more physical data channels form a single Coded Composite Transport Channel (CCTrCh). There can be more than one CCTrCh on a given connection but only one physical layer control channel is transmitted in such a case.

Two types of transport channel exist: dedicated channels and common channels. The main difference between them is that a common channel is a resource divided between all or a group of users in a cell, whereas a dedicated channel resource, identified by a certain code on a certain frequency, is reserved for a single user only.

The only dedicated transport channel is the dedicated channel, for which the term DCH is used. The dedicated transport channel carries all the information intended for the given user coming from layers above the physical layer, including data for the actual service as well as higher layer control information. The content of the information carried on the DCH is not visible to the physical layer, thus higher layer control information and user data are treated in the same way.

The dedicated transport channel is characterised by features such as fast power control, fast data rate change on a frame-by-frame basis, and the possibility of transmission to a certain part of the cell or sector with varying antenna weights with adaptive antenna systems. The dedicated channel supports soft handover.

There are six different common transport channel types defined for UTRA. Common channels do not have soft handover but some of them can have fast power control.

The Broadcast Channel (BCH) is a transport channel that is used to transmit information specific to the UTRA network or for a given cell. The most typical data needed in every network is the available random access codes and access slots in the cell, or the types of transmit diversity method used with other channels for that cell. As the terminal cannot register to the cell without the possibility of decoding the broadcast channel, this channel is needed for transmission with relatively high power in order to reach all the users within the intended coverage area.

The Forward Access Channel (FACH) is a downlink transport channel that carries control information to terminals known to be located in the given cell. This is used, for example, after a random access message has been received by the base station.

The Paging Channel (PCH) is a downlink transport channel that carries data relevant to the paging procedure, that is, when the network wants to initiate communication with the terminal. The simplest example is a speech call to the terminal: the network transmits the paging message to the terminal on the paging channel of those cells belonging to the location area that the terminal is expected to be in. The terminals must be able to receive the paging information in the whole cell area.

The Random Access Channel (RACH) is an uplink transport channel intended to be used to carry control information from the terminal, such as requests to set up a connection. It can also be used to send small amounts of packet data from the terminal to the network. For proper system operation the random access channel must be heard from the whole desired cell coverage area,

The uplink common packet channel (CPCH) is an extension to the RACH channel that is intended to carry packet-based user data in the uplink direction. The reciprocal channel providing the data in the downlink direction is the FACH.

The downlink shared channel (DSCH) is a transport channel intended to carry dedicated user data and/or control information; it can be shared by several users The DSCH does not need to be heard in the whole cell area and can employ the different modes of transmit antenna diversity methods that are used with the associated downlink DCH.

The common transport channels needed for basic network operation are RACH, FACH and PCH, while the use of DSCH and CPCH is optional and can be decided by the network.

Mapping of Transport Channels onto the Physical Channels

The different transport channels are mapped to different physical channels, though some of the transport channels are carried by identical (or even the same) physical channel. The transport channel to physical channel mapping is illustrated in Figure PIII 2.

In addition to the transport channels, there exist physical channels to carry only information relevant to physical layer procedures. The Synchronisation Channel (SCH), the Common Pilot Channel (CPICH) and the Acquisition Indication Channel (AICH) are not directly visible to higher layers and are mandatory from the system function point of view, to be transmitted from every base station. The CPCH Status Indication Channel (CSICH) and the Collision Detection/Channel Assignment Indication Channel (CD/CA-ICH) are needed if CPCH is used.

A dedicated channel (DCH) is mapped onto two physical channels. The Dedicated Physical Data Channel (DPDCH) carries higher layer information, including user data, while the Dedicated Physical Control Channel (DPCCH) carries the necessary physical layer control information. These two dedicated physical channels are needed to support efficiently the variable bit rate in the physical layer. The bit rate of the DPCCH is constant, while the bit rate of DPDCH can change from frame to frame.

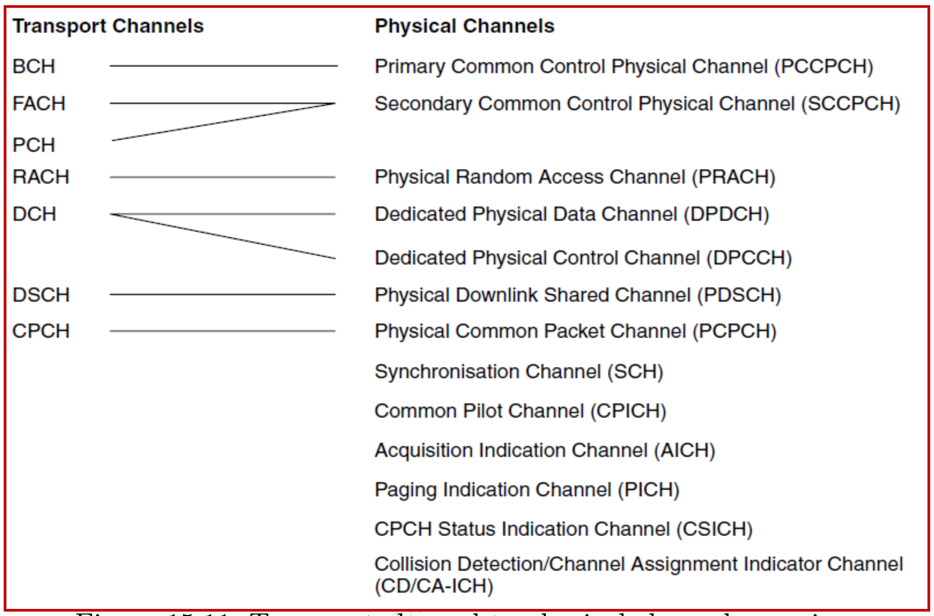


Figure 15.11: Transport channel to physical channel mapping.

UTRA channels use a 10 ms radio frame structure. The frame structure also employs a longer period, called the system frame period. The System Frame Number (SFN) is a 12-bit number and is used by procedures that span more than a single frame. Physical layer procedures, such as the paging procedure or random access procedure, are examples of procedures that need a longer period than 10 ms for correct definition.

Transmitter Characteristics: ACLR

The pulse shaping method applied to the transmitted symbols is root-raised cosine filtering with a roll-off factor of 0.22. The same roll-off is valid for both the terminals and the base stations. The nominal carrier spacing in WCDMA is 5 MHz but the carrier frequency in WCDMA can be adjusted with a 200 kHz raster. The central frequency of each WCDMA carrier is indicated with an accuracy of 200 kHz.

The Adjacent Channel Leakage Ratio (ACLR) determines how much of the transmitted power is allowed to leak into the first or second neighbouring carrier. The concept of ACLR is illustrated in Figure PIII 3, where ACLR1 and ACLR2 correspond to the power level integrated over the first and second adjacent carriers, with 5 MHz and 10 MHz carrier separation respectively. On the terminal side the ACLR values for the power classes of 21 dBm and 24 dBm have been set to 33 dB and 43 dB for ACLR1 and ACLR2 respectively. On the base station side the corresponding values are 45 dB and 50 dB.

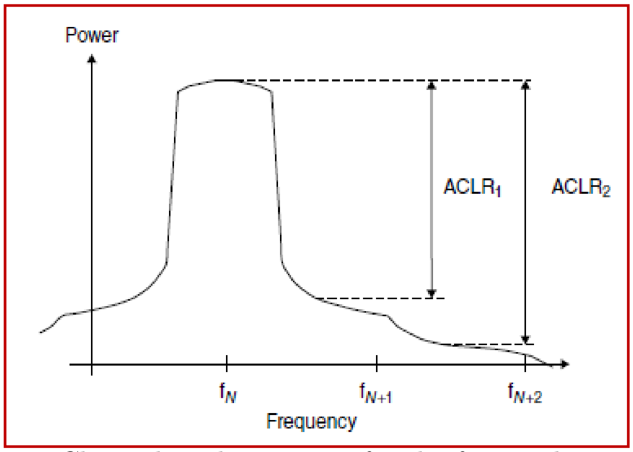


Figure 15.12: Adjacent Channel Leakage Ratio for the first and second adjacent carriers.

The frequency accuracy requirements are also directly related to the implementation cost, especially on the terminal side. The terminal frequency accuracy has been defined to be 0.1 ppm when compared to the received carrier frequency. On the base station side the requirement is tighter: 0.05 ppm.
The baseband timing is tied to the same timing reference as RF. The base station value needs to be tighter than the terminal value, since the base station carrier frequency is the reference for the terminal accuracy. The terminal needs also to be able to search the total frequency uncertainty area caused by the base station frequency error tolerance on top of the terminal tolerances and the error caused by terminal movement.

Multiplexing: Uplink

In the uplink direction the services are multiplexed dynamically so that the data stream is continuous with the exception of zero rate. The symbols on the DPDCH are sent with equal power level for all services, i.e. the service coding and channel multiplexing needs, in some cases, to adjust the relative symbol rates for different services in order to balance the power level requirements for the channel symbols. The rate matching function in the multiplexing chain can be used for such quality balancing operations between services on a single DPDCH. For the uplink DPDCH there do not exist fixed positions for different services, but the frame is filled according to the outcome of the rate matching and interleaving operation(s). The uplink multiplexing is done in 11 steps, as illustrated in Figure PIII 4.

After receiving a transport block from higher layers, (1) the first operation is CRC attachment. The CRC (Cyclic Redundancy Check) is used for error checking of the transport blocks at the receiving end. The CRC length that can be inserted has four different values: 0, 8, 12, 16 and 24 bits. The more bits the CRC contains, the lower is the probability of an undetected error in the transport block in the receiver. The physical layer provides the transport block to higher layers together with the error indication from the CRC check. (2) After the CRC attachment, the transport blocks are either concatenated together or segmented to different coding blocks. This depends on whether the transport block fits the available code block size as defined for the channel coding method. (3) The channel encoding is performed on the coding blocks after the concatenation or segmentation operation. (4) The function of radio frame equalisation is to ensure that data can be divided into equalized blocks when transmitted over more than a single 10 ms radio frame. This is done by padding the necessary number of bits until the data can be in equal-sized blocks per frame.

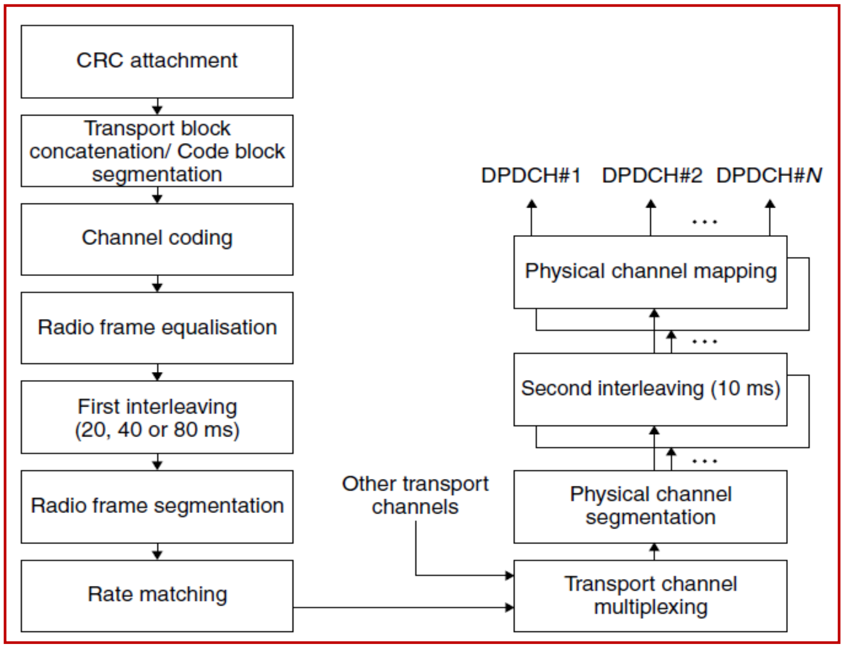


Figure 15.13: Uplink multiplexing and channel coding chain.

(5) The first interleaving or inter-frame interleaving is used when the delay budget allows more than 10 ms of interleaving. The interlayer length of the first interleaving has been defined to be 20, 40 and 80 ms. The interleaving period is directly related to the Transmission Time Interval (TTI), which indicates how often data arrives from higher layers to the physical layer. (6) The timing relation with different TTIs is illustrated in Figure PIII 5. If the first interleaving is used, the frame segmentation will distribute the data coming from the first interleaving over two, four or eight consecutive frames in line with the interleaving length. (7) Rate matching is used to match the number of bits to be transmitted to the number available on a single frame. This is achieved either by puncturing or by repetition. When the data rate of the service with lowest TTI varies, as in Figure PIII 5, the dynamic rate matching adjusts the rate matching parameters for other transport channels as well, so that all the symbols in the radio frame are used.

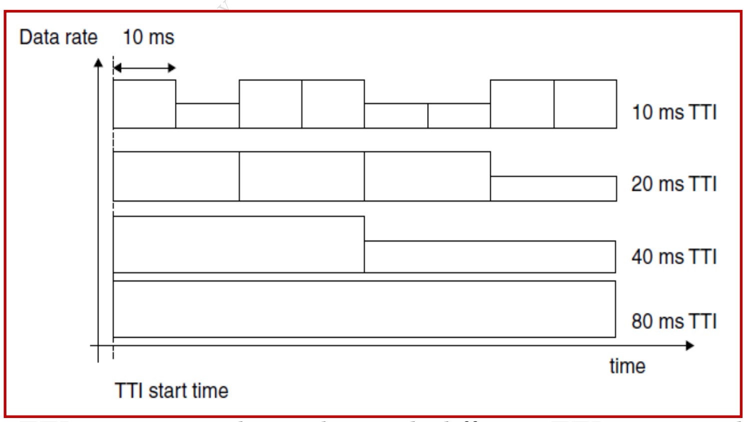


Figure 15.14: TTI start time relationship with different TTIs on a single connection

(8) The different transport channels are multiplexed together by the transport channel multiplexing operation. This is a simple serial multiplexing on a frame-by-frame basis. Each transport channel provides data in 10 ms blocks for this multiplexing. (9) In case more than one physical channel (spreading code) is used, physical channel segmentation is used. This operation simply divides the data evenly on the available spreading codes, as currently no cases have been specified where the spreading factors would be different in multi-code transmissions. (10) The second interleaving performs 10 ms radio frame interleaving, sometimes called intraframe interleaving. This is a block interleaver with inter-column permutations applied to the 30 columns of the interleaver. (11) From the output of the second interleaver the bits are mapped on the physical channels. The number of bits given for a physical channel at this stage is exactly the number that

the spreading factor of that frame can transmit. Alternatively, the number of bits to transmit is zero and the physical channel is not transmitted at all.

Multiplexing: Downlink

As in the uplink, the interleaving is implemented in two parts, covering both intra-frame and inter-frame interleaving. Also the rate matching allows one to balance the required channel symbol energy for different service qualities. However, There are differences in the order in which rate matching and segmentation functions are performed (Figure PIII 6). Whether fixed or flexible bit positions are used determines the DTX indication insertion point. The DTX indication bits are not transmitted over the air; they are just inserted to inform the transmitter at which bit positions the transmission should be turned off.

The use of fixed positions means that for a given transport channel, the same symbols are always used. If the transmission rate is below the maximum, then DTX indication bits are used for those symbols. With flexible positions the situation is different since now the channel bits unused by one service may be utilised by another. This is useful when it is possible to have such a transport channel combination that they do not all need to be able to reach the full data rate simultaneously, but can alternate with the need for full rate transmission. This allows the necessary spreading code occupancy in the downlink to be reduced (Figure PIII 7).

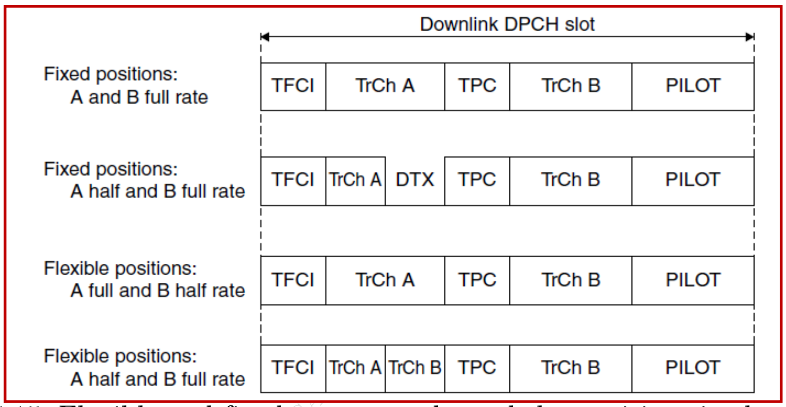


Figure 15.15: Flexible and fixed transport channel slot positions in the downlink.

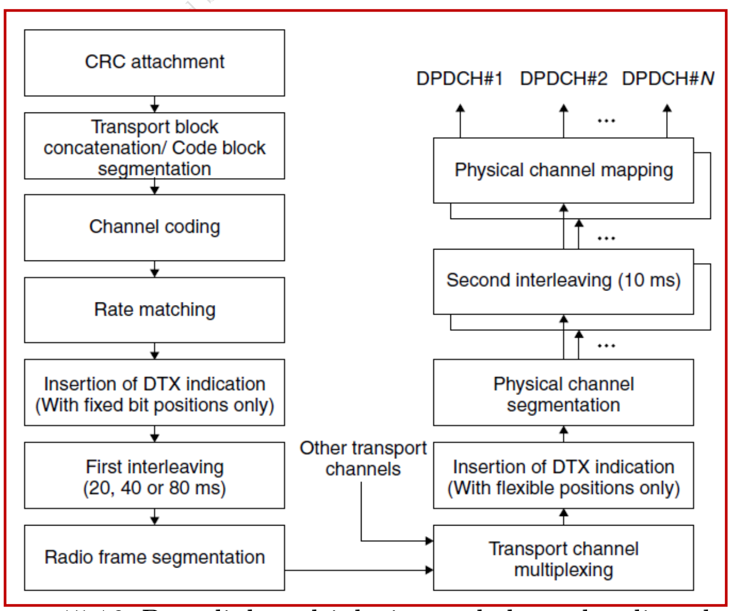


Figure 15.16: Downlink multiplexing and channel coding chain.

Signaling

234

For signaling purposes a lot of information needs to be transmitted between the network and the terminals. methods for transmitting signaling messages generated in the physical layer as well as the physical layer control channels for system operation are shortly described in the followings.

Common Pilot Channel (CPICH): The common pilot channel is an unmodulated code channel, which is scrambled with the cell-specific primary scrambling code. The function of the CPICH is to aid the channel estimation at the terminal for the dedicated channel and to provide the channel estimation reference for the common channels when they are not associated with the dedicated channels or not involved in the adaptive antenna techniques. UTRA has two types of common pilot channel, primary and secondary. the Primary CPICH is always under the primary scrambling code with a fixed channelization code allocation and there is only one such channel for a cell or sector. the primary common pilot channel can be used for the measurements for the handover and cell selection/ reselection.

Secondary CPICH may have any channelization code of length 256 and may be under a secondary scrambling code as well. The typical area of Secondary CPICH usage would be operations with narrow antenna beams intended for service provision at specific 'hot spots' or places with high traffic density.

Synchronization Channel (SCH): Synchronization Channel (SCH) is needed for the cell search. It consists of two channels, the primary and secondary synchronization channels. The Primary SCH uses a 256-chip spreading sequence identical in every cell. The system wide sequence has been optimized for matched filter implementations The Secondary SCH uses sequences with different code word combination possibilities representing different code groups. Once the terminal has identified the secondary synchronization channel, it has obtained frame and slot synchronization as well as information on the group the cell belongs to.

Primary Common Control Physical Channel (Primary CCPCH):The Primary Common Control Physical Channel (Primary CCPCH) is the physical channel carrying the Broadcast Channel (BCH). It needs to be demodulated by all the terminals in the system. As a result, the parameters with respect to, for example, the channel coding and spreading code contain no flexibility, as they need to be known by all terminals

Secondary Common Control Physical Channel (Secondary CCPCH): The Secondary Common Control Physical Channel (Secondary CCPCH) carries two different common transport channels: the Forward Access Channel (FACH) and the Paging Channel (PCH). The two channels can share a single Secondary CCPCH or can use different physical channels. This means that in the minimum configuration each cell has at least one Secondary CCPCH.

Random Access Channel (RACH): The Random Access Channel (RACH) is typically used for signalling purposes, to register the terminal after power-on to the network or to perform location update after moving from one location area to another or to initiate a call.

- Acquisition Indicator Channel (AICH): the Acquisition Indicator Channel (AICH) is used to indicate from the base station the reception of the random access channel signature sequence in connection with the Random Access Channel. The AICH uses an identical signature sequence as the RACH on one of the downlink channelisation codes of the base station to which the RACH belongs. Once the base station has detected the preamble with the random access attempt, then the same signature sequence that has been used on the preamble will be echoed back on AICH.
- Paging Indicator Channel (PICH): The Paging Channel (PCH) is operated together with the Paging Indicator Channel (PICH) to provide terminals with efficient sleep mode operation. The paging indicators use a channelisation code of length 256. The paging indicators occur once per slot on the corresponding physical channel, the Paging Indicator Channel (PICH).
- Physical Channels for the CPCH Access Procedure: For the CPCH access procedure, a set of CPCH specific physical channels has been specified.

These channels carry no transport channels, but only information needed in the CPCH access procedure. The channels are:

- CPCH Status Indication Channel (CSICH):
- CPCH Collision Detection Indicator Channel (CD-ICH);
- CPCH Channel Assignment Indicator Channel (CA-ICH);
- CPCH Access Preamble Acquisition Channel (AP-AICH).

CPCH Status Indication Channel (CSICH) uses the part of the AICH channel that is defined as unused. CSICH bits indicate the availability of each physical CPCH channel and are used to tell the terminal to initiate access only on a free channel but, on the other hand, to accept a channel assignment command to an unused channel. The CD-ICH carries the collision detection information to the terminal. When the CAICH channel is used, the CD-ICH and CA-ICH are sent in parallel to the terminal. Both have 16 different bit patterns specified. The AP-AICH is identical to the AICH used with RACH and may share the same channelization code when sharing access resources with RACH.

15.8 PHYSICAL LAYER PROCEDURES

In the physical layer of a CDMA system there are many procedures essential for system operation such as fast power control, random access, paging, handover measurements and operation with transmit diversity. In the following, these procedures are briefly overviewed.

Fast Closed Loop Power Control Procedure: The fast power control operation operates on a basis of one command per slot, resulting in a 1500 Hz command rate. The basic step size is 1 dB. Additionally, multiples of that step size can be used and smaller step sizes can be emulated. The emulated step size means that the 1 dB step is used, for example, only every second slot, thus emulating the 0.5 dB step size. The specifications define the relative accuracy for a 1 dB power control step to be 0.5 dB. Fast power control operation has two special cases: operation with soft handover and with compressed mode in connection with handover measurements.

In soft handover the main issue for terminals is how to react to multiple power control commands from several sources that has been solved by specifying the operation such that the terminal combines the commands but also takes the reliability of each individual command decision into account (in deciding whether to increase or decrease the power). In the compressed mode case, the fast power control uses a larger step size for a short period after a compressed frame. This allows the power level to converge more quickly to the correct value after a break in the control stream.

Open Loop Power Control: In UTRA FDD open loop power control is applied only prior to initiating the transmission on the RACH or CPCH. Open loop power control is not very accurate, since it is difficult to measure large power dynamics accurately in the terminal equipment. The requirement for open loop power control accuracy is specified to be within 9 dB in normal conditions

Paging Procedure: The Paging Channel (PCH) operation is organised as follows. A terminal, once registered to a network, has been allocated a paging group. For the paging group there are Paging Indicators (PI) which appear periodically on the Paging Indicator Channel (PICH) when there are paging messages for any of the terminals belonging to that paging group. Once a PI has been detected, the terminal decodes the next PCH frame transmitted on the Secondary CCPCH to see whether there was a paging message intended for it. The terminal may also need to decode the PCH in case the PI reception indicates low reliability of the decision (Figure 15.17).

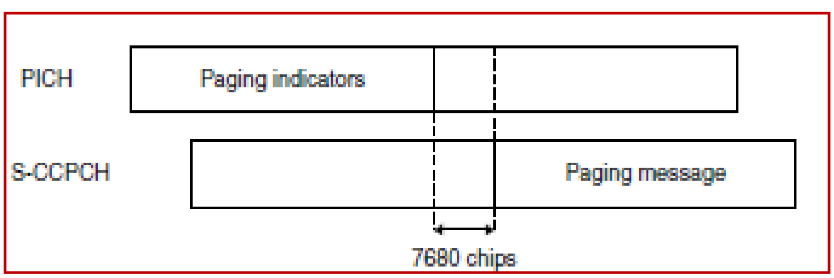


Figure 15.17: PICH relationship to PCH.

RACH Procedure: The Random Access procedure in a CDMA system has to cope with the near–far problem, as when initiating the transmission there is no exact knowledge of the required transmission power. In UTRA the RACH procedure has the following phases:

- The terminal decodes the BCH to find out the available RACH sub-channels and their scrambling codes and signatures.
- The terminal selects randomly one of the RACH sub-channels from the group its access class allows it to use. Furthermore, the signature is also selected randomly from among the available signatures.

- The downlink power level is measured and the initial RACH power level is set with the proper margin due to the open loop inaccuracy.
- A 1-ms RACH preamble is sent with the selected signature.
- The terminal decodes AICH to see whether the base station has detected the preamble.
- In case no AICH is detected, the terminal increases the preamble transmission power by a step given by the base station, as multiples of 1 dB. The preamble is retransmitted in the next available access slot.
- When an AICH transmission is detected from the base station, the terminal transmits the 10 ms or 20 ms message part of the RACH transmission.

The RACH procedure is illustrated in Figure PIII 9, where the terminal transmits the preamble until acknowledgement is received on AICH, and then the message part follows.

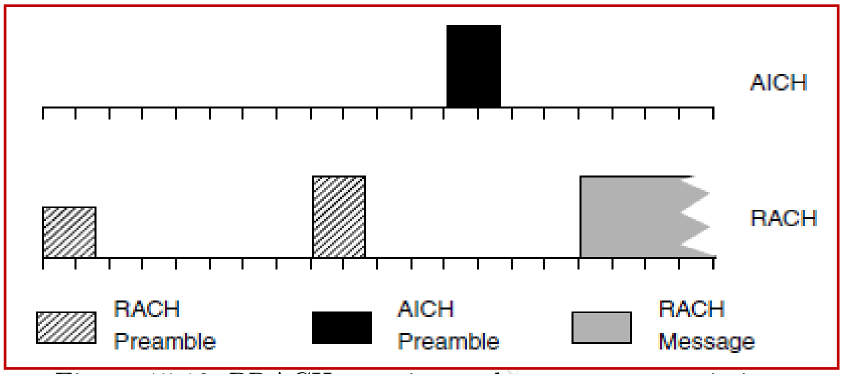


Figure 15.18: PRACH ramping and message transmission.

CPCH Operation: The operation follows the RACH procedure until the terminal detects AICH. After that a Collision Detection (CD) preamble with the same power level is still sent back with another signature, randomly chosen from a given set. Then the base station is expected to echo this signature back to the terminal on the CD Indication Channel (CD-ICH) and in this way to create a method of reducing the collision probability on Layer 1. After the correct preamble has been sent by the base station on the collision detection procedure, the terminal starts the transmission, which may last over several frames. Uplink Common Packet Channel (CPCH) operation is illustrated in Figure PIII 10.

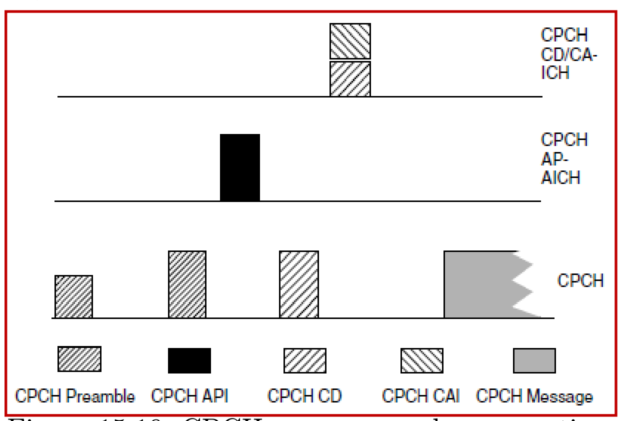


Figure 15.19: CPCH access procedure operation.

Cell Search Procedure: The cell search procedure using the synchronisation channel has basically three steps which are typically as follows.

- The terminal searches the 256-chip primary synchronisation code, being identical for all cells. As the primary synchronisation code is the same in every slot, the peak detected corresponds to the slot boundary.

- Based on the peaks detected for the primary synchronisation code, the terminal seeks the largest peak from the Secondary SCH code word. There are 64 possibilities for the secondary synchronisation code word. The terminal needs to check all 15 positions, as the frame boundary is not available before Secondary SCH code word detection.
- Once the Secondary SCH code word has been detected, the frame timing is known. The terminal then seeks the primary scrambling codes that belong to that particular code group. Each group consists of eight primary scrambling codes. These need to be tested for a single position only, as the starting point is known already.

Transmit Diversity Procedure: UTRA uses two types of transmit diversity transmission for user data performance improvement. These methods are classified as open loop and closed loop methods. In the case of closed loop transmit diversity, the base station uses two antennas to transmit the user information. The use of these two antennas is based on the feedback from the terminal, transmitted in the Feedback (FB) bits in the uplink DPCCH. The closed loop transmit diversity itself has two modes of operation. In mode 1, the terminal feedback commands control the phase adjustments that are expected to maximise the power received by the terminal. The base station thus maintains the phase with antenna 1 and then adjusts the phase of antenna 2 based on the sliding averaging over two consecutive feedback commands. With this method, four different phase settings are applied to antenna 2. In mode 2, the amplitude is adjusted in addition to the phase adjustment. The same signalling rate is used, but now the command is spread over four bits in four uplink DPCCH slots, with a single bit for amplitude and three bits for phase adjustment

The amplitude values have been defined to be 0.2 and 0.8, while the phase values are naturally distributed evenly for the antenna phase offsets, from 135 to þ180 phase offset.

Handover Measurements Procedure: Within the UTRA FDD the possible handovers are as follows:

- Intra-mode handover, which can be soft handover, softer handover or hard handover. Hard handover may take place as intra- or inter-frequency handover. The UTRA FDD intra-mode handover relies on the Ec/N0 measurement performed from the common pilot channel (CPICH).
- Inter-mode handover as handover to the UTRA TDD mode: On request from UTRAN, the dual-mode FDD–TDD terminals operating in FDD measure the power level from the TDD cells available in the area. 238
- Inter-system handover, which in Release '99 means only GSM handover. The GSM handover may take place to a GSM system operating at 850 MHz, 900 MHz, 1800 MHz and 1900 MHz.

Compressed Mode Measurement Procedure: The compressed mode, often referred to as the slotted mode, is needed when making measurements from another frequency in a CDMA system without a full dual receiver terminal. The compressed mode means that transmission and reception are halted for a short time, in the order of a few milliseconds, in order to perform measurements on the other frequencies. The intention is not to lose data but to compress the data transmission in the time domain.

Other Measurements: In the base station other measurements are needed to give RNC sufficient information on uplink status and base station transmission power resource usage. The following have been specified for the base station, to be supported by signalling between base station and RNC.

- RSSI, to give information on the uplink load;
- Uplink SIR on the DPCCH;
- Total transmission power on a single carrier at a base station transmitter, giving information on the available power resources at the base station;
- The transmission code on a single code for one terminal. This is used, for example, in balancing power between radio links in soft handover;
- Block Error Rate (BLER) and Bit Error Rate (BER) estimates for different physical channels.

Operation with Adaptive Antennas: UTRA has been designed to allow the use of adaptive antennas, also known as beamforming, in both the uplink and downlink directions. Basically there are two types of beamforming one may use. Either a beam may use the secondary common pilot channel (S-CPICH) or then a beam may use only the dedicated pilot symbols. What kind of beamforming may be applied with different channels depends on whether the channel contains dedicated pilot symbols or not.

A summary of the use of beamforming on different downlink channel types is given in Table PIII 1. Figure PIII 11 shows Release 6 beamforming enhancement method.

Table: Application of beamforming concepts on downlink physical channel types.

Physical channel type	Beamforming with S-CPICH	Beamforming without S-CPICH
P-CCPCH	No	No
SCH	No	No
S-CCPCH	No	No
DPCH	Yes	Yes
PICH	No	No
PDSCH, HS-PDSCH and HS-SCCH (with associated DPCH)	Yes	Yes[1]
AICH	No	No
CSICH	No	No

[1]UE capability with HSDPA

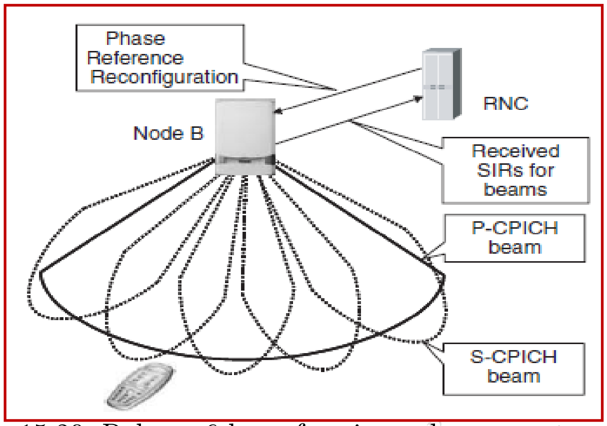


Figure 15.20: Release 6 beamforming enhancement method.

15.9 WCDMA: RADIO INTERFACE PROTOCOLS

Preface

This part named WCDMA Part IV Radio Interface Protocols describes the general radio interface protocol architecture is described. For each protocol, the logical architecture and main functions are described. In the MAC section, the logical channels (services offered by MAC) and mapping between logical channels and transport channels are also explained. In the RRC section, the RRC service states are described, together with the main (RRC) functions and signaling procedures. For Further detail, please consult with Ref. 1.

Introduction

The radio interface protocols are needed to set up, reconfigure and release the Radio Bearer services. The protocol layers above the physical layer are called the data link layer (Layer 2) and the network layer (Layer 3). In the UTRA FDD radio interface, Layer 2 is split into sub layers. In the control plane, Layer 2 contains two sub-layers – Medium Access Control (MAC) protocol and Radio Link Control (RLC) protocol. In the user plane, in addition to MAC and RLC, two additional service-dependent protocols exist: Packet Data Convergence Protocol (PDCP) and Broadcast/Multicast Control Protocol (BMC). Layer 3 consists of one protocol, called Radio Resource Control (RRC), which belongs to the control plane. The other network layer protocols, such as Call Control, Mobility Management, Short Message Service, and so on.

Protocol Architecture

- The overall radio interface protocol architecture is illustrated in Figure 15.21. which mainly contains only the protocols that are visible in UTRAN. The physical layer offers services to the MAC layer via transport channels that were characterized by how and with what characteristics data is transferred.
- The MAC layer, in turn, offers services to the RLC layer by means of logical channels. The logical channels are characterized by what type of data is transmitted.

- The RLC layer offers services to higher layers via service access points (SAPs), which describe how the RLC layer handles the data packets and if, for example, the automatic repeat request (ARQ) function is used. On the control plane, the RLC services are used by the RRC layer for signaling transport. On the user plane, the RLC services are used either by the service-specific protocol layers PDCP or BMC or by other higher-layer u-plane functions (e.g. speech codec). The RLC services are called Signaling Radio Bearers in the control plane and Radio Bearers in the user plane for services not utilizing the PDCP or BMC protocols.

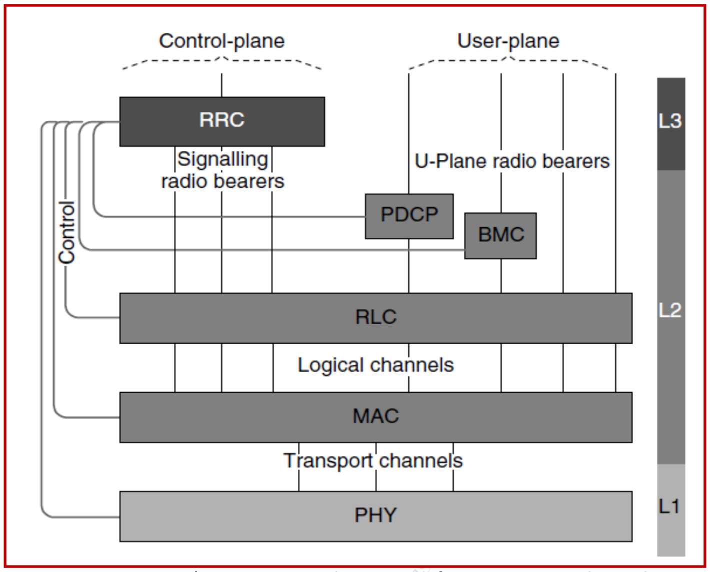


Figure 15.21: UTRA FDD Radio Interface protocol architecture.

- The Packet Data Convergence Protocol (PDCP) exists only for the PS domain services. Its main function is header compression. Services offered by PDCP are called Radio Bearers.
- The Broadcast Multicast Control protocol (BMC) is used to convey over the radio interface messages originating from the Cell Broadcast Centre. The service offered by BMC protocol is also called a Radio Bearer.
- The RRC layer offers services to higher layers (to the Non-Access Stratum) via service access points, which are used by the higher layer protocols in the UE side and by the Iu RANAP protocol in the UTRAN side. All higher layer signaling (mobility management, call control, session management, and so on) is encapsulated into RRC messages for transmission over the radio interface.
- The control interfaces between the RRC and all the lower layer protocols are used by the RRC layer to configure characteristics of the lower layer protocol entities, including parameters for the physical, transport and logical channels.

The Medium Access Control Protocol

In the Medium Access Control (MAC) layer [3] the logical channels are mapped to the transport channels. The MAC layer is also responsible for selecting an appropriate transport format (TF) for each transport channel depending on the instantaneous source rate(s) of the logical channels. The transport format is selected with respect to the transport format combination set (TFCS) which is defined by the admission control for each connection.

MAC Layer Architecture

The MAC layer logical architecture is shown in 15.22. The MAC layer consists of three logical entities:

MAC-b handles the broadcast channel (BCH). There is one MAC-b entity in each UE and one MAC-b in the UTRAN (located in Node B) for each cell.

MAC-c/sh handles the common channels and shared channels – paging channel (PCH), forward link access channel (FACH), random access channel (RACH), uplink Common Packet Channel (CPCH) and Downlink Shared.

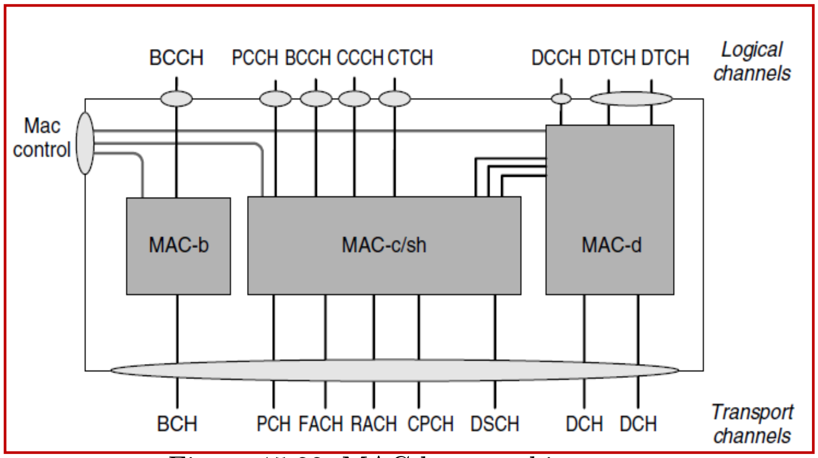


Figure 15.22: MAC layer architecture.

Channel (DSCH). There is one MAC-c/sh entity in each UE that is using shared channel(s) and one MAC-c/sh in the UTRAN (located in the controlling RNC) for each cell.

MAC-d is responsible for handling dedicated channels (DCH) allocated to a UE in connected mode. There is one MAC-d entity in the UE and one MAC-d entity in the UTRAN (in the serving RNC) for each UE.

MAC Functions

- Mapping between logical channels and transport channels.
- Selection of appropriate Transport Format (from the Transport Format Combination Set) for each Transport Channel, depending on the instantaneous source rate.
- Priority handling between data flows of one UE.
- Priority handling between UEs by means of dynamic scheduling.
- Identification of UEs on common transport channels
- Multiplexing/demultiplexing of higher layer PDUs into/from transport blocks delivered to/from the physical layer on common transport channels.
- Multiplexing/demultiplexing of higher layer PDUs into/from transport block sets delivered to/from the physical layer on dedicated transport channels.
- Traffic volume monitoring.
- Dynamic Transport Channel type switching.
- Ciphering.
- Access Service Class (ASC) selection for RACH transmission.

MAC Layer Logical Channels

The data transfer services of the MAC layer are provided on logical channels. A set of logical channel types is defined for the different kinds of data transfer service offered by MAC. A general classification of logical channels is into two groups: Control Channels and Traffic Channels. Control Channels are used to transfer control plane information, and Traffic Channels for user plane information. The Control Channels are:

- Broadcast Control Channel (BCCH). A downlink channel for broadcasting system control information.
- Paging Control Channel (PCCH). A downlink channel that transfers paging information.
- Dedicated Control Channel (DCCH). A point-to-point bidirectional channel that transmits dedicated control information between a UE and the RNC. This channel is established during the RRC connection establishment procedure.

- Common Control Channel (CCCH). A bidirectional channel for transmitting control information between the network and Ues.
- The Traffic Channels are:
- Dedicated Traffic Channel (DTCH). A Dedicated Traffic Channel (DTCH) is a point-to-point channel, dedicated to one UE, for the transfer of user information. A DTCH can exist in both uplink and downlink.
- Common Traffic Channel (CTCH). A point-to-multipoint downlink channel for transfer of dedicated user information for all, or a group of specified, UEs.

Mapping Between Logical Channels and Transport Channels

The mapping between logical channels and transport channels is shown in Figure 15.23. The following connections between logical channels and transport channels exist.

- PCCH is connected to PCH.
- BCCH is connected to BCH and may also be connected to FACH.
- DCCH and DTCH can be connected to either RACH and FACH, to CPCH and FACH, to RACH and DSCH, to DCH and DSCH, or to a DCH and DCH.
- CCCH is connected to RACH and FACH.
- CTCH is connected to FACH.

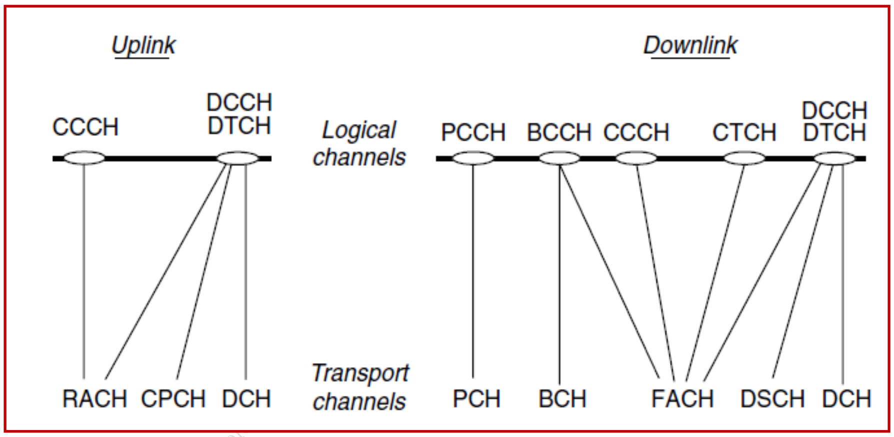


Figure 15.23: Mapping between logical channels and transport channels, uplink and downlink directions.

The Radio Link Control Protocol

The radio link control protocol provides segmentation and retransmission services for both user and control data. Each RLC instance is configured by RRC to operate in one of three modes: transparent mode (Tr), unacknowledged mode (UM) or acknowledged mode (AM). The service the RLC layer provides in the control plane is called Signalling Radio Bearer (SRB). In the user plane, the service provided by the RLC layer is called a Radio Bearer (RB) only if the PDCP and BMC protocols are not used by that service, otherwise the RB service is provided by the PDCP or BMC.

RLC Layer Architecture

The RLC layer architecture is shown in Figure 15.24. Note that the transparent and unacknowledged mode RLC entities are defined to be unidirectional, whereas the acknowledged mode entities are described as bidirectional.

In transparent mode no protocol overhead is added to higher layer data. Erroneous protocol data units (PDUs) can be discarded or marked erroneous. Transmission can be of the streaming type, in which higher layer data is not segmented, though in special cases transmission with limited segmentation/reassembly capability can be accomplished.

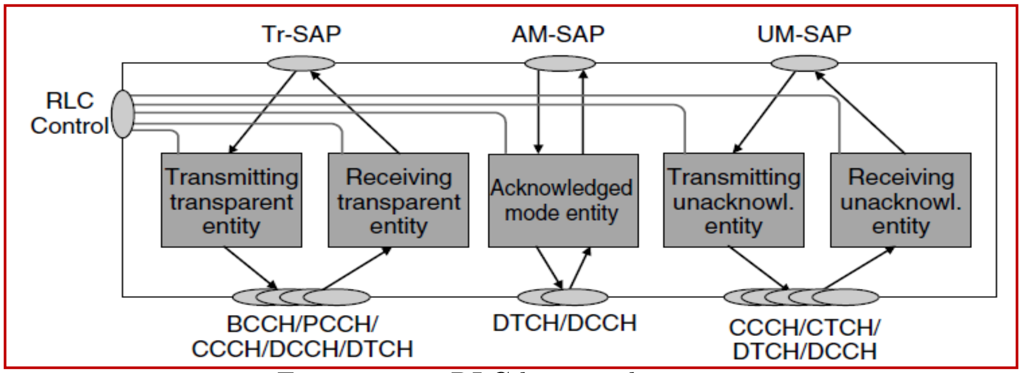


Figure 15.24: RLC layer architecture.

In unacknowledged mode no retransmission protocol is in use and data delivery is not guaranteed. Received erroneous data is either marked or discarded depending on the configuration.

In the acknowledged mode an automatic repeat request (ARQ) mechanism is used for error correction. The quality vs. delay performance of the RLC can be controlled by RRC through configuration of the number of retransmissions provided by RLC. In case RLC is unable to deliver the data correctly (max number of retransmissions reached or the transmission time exceeded), the upper layer is notified and the RLC SDU is discarded.
RLC Functions

- Segmentation and reassembly
- Concatenation
- Padding
- Transfer of user data
- Errorcorrection

In-sequence delivery of higher layer PDUs

- Duplicate detection.
- Flow control
- Sequence number check (Unacknowledged data transfer mode).
- Protocol error detection and recovery
- Ciphering
- Suspend/resume function for data transfer

The Packet Data Convergence Protocol

The Packet Data Convergence Protocol (PDCP) exists only in the user plane and only for services from the PS domain. The PDCP contains compression methods, which are needed to get better spectral efficiency for services requiring IP packets to be transmitted over the radio. As an example of why header compression is valuable, the size of the combined RTP/UDP/IP headers is at least 40 bytes for IPv4 and at least 60 bytes for IPv6, while the payload, for example for IP voice service, can be about 20 bytes or less.

PDCP Layer Architecture

Every PDCP entity uses zero, one or several header compression algorithm types with a set of configurable parameters. Several PDCP entities may use the same algorithm types. The algorithm types and their parameters are negotiated during the RRC Radio Bearer establishment or reconfiguration procedures and indicated to the PDCP through the PDCP Control Service Access Point (Figure 15.25).

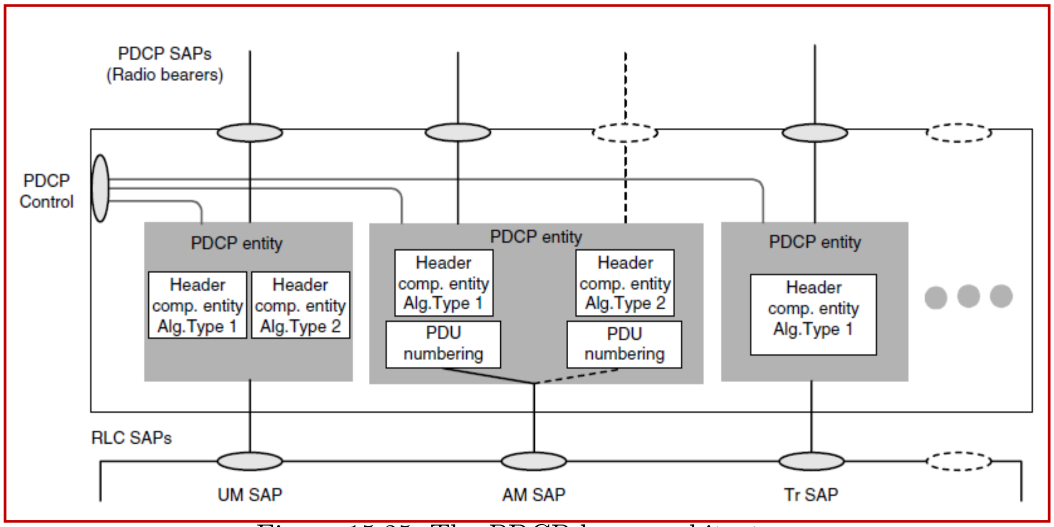


Figure 15.25: The PDCP layer architecture

PDCP Functions

- Compression of redundant protocol control information (e.g. TCP/IP and RTP/UDP/IP headers) at the transmitting entity, and decompression at the receiving entity
- Transfer of user data
- Support for lossless SRNS relocation

The Broadcast/Multicast Control Protocol

The Broadcast/Multicast Control (BMC) protocol – exists also only in the user plane. This protocol is designed to adapt broadcast and multicast services, originating from the Broadcast domain, on the radio interface. The BMC protocol, shown in Figure 15.26, does not have any special logical architecture. 244

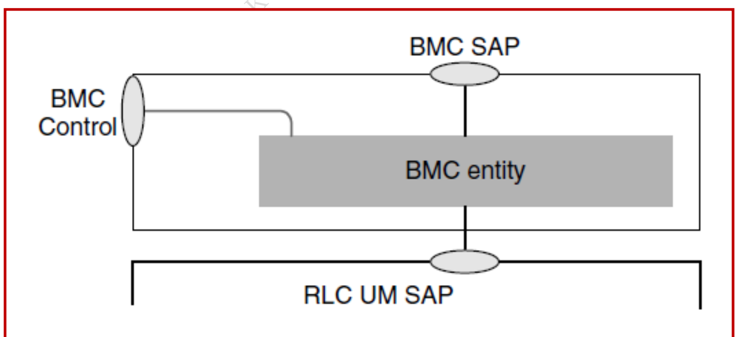


Figure 15.26: The Broadcast/Multicast Control layer architecture.

BMC Functions

- Storage of Cell Broadcast messages.
- Traffic volume monitoring and radio resource request for CBS.
- Scheduling of BMC messages
- Transmission of BMC messages to UE.
- Delivery of Cell Broadcast messages to the upper layer

The Radio Resource Control Protocol

The major part of the control signaling between UE and UTRAN is Radio Resource Control (RRC) messages. RRC messages carry all parameters required to set up, modify and release Layer 2 and Layer 1 protocol entities. RRC messages carry in their payload also all higher layer signaling (MM, CM, SM, etc.).

The mobility of user equipment in the connected mode is controlled by RRC signaling (measurements, handovers, cell updates, etc.).

RRC Layer Logical Architecture

The RRC layer logical architecture is shown in Figure PIV 8. The RRC layer can be described with four functional entities.

- The Dedicated Control Function Entity (DCFE) handles all functions and signalling specific to one UE.
- The Paging and Notification control Function Entity (PNFE) handles paging of idle mode UE(s). There is at least one PNFE in the RNC for each cell controlled by this RNC.

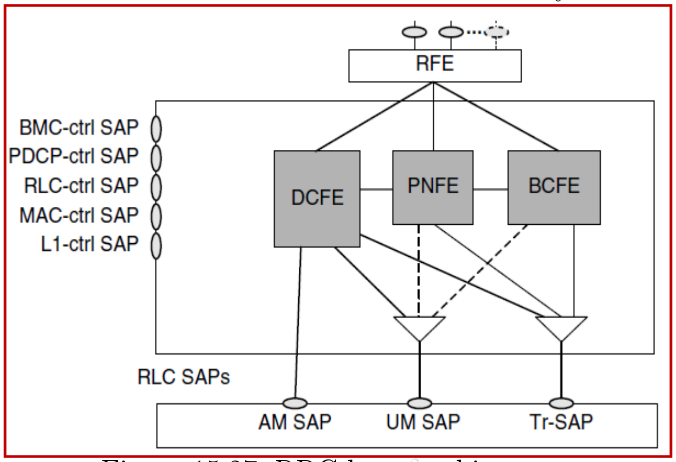


Figure 15.27: RRC layer architecture.

- The broadcast control function entity (BCFE) handles the system information broadcasting. There is at least one BCFE for each cell in the RNC.
- The fourth entity is normally drawn outside of the RRC protocol, but still belonging to access stratum and 'logically' to the RRC layer, since the information required by this entity is part of RRC messages. The entity is called Routing Function Entity (RFE) and its task is the routing of higher layer (non-access stratum) messages to different MM/CM entities (UE side) or different core network domains (UTRAN side).

RRC Functions

- Broadcast of system information, related to access stratum and non-access stratum;
- Paging;
- Initial cell selection and reselection in idle mode;
- Establishment, maintenance and release of an RRC connection between the UE and
- UTRAN;
- Control of Radio Bearers, transport channels and physical channels;
- Control of security functions (ciphering and integrity protection);
- Integrity protection of signalling messages;
- UE measurement reporting and control of the reporting;
- RRC connection mobility functions;
- Support of SRNS relocation;
- Support for downlink outer loop power control in the UE;
- Open loop power control;
- Cell broadcast service related functions;
- Support for UE Positioning functions.

Preface

This part named WCDMA Part V Radio Interface Protocols describes The WCDMA RRM, including handover control, power control, admission control, load control, and packet scheduling functionalities. For Further detail, please consult with Ref. 1.

Introduction

- Radio Resource Management (RRM) algorithms are responsible for efficient utilisation of the air interface resources. RRM is needed to guarantee Quality of Service (QoS), to maintain the planned coverage area, and to offer high capacity. The family of RRM algorithms can be divided into handover control, power control, admission control, load control, and packet scheduling functionalities.
- Power control is needed to keep the interference levels at minimum in the air interface and to provide the required quality of service.
- Handovers are needed in cellular systems to handle the mobility of the UEs across cell boundaries.
- admission control, load control and packet scheduling – are required to guarantee the quality of service and to maximize the system throughput with a mix of different bit rates, services and quality requirements. Typical locations of the RRM algorithms in a WCDMA network are shown in Figure 15.28.

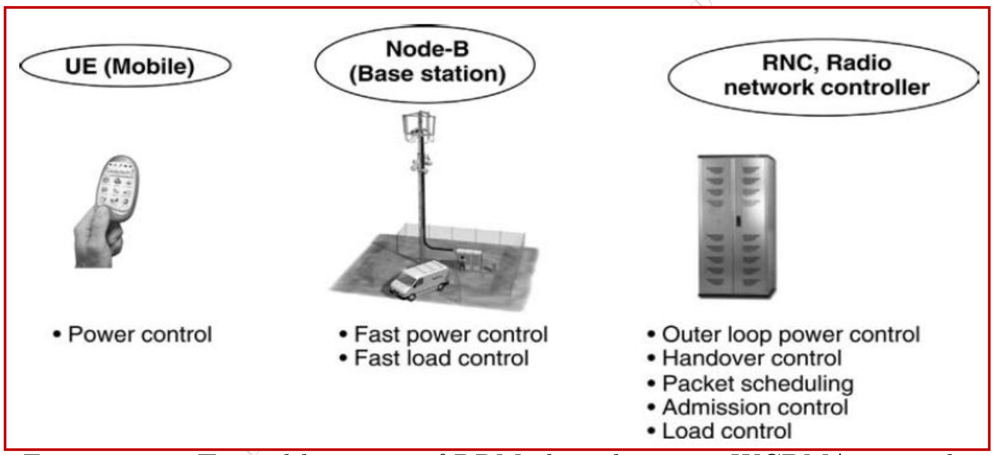


Figure 15.28: Typical locations of RRM algorithms in a WCDMA network.

Power Control

Open loop power control mechanisms that attempt to make a rough estimate of path loss by means of a downlink beacon signal, such a method would be far too inaccurate. The prime reason for this is that the fast fading is essentially uncorrelated between uplink and downlink, due to the large frequency separation of the uplink and downlink bands of the WCDMA FDD mode. Open loop power control is, however, used in WCDMA, but only to provide a coarse initial power setting of the mobile station at the beginning of a connection.

Tight and fast power control is perhaps the most important aspect in WCDMA, in particular on the uplink. Figure 15.29 depicts the problem and the solution in the form of closed loop transmission power control. Mobile stations MS1 and MS2 operate within the same frequency, separable at the base station only by their respective spreading codes. It may happen that MS1 at the cell edge suffers a path loss, say 70 dB above that of MS2 which is near the base station BS. If there were no mechanism for MS1 and MS2 to be power-controlled to the same level at the base station, MS2 could easily overshout MS1 and thus block a large part of the cell, giving rise to the so-called near–far problem of CDMA.

The optimum strategy in the sense of maximising capacity is to equalise the received power per bit of all mobile stations at all times.

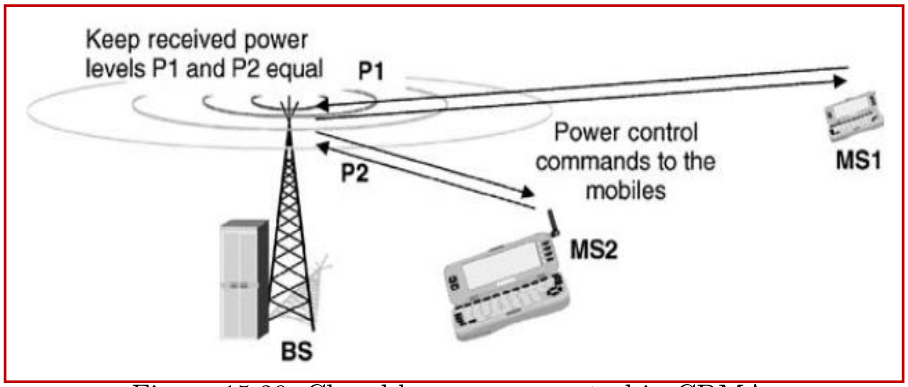


Figure 15.29: Closed loop power control in CDMA.

The solution to power control in WCDMA is fast closed loop power control, also shown in Figure PV 3. In closed loop power control in the uplink, the base station performs frequent estimates of the received Signal-to-Interference Ratio (SIR) and compares it to a target SIR. If the measured SIR is higher than the target SIR, the base station will command the mobile station to lower the power; if it is too low it will command the mobile station to increase its power. This measure–command–react cycle is executed at a rate of 1500 times per second (1.5 kHz) for each mobile station and thus operates faster than any significant change of path loss could possibly happen and, indeed, even faster than the speed of fast Rayleigh fading for low to moderate mobile speeds. Thus, closed loop power control will prevent any power imbalance among all the uplink signals received at the base station.

on the downlink there is no near–far problem due to the one-to-many scenario. All the signals within one cell originate from the one base station to all mobiles. It is, however, desirable to provide a marginal amount of additional power to mobile stations at the cell edge, as they suffer from increased other-cell interference. Also on the downlink a method of enhancing weak signals caused by Rayleigh fading with additional power is needed at low speeds when other error-correcting methods based on interleaving and error correcting codes do not yet work effectively.

Figure 15.30 shows how uplink closed loop power control works on a fading channel at low speed. Closed loop power control commands the mobile station to use a transmit power proportional to the inverse of the received power (or SIR).

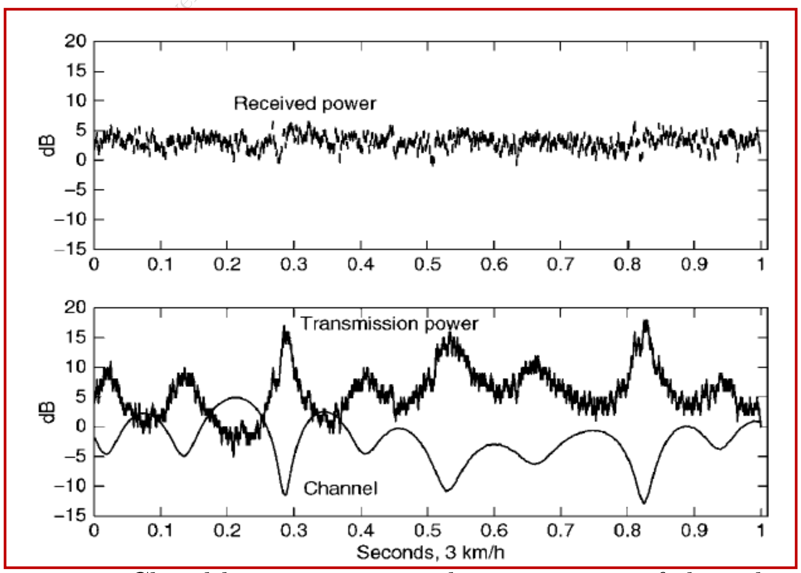


Figure 15.30: Closed-loop power control compensates a fading channel.

Outer loop control is typically implemented by having the base station tag each uplink user data frame with a frame reliability indicator, such as a CRC check result obtained during decoding of that particular user data frame. Should the frame quality indicator indicate to the Radio Network Controller (RNC) that the transmission quality is decreasing, the RNC in turn will command the base station to increase the target SIR set point by a certain amount. The reason for having outer loop control reside in the RNC is that this function should be performed after a possible soft handover combining.

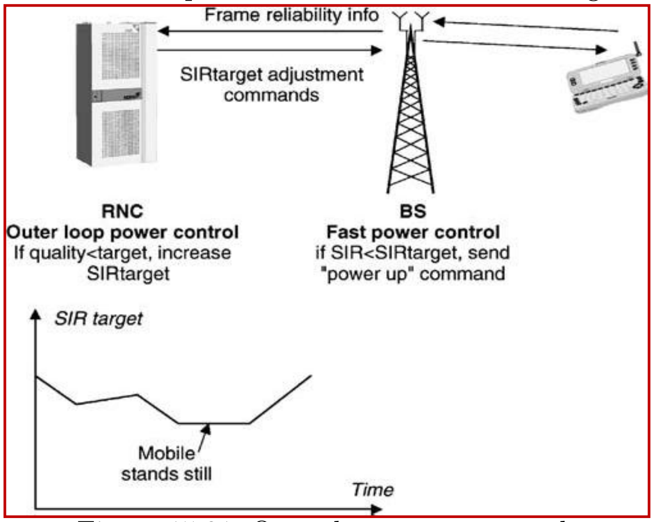


Figure 15.31: Outer loop power control.

Softer and Soft Handovers

In softer handover, a mobile station is in the overlapping cell coverage area of two adjacent sectors of a base station. The communications between mobile station and base station take place concurrently via two air interface channels, one for each sector separately. This requires the use of two separate codes in the downlink direction, so that the mobile station can distinguish the signals. The two signals are received in the mobile station by means of Rake processing. Figure 15.32 shows the softer handover scenario.

In the uplink direction a similar process takes place at the base station: the code channel of the mobile station is received in each sector, then routed to the same baseband Rake receiver and the maximal ratio combined there in the usual way. During softer handover only one power control loop per connection is active. Softer handover typically occurs in about 5–15% of connections.

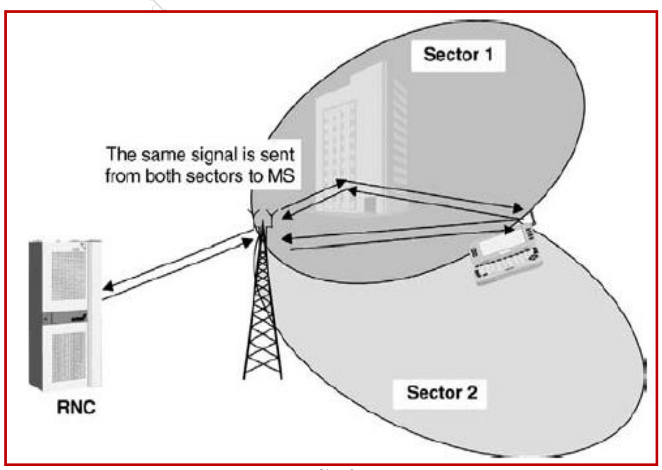


Figure 15.32: Softer handover.

Figure 15.33 shows soft handover. During soft handover, a mobile station is in the overlapping cell coverage area of two sectors belonging to different base stations. As in softer handover, the communications between mobile station and base station take place concurrently via two air interface channels from each base station separately. As in softer handover, both channels (signals) are received at the mobile station by maximal ratio combining Rake processing.

In the uplink direction soft handover differs significantly from softer handover (though does not differ much in the downlink seen from the MS). The code channel of the mobile station is received from both base stations, but the received data is then routed to the RNC for combining. This is typically done so that the same frame reliability indicator as provided for outer loop power control is used to select the better frame between the two possible candidates within the RNC. Note that during soft handover two power control loops per connection are active, one for each base station. Soft handover occurs in about 20–40% of connections.

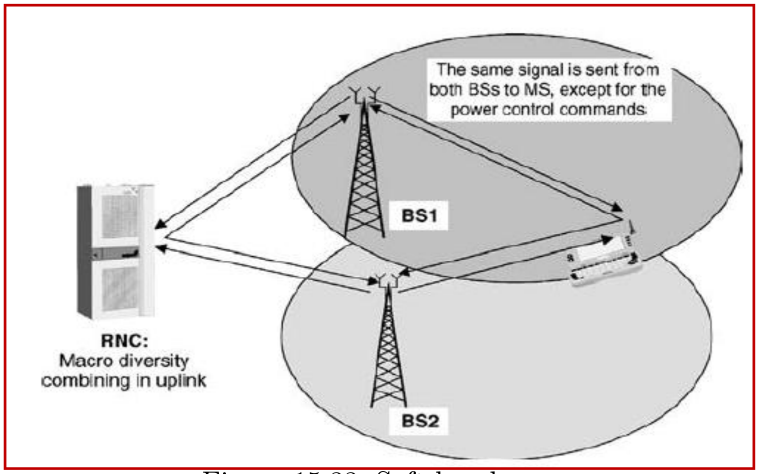


Figure 15.33: Soft handover.

These CDMA-specific handover types are needed for similar reasons as closed loop power control: without soft/softer handover there would be near–far scenarios of a mobile station penetrating from one cell deeply into an adjacent cell without being power-controlled by the latter. Very fast and frequent hard handovers could largely avoid this problem; however, they can be executed only with certain delays during which the near–far problem could develop. So, as with fast power control, soft/softer handovers are an essential interference-mitigating tool in WCDMA.

- In addition to soft /softer handover, WCDMA provides other handover types: Inter-frequency hard handovers that can be used, for example, to hand a mobile over from one WCDMA frequency carrier to another. One application for this is high capacity base stations with several carriers.
- Inter-system hard handovers that take place between the WCDMA FDD system and another system, such as WCDMA TDD or GSM.

Inter-frequency handovers

Most UMTS operators have two or three FDD frequencies available. The operation can be started using one frequency and the second and the third frequency are needed later to enhance the capacity. Several frequencies can be used in two different ways, as shown in Figure 15.34: several frequencies on the same sites for high capacity sites or macro and micro layers using different frequencies. Inter-frequency handovers between WCDMA carrier.

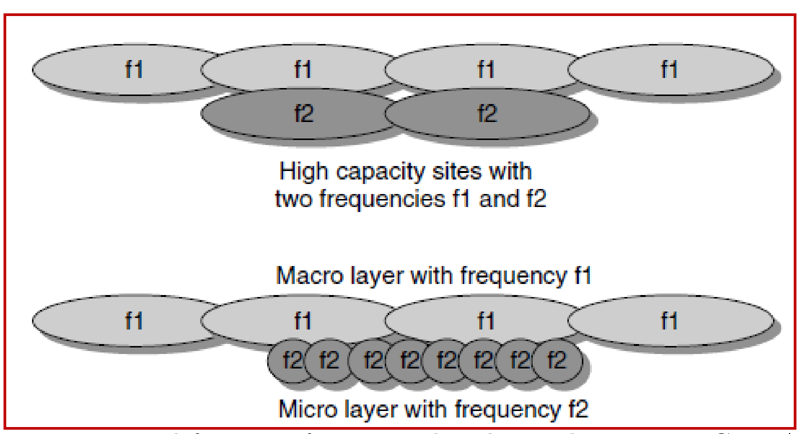


Figure 15.34: Need for inter-frequency handovers between WCDMA

carriers are needed to support these scenarios.

The inter-frequency handover procedure is shown in Figure 15.35. The UE uses the same WCDMA synchronisation procedure as for intra-frequency handovers to identify the cells on the target frequency. identification time depends mainly on the number of cells and multipath components that the UE can receive, in the same way as with intra-frequency handovers. The requirement for the cell identification in 3GPP is 5 seconds with CPICH E_c/I_0 ¿ - 20 dB.

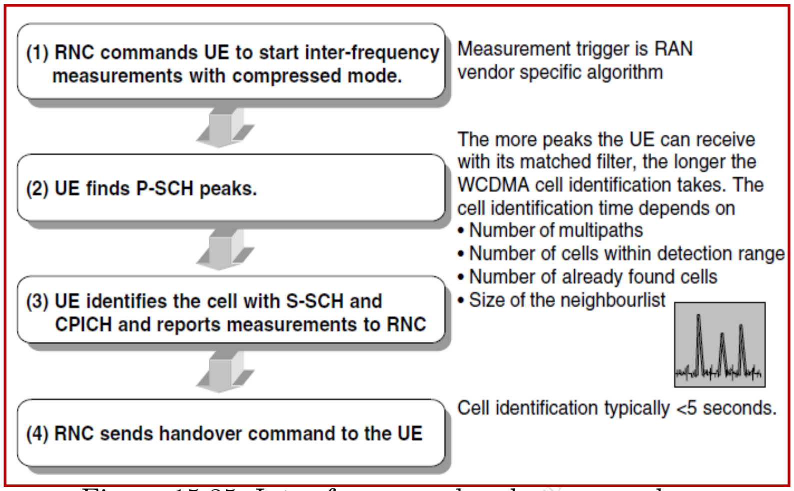


Figure 15.35: Inter-frequency handover procedure.

Inter-system handovers

WCDMA and GSM standards support handovers both ways between WCDMA and GSM. These handovers can be used for coverage or load balancing reasons. At the start of WCDMA deployment, handovers to GSM are needed to provide continuous coverage, and handovers from GSM to WCDMA can be used to lower the loading in GSM cells. This scenario is shown in Figure 15.36. When the traffic in WCDMA networks increases, it is important to have load reason handovers in both directions. The inter-system handovers are triggered in the source RNC/BSC, and from the receiving system point of view, the intersystem handover is similar to inter-RNC or inter-BSC handover.

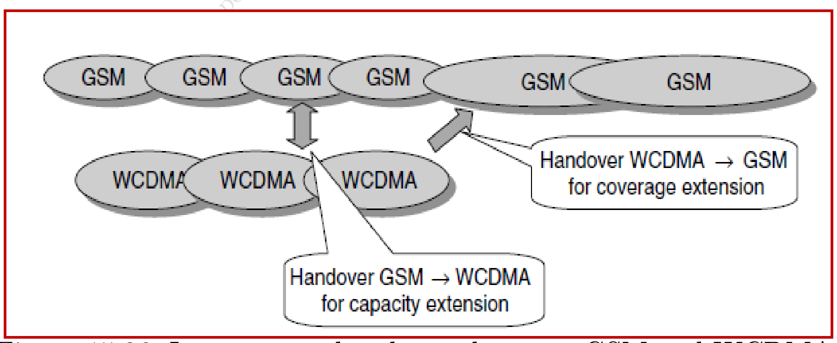


Figure 15.36: Inter-system handovers between GSM and WCDMA.

An example handover scenario

Inter-frequency and inter-system measurements are typically initiated only when there is a need to make inter-system and inter-frequency handovers. Inter-frequency handovers are needed to balance loading between WCDMA carriers and cell layers, and to extend the coverage area if the other frequency does not have continuous coverage. Inter-system handovers to GSM are needed to extend the WCDMA coverage area, to balance load between systems and to direct services to the most suitable systems.

An example handover scenario is shown in Figure 15.37. The UE is first connected to cell1 on frequency f1. When it moves, intra-frequency handover to cell2 is made. The cell2, however, happens to have a high load, and RNC commands load reason inter-frequency handover to cell5 on frequency f2. The UE remains on frequency f2 and continues to cell6. When it runs out of the coverage area of frequency f2, coverage reason inter-frequency handover is made to cell4 on frequency f1.

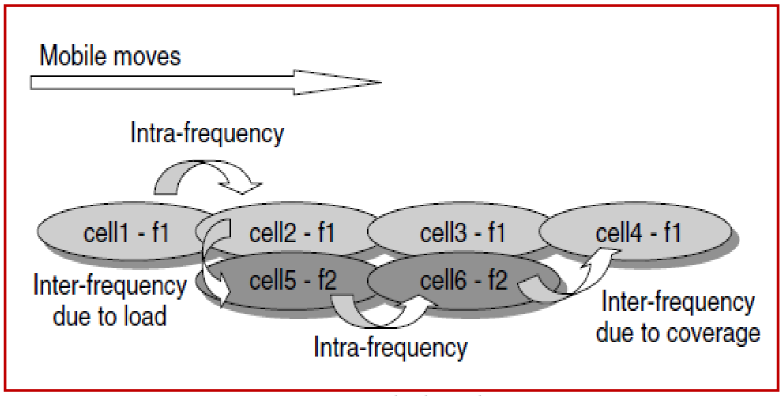


Figure 15.37: Example handover scenario.

Admission Control and Load Control

If the air interface loading is allowed to increase excessively, the coverage area of the cell is reduced below the planned values, and the quality of service of the existing connections cannot be guaranteed. Before admitting a new UE, admission control needs to check that the admittance will not sacrifice the planned coverage area or the quality of the existing connections. Admission control accepts or rejects a request to establish a radio access bearer in the radio access network. The admission control algorithm is executed when a bearer is set up or modified. The admission control functionality is located in RNC where the load information from several cells can be obtained. The admission control algorithm estimates the load increase that the establishment of the bearer would cause in the radio network. This has to be estimated separately for the uplink and downlink directions. The requesting bearer can be admitted only if both uplink and downlink admission control admit it, otherwise it is rejected because of the excessive interference that it would produce in the network. The limits for admission control are set by the radio network planning.

An example the primary uplink admission control decision criterion would be noise rise: ratio between the total received wideband power and the noise level.

Load Control

One important task of the RRM functionality is to ensure that the system is not overloaded and remains stable. If the system is properly planned, and the admission control and packets scheduler work sufficiently well. If overload is encountered, however, the load control functionality returns the system quickly and controllably back to the targeted load, which is defined by the radio network planning. The possible load control actions in order to reduce load are listed below:

- Downlink fast load control: Deny downlink power-up commands received from the UE.
- Uplink fast load control: Reduce the uplink E_b=N_0 target used by the uplink fast power control.
- Reduce the throughput of packet data traffic.
- Handover to another WCDMA carrier.
- Handover to GSM.
- Decrease bit rates of real time UEs, e.g. AMR speech codec.
- Drop low priority calls in a controlled fashion.

The first two in this list are fast actions that are carried out within a Node B. These actions can take place within one time slot, i.e. with 1.5 kHz frequency, and provide fast prioritization of the different services. The instantaneous frame error rate of the non-delay sensitive connections can be allowed to increase in order to maintain the quality of those services that cannot tolerate retransmission. These actions only cause increased delay of packet data services while the quality of the conversational services, such as speech and video telephony, is maintained. The other load control actions are typically slower. Packet traffic is reduced by the packet scheduler. Inter-frequency and inter-system handovers can also be used as load balancing and load control algorithms.

Packet Scheduling: User Specific

The WCDMA packet scheduler is located in RNC. The base station provides the air interface load measurements and the mobile provides uplink traffic volume measurements for the packet scheduler (Figure 15.38). The user-specific part controls the utilisation of Radio resource control (RRC) states, transport channels and their bit rates according to the traffic volume. WCDMA supports three types of transport channel that can be used to transmit packet data: common, dedicated and shared transport channels. The transport channels for packet data are summarised in Table 15.4.

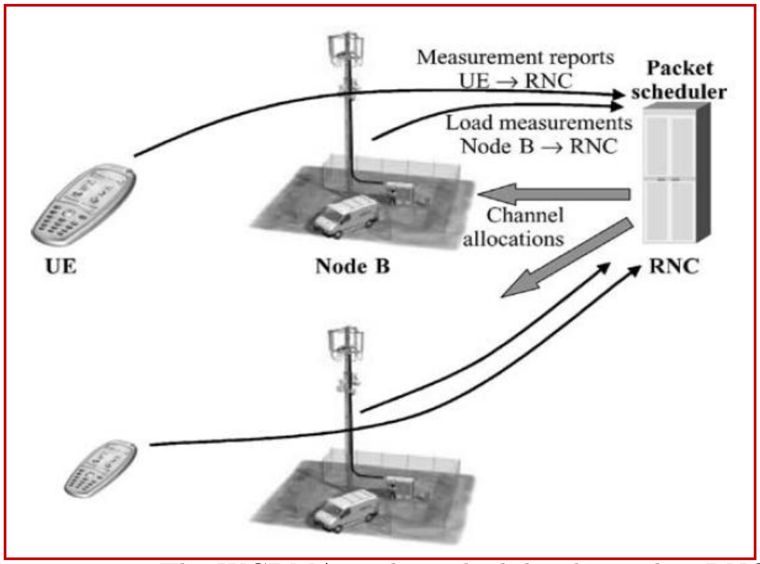


Figure 15.38: The WCDMA packet scheduler, located in RNC.

Table 15.4: Selection of WCDMA transport channels for packet data.

	Dedicated channels DCH	Shared channel DSCH	Common channels FACH	Common channels RACH	Common channels CPCH
RRC state	Cell_DCH	Cell_DCH	Cell_FACH	Cell_FACH	Cell_FACH
Uplink/ Downlink	Both	Downlink	Downlink	Uplink	Uplink
Code usage	According to maximum bit rate	Code shared between users	Fixed codes per cell	Fixed codes per cell	Fixed codes per cell
Fast power control	Yes	Yes	No	No	Yes
Soft handover	Yes	No	No	No	No
Suited for	Medium or large data amounts	Medium or large data amounts	Small data amounts	Small data amounts	Small or medium data amounts
Suited for bursty data	No	Yes	Yes	Yes	Yes
Available in first networks and terminals	Yes	No	Yes	Yes	No

Packet Scheduling: Cell-Specific

- The cell-specific packet scheduler divides the non-real time capacity between simultaneous users. The cell-specific scheduler operates periodically. This period is a configuration parameter and its value typically ranges from 100 ms to 1 s. If the load exceeds the target, the packet scheduler can decrease the load by decreasing the bit rates of packet bearers; if the load is less than the target, it can increase the load by allocating higher bit rates, as shown in Figure 15.39. The target of the scheduling is to use efficiently all remaining cell capacity for non-real time connections but also maintain interference levels within planned values so that real time connections are not affected. The cell-specific packet scheduler uses the following input information:
- Total Node B power. The total load is estimated using power-based load estimation.

- Capacity used by non-real time bearers. This capacity can be estimated using throughput based load estimation.

Inter-system handovers

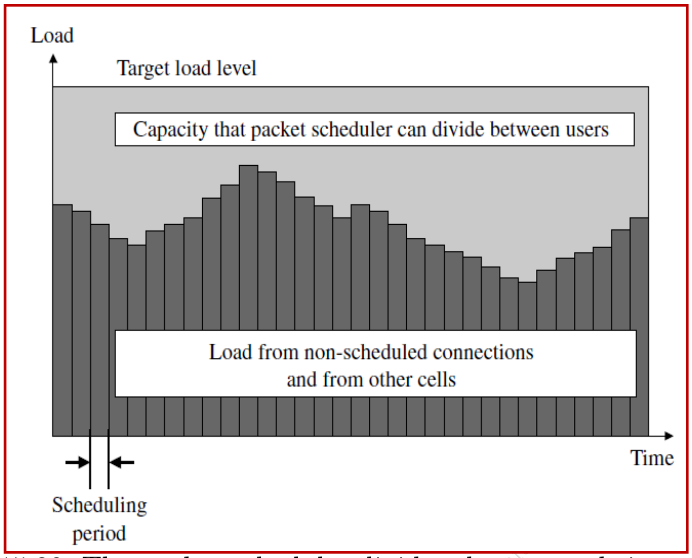


Figure 15.39: The packet scheduler divides the non-real time capacity between non real time data users.

- Target load level from network planning parameters. This parameter defines the maximum interference level that can be tolerated in the cell without affecting the real time connection.
- Bit rate upgrade requests from the user-specific packet scheduler.

The input information and the calculation principles are illustrated in Figure 15.40. Node B provides periodic total power and link power measurements in radio resource reporting to RNC. The total power measurements normally have a faster reporting cycle than link power measurements. The packet scheduler can estimate the total power used by noncontrollable traffic, which consists of real time connections and inter-cell interference. This part of the interference cannot be affected by the packet scheduler. The remaining capacity can be divided by the packet scheduler between the simultaneous users. The bit rate upgrade information from the user-specific scheduler is taken into account when changing the bit rates.

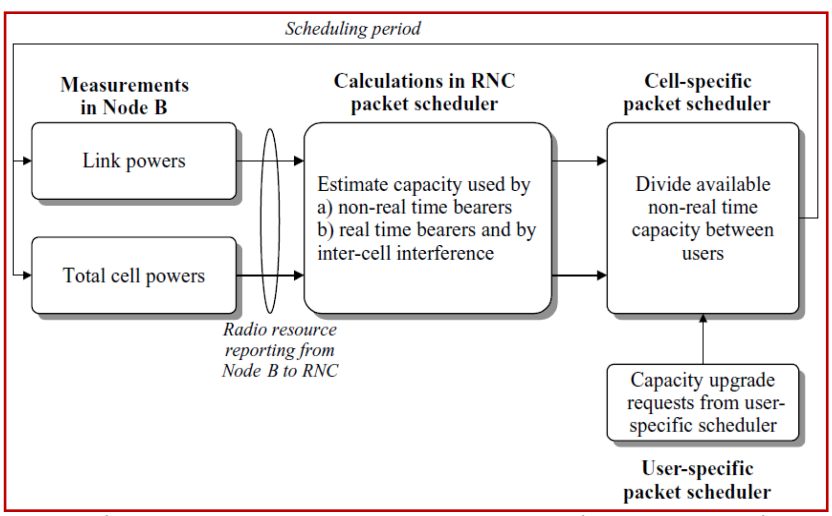


Figure 15.40: Input information and calculation principles of the cell-specific packet scheduler.

Packet Data Application Performance

The performance of the considered person-to-person applications is mainly defined by the delay: a short delay is required for these applications. WCDMA is able to support these applications by providing a lower delay than second generation systems. The content-to-person applications benefit both from high bit rate capability and from low delay. The presented business applications are sensitive to the delay because of the application signaling involved. The business applications also benefit from the high bit rate capabilities to provide close to DSL/WLAN levels of performance.

The performance estimates [1] show that GPRS is well suited for background downloads without strict delay requirements, and for downloads of small WAP/web pages. GPRS can also be used for narrowband streaming, like audio streaming. EDGE brings a clear improvement in performance for content-to-person and for business applications. EDGE performance is better than dial-up connection and it allows video streaming bit rates. WCDMA enables a number of person-to-person applications with its short delay.

WCDMA also improves business applications and download performance. High Speed Downlink Packet Access, HSDPA, brings a further improvement in end user performance for downlink packet data. The application areas are shown in Figure 15.41.

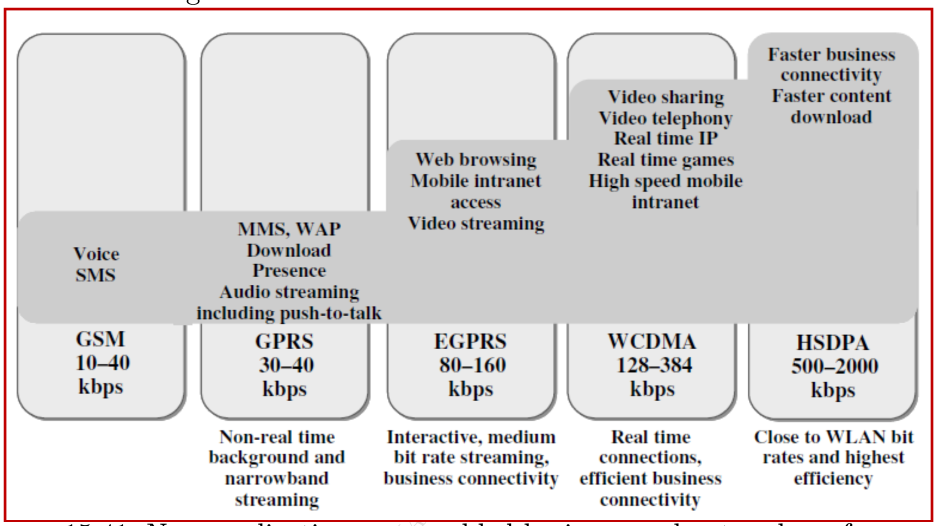


Figure 15.41: New applications are enabled by improved network performance.

CHAPTER 16
HSPA AND HSPA EVOLUTION

16.1 PREFACE

The first step in the evolution of WCDMA radio access is the introduction of High-Speed Downlink Packet Access (HSDPA) in Release 5 of the 3GPP/ WCDMA specifications. Although packet-data communication is supported already in the first release of the WCDMA standard, HSDPA brings further enhancements to the provisioning of packet-data services in WCDMA, both in terms of system and end-user performance. The downlink packet-data enhancements of HSDPA are complemented by Enhanced Uplink, introduced in Release 6 of the 3GPP/WCDMA specifications. HSDPA and Enhanced Uplink are often jointly referred to as High-Speed Packet Access (HSPA) [3].

Important requirements for cellular systems providing packet-data services are high data rates and low delays while, as at the same time, maintaining good coverage and providing high capacity. To achieve this, HSPA introduces several of the basic techniques described in Part I into WCDMA, such as higher order modulation, fast (channel-dependent) scheduling, rate control, and fast hybrid ARQ with soft combining. Altogether, HSPA provides downlink and uplink data rates up to approximately 14 Mbit/s and 5.7 Mbit/s respectively and significantly reduced round trip time and improved capacity, compared to Release 99.

3GPP Release 6 also brings efficient support for broadcast services into WCDMA through the introduction of Multimedia Broadcast Multicast Services (MBMS), suitable for applications like mobile TV. With MBMS, multiple terminals may receive the same broadcast transmission instead of the network transmitting the same information individually to each of the users. Naturally, this will lead to an improvement in the resource utilization when several users in a cell receive the same content. As a consequence of the broadcast transmission, user-specific adaptation of the transmission parameters cannot be used, and diversity as a mean to maintain good coverage is important. For MBMS, however, macro-diversity through multi-cell reception is employed for this purpose.

The evolution of the WCDMA (Figure 16.1) radio access continues and will continue also in the future. For example, 3GPP Release 7 introduces several new features. MIMO is a tool to further improve capacity and especially the HSPA peak data rates. Continuous Packet Connectivity aims at providing an 'always-on' service perception for the end-user. In parallel to enhancements of the radio-access specifications, there is an ongoing effort in advanced receiver structures. These improvements, which can provide a significant gain in both system and end-user performance, are to a large extent implementation specific, although the receiver-performance requirements as such are subject to standardization.

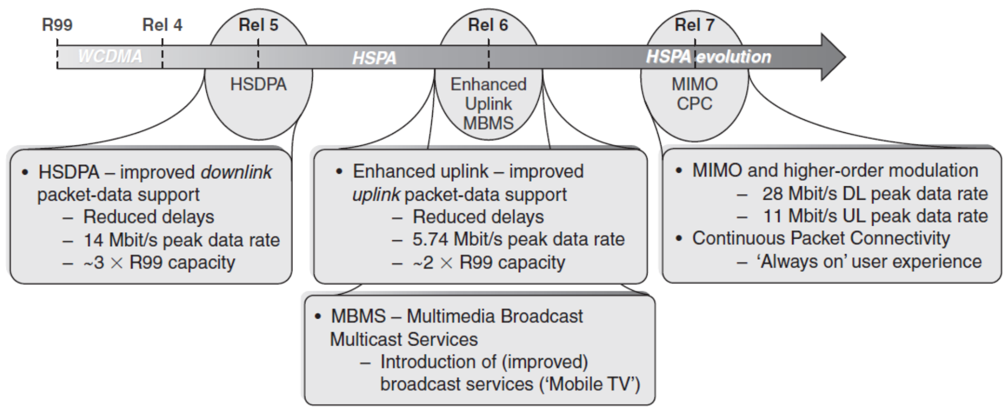


Figure 16.1: WCDMA Evolution.

16.2 INTRODUCTION

The rapid widespread deployment of WCDMA and an increasing uptake of third-generation services are raising expectations with regard to new services. Packet data services such as Web surfing and file transfer are already provided in the first release of WCDMA networks, release 99. Although this is a significant improvement compared to 2G networks, where such services have no or limited support, WCDMA is continuously evolving to provide even better performance.

For high-speed data access, data typically arrives in bursts, posing rapidly varying requirements on the amount of radio resources required. The transmission is typically bidirectional and low delays are required for a good end-user experience. Broadcast/multicast services carry data intended for multiple users. Consequently, user specific adaptation of the transmission parameters is cumbersome and diversity not requiring feedback is crucial. Due to the unidirectional nature of broadcasted data, low delays for transmission are not as important as for high-speed data access.

High-Speed Packet Access (HSPA) is an evolution of the wideband Code-Division Multiple Access (WCDMA) radio interface that offers significantly increased performance for packet data and broadcast services through the introduction of high-speed downlink packet access (HSDPA), enhanced uplink, and multimedia broadcast multicast services (MBMS).

Although building upon the basic WCDMA structure sets some constraints (on what it is possible to introduce compared to a clean slate design such as LTE), it provides the possibility to gradually improve the performance of an already deployed network: an important aspect from an operator's perspective.

HSPA exists in both FDD and TDD versions, and the two duplex schemes are designed more or less independently one another, which results in significant differences in the physical-layer design.

Technologies such as shared-channel transmission, channel-dependent scheduling, link adaptation, and hybrid ARQ with soft combining are employed in achieving so. Figure 16.2 outlines the evolutionary steps of HSPA that includes further new features such as spatial multiplexing, carrier aggregation, and broadcast support. A brief explanation is the following.

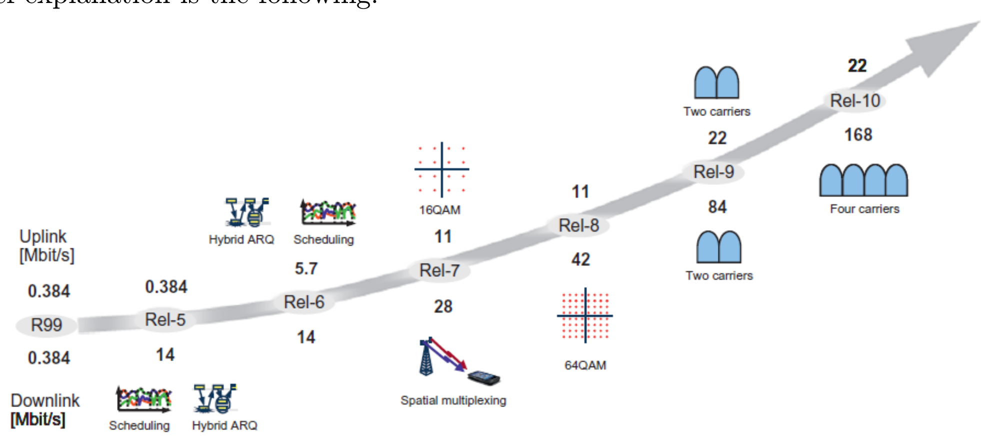


Figure 16.2: Evolution of HSPA.

257

Release 5 was the first release of HSPA. In that basic set of features, including shared-channel transmission, channel-dependent scheduling, link adaptation, and hybrid ARQ with soft combining in the downlink were introduced. In addition, peak data rates up to 14 Mbps in the downlink were supported.

In Release 6, the basic set of features for uplink transmissions were introduced that includes an uplink peak data rate of 5.7 Mbps.

Release 7 introduced the spatial multiplexing of two layers to the downlink, modulation schemes such as downlink 64QAM and uplink 16QAM to HSPA. Further enhancements include in continuous packet connectivity (to provide better DRX/DTX capabilities) and enhanced CELL˙FACH (to reduce the latency associated with state switching).

In Release 8 simultaneous usage of spatial multiplexing and 64QAM modulation in the downlink were supported.

Release 9 introduced carrier aggregation (in similar to what is done for LTE. This turns out with an increase in the HSPA bandwidth to 10 MHz.

In Release 10, further increase in bandwidth was supported using four carriers in the downlink (of HSPA) to 20 MHz that provides downlink peak data rate of 168 Mbps.

A brief overview of the most important technology components in the FDD HSPA is provided.

16.3 HSPA ARCHITECTURE

Traditional cellular systems have typically allocated resources in a relatively static way, where the data rate for a user is changed slowly or not at all. This approach is efficient for applications with a relatively constant data rate such as voice. For data with a bursty nature and rapidly varying resource requirements, fast allocation of shared resources is more efficient. In WCDMA, the shared downlink resource consists of transmission power and channelization codes in node B (the base station); while in the uplink the shared radio resource is the interference at the base station.

HSPA builds upon the WCDMA radio-access network architecture, with the existence of a Radio Network Controller (RNC), to support macro diversity like WCDMA. The functional split between the RNC and the NodeB (base station) is such that the NodeB handles most of the physical-layer processing, while the RNC handles the higher-layer protocols and all the radio-resource management. Figure 16.3 illustrates the WCDMA/HSPA radio-access network architecture.

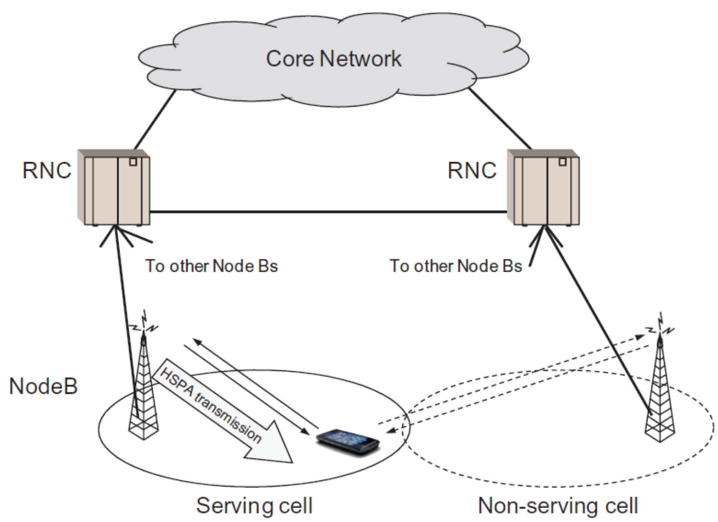


Figure 16.3: WCDMA/HSPA radio-access network architecture.

Since HSPA, to a large extent, relies on features such as channel-dependent scheduling, hybrid ARQ with soft combining, rate adaptation, etc., which to a large extent rely on rapid adaptation to the instantaneous radio conditions. This necessitates their existence in the NodeB. However, the higher-layer protocols reside in the RNC. Fast scheduling is used to control allocation of the shared resource among users on a rapid basis. Additionally, fast hybrid ARQ with soft combining enables fast retransmission of erroneous data packets. A short transmission time interval (TTI) is also employed to reduce the delays and allow the other features to adapt rapidly [2].

An important design objective of HSPA is to retain the original WCDMA functional split between layers and nodes. Minimization of the architectural changes is desirable since it simplifies introduction of

HSPA in already deployed networks and also secures operation in environments, where not all cells have been upgraded with HSPA functionality. Hence, HSPA introduces new MAC sub layers in the NodeB to address such operations as scheduling, rate adaptation, and hybrid-ARQ retransmissions.

Any terminal using HSPA will receive data from one cell called the serving cell that is responsible for scheduling. In addition, uplink soft handover is supported explicitly where the uplink data transmission is received in multiple cells, and the terminal receives power control commands from multiple cells.

16.4 HIGH-SPEED DOWNLINK PACKET ACCESS

To meet the requirement on low delays and rapid resource (re)allocation, the corresponding functionality must be located close to the air interface. In WCDMA this has been solved by locating the enhancements in the base station as part of additions to the MAC layer. An illustration of this can be found in Figure PII 2.3, where the overall UTRAN architecture with high-speed downlink packet access (HSDPA) and enhanced uplink is illustrated [2].

A number of radio network controllers (RNCs) are connected to the core network. Each RNC controls one or several node Bs, which in turn communicate with the user equipment (UE). The radio link control (RLC) entity in the RNC is unchanged compared to previous versions of WCDMA; it provides ciphering and also guarantees lossless data delivery if the hybrid ARQ protocol fails, for example, at an HSDPA cell change, where the node B buffers are flushed. Some functionality has also been added to the existing MAC functionality in the RNC to support flow control between the RNC and node B for HSDPA, and reordering and selection combining for enhanced uplink.

Furthermore, the RNC handles mobility, for example, channel switching when a user is moving from a cell supporting the enhancements into a cell where a previous release of WCDMA is used. The RNC is also responsible for the overall radio resource management, for example, setting limits on the amount of resources to be used for HSDPA and enhanced uplink.

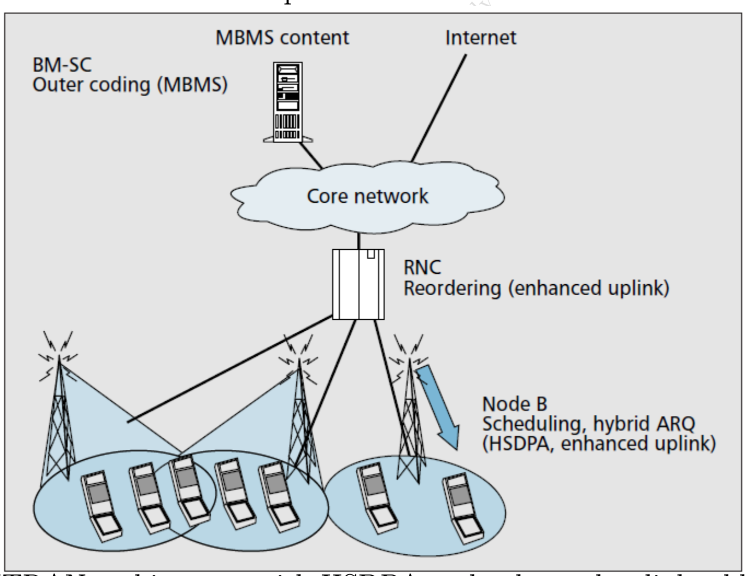


Figure 16.4: The UTRAN architecture with HSDPA and enhanced uplink additions, and MBMS enhancements. HSDPA, enhanced uplink, and MBMS can simultaneously be present in a cell, although for illustrative purposes they are shown in different cells in the figure.

A key characteristic of HSDPA is the use of shared-channel transmission. This implies that a certain fraction of the total downlink radio resources available within a cell, channelization codes, and transmission power are seen as a common resource that are dynamically shared between users, primarily in the time domain. The use of shared-channel transmission, in WCDMA implemented through the high-speed downlink shared channel (HS-DSCH), enables the possibility to rapidly allocate a large amount of the downlink resources to a user when needed.

The basic HS-DSCH code and time structure is illustrated in Figure PII 2.4. The HS-DSCH code resource consists of a number of codes of spreading factor 16 and the number of codes is configurable between 1 and 15. Codes not reserved for HS-DSCH transmission are used for other purposes (e.g., related control signaling, MBMS, and circuit-switched services such as voice). Allocation of the HS-DSCH code resource is done on a 2 ms TTI basis. The use of a short TTI reduces the overall delay and improves the tracking of fast channel variations exploited by the link adaptation and the channel-dependent scheduling as discussed below. Although the common code resource is shared primarily in the time domain, sharing in the code domain is also possible, as illustrated in Figure 16.5. The reasons are twofold: support of terminals not able to despread the full set of codes, and efficient support of small payloads (i.e., when the transmitted data does not require the full set of allocated HS-DSCH codes).

In addition to being allocated a part of the overall code resource, a certain part of the total available cell power should also be allocated for HS-DSCH transmission. Note that the HSDSCH is not power controlled but rate controlled, as discussed below. This allows the remaining power (after serving other, power controlled channels) to be used for HS-DSCH transmission and enables efficient exploitation of the shared power resource.

Soft handover is not used for the HS-DSCH for two reasons: first, the diversity gains traditionally exploited through downlink soft handover are instead exploited in the scheduling process. Second, as the scheduler is located in the node B, inter-node-B soft handover is not possible.

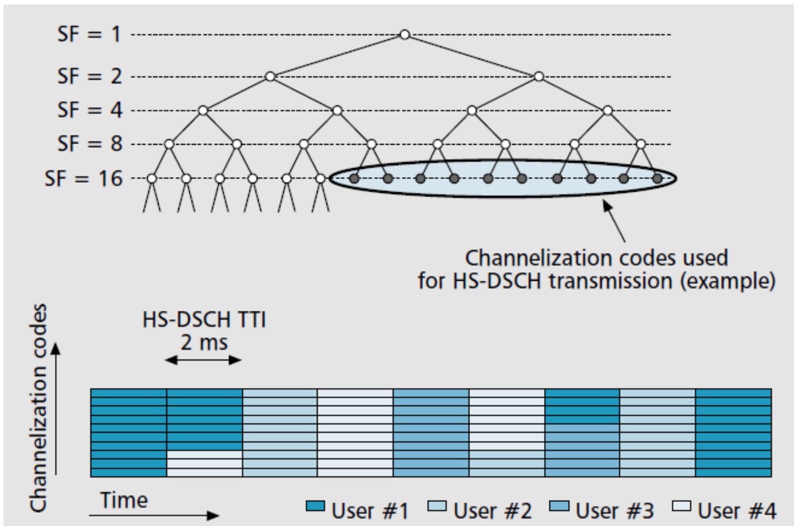


Figure 16.5: Code and time domain structure for HS-DSCH.

Control signaling necessary for successful reception of the HS-DSCH at the terminal is carried on shared control channels. There is also a need for transmitting power-control commands for the uplink in the downlink. These are either carried on a conventional dedicated channel, which can also carry non-HS-DSCH services, or on a new type of dedicated channel introduced in release 6, optimized to carry power control commands only and reducing the code space required by up to a factor of ten.

Key Features of HSDPA

Link Adaptation: Traditionally, fast closed loop power control has been used in CDMA systems to combat the fading variations in the radio channel and to maintain a constant E_b/N_0. For services requiring a constant data rate (e.g., circuit- switched voice services), this is clearly a suitable approach. However, for services that can tolerate some jitter in the data rate, it is more efficient to control the E_b/N_0 by adjusting the data rate while keeping the transmission power constant. This is commonly known as link adaptation or rate adaptation.

Link adaptation is implemented by adjusting the channel-coding rate, and selecting between QPSK and 16-QAM. Higher-order modulation such as 16-QAM makes more efficient use of bandwidth than QPSK, but requires greater received E_b/N_0. Consequently, 16-QAM is mainly useful in advantageous channel conditions. In addition, the data rate also depends on the number of channelization codes assigned for HS-DSCH transmission in a TTI. The data rate is selected independently for each 2 ms TTI by node B, and the link-adaptation mechanism can therefore track rapid channel variations.

To provide node B with information about the instantaneous channel conditions (information necessary to select a suitable data rate), each terminal regularly transmits a channel quality indicator (CQI) using an uplink control channel. The reporting frequency is configurable, but reports can be sent as often as every 2 ms. To allow for different receiver implementations, the CQI is expressed as a recommended data rate given the current channel conditions. Therefore, a terminal implementing an interference-suppressing receiver, will typically report a higher CQI, thus leading to a higher data rate for upcoming transmissions.

Hence, in contrast to power-controlled channels, where an advanced receiver only will benefit the network operator in terms of a lower transmission power, there is an incentive for the end user to buy a terminal with a more advanced receiver.

Scheduling: The scheduler is a key element and to a large extent determines the overall downlink performance, especially in a highly loaded network. In each TTI, the scheduler decides to which user(s) the HS-DSCH should be transmitted and, in close cooperation with the link-adaptation mechanism, at what data rate. A significant increase in capacity can be obtained if channel-dependent scheduling is used.

Since the radio conditions for the users typically vary independently, at each point in time there is almost always a user whose channel quality is near its peak. The gain obtained by transmitting to users with favorable conditions is commonly known as multi-user diversity and the gains are larger with larger channel variations and a larger number of users. Thus, in contrast to the traditional view that fading is an undesirable effect that has to be combated, fading is in fact desirable and should be exploited. This is illustrated in Figure 16.6.

Practical scheduler strategy exploits the short-term variations (e.g., due to multipath fading and fast interference variations) while maintaining some degree of long-term fairness between the users. In principle, the larger the long-term unfairness, the higher the cell capacity and trade-off between the two are required. Additionally, traffic priorities should also be taken into account, for example, to prioritize streaming services before a file download. The scheduler algorithm is implementation specific, but several different strategies are discussed in the literature. Information about each user's instantaneous radio-channel quality is obtained from the CQI.

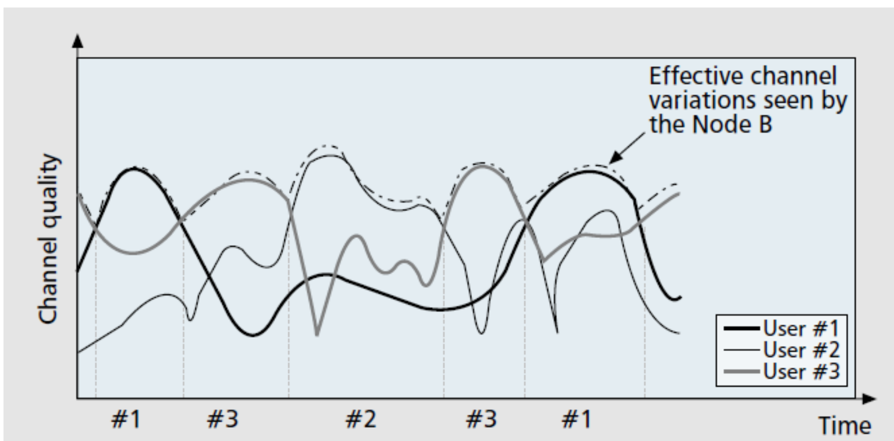


Figure 16.6: Surfing the fading peaks to exploit the channel variations.

Hybrid ARQ: The third key feature of HSDPA is hybrid ARQ with soft combining, which allows the terminal to rapidly request retransmission of erroneously received transport blocks, essentially fine-tuning the effective code rate and compensating for errors made by the link-adaptation mechanism. The terminal attempts to decode each transport block it receives and reports to node B its success or failure 5 ms after the reception of the transport block.

Soft combining implies that the terminal does not discard soft information in case it cannot decode a data block as in traditional hybrid ARQ protocols, but combines soft information from previous transmission attempts with the current retransmission to increase the probability of successful decoding.

Incremental redundancy (IR) is used as the basis for soft combining, that is, the retransmissions may contain parity bits not included in the original transmission. It is well known that IR can provide significant gains when the code rate for the initial transmission attempts is high, as the additional parity bits in the retransmission result in a lower overall code rate. Thus, IR is mainly useful in band limited situations, for example, when the terminal is close to the base station and the amount of channelization codes (and not the transmission power) limits the achievable data rate.

The set of coded bits to use for the retransmission is controlled by node B, taking the available UE memory into account. Due to the properties of the hybrid ARQ protocol, out-of sequence reception of data can occur. A reordering functionality in the UE ensures that data are delivered in sequence to higher layers.

16.5 ENHANCED UPLINK

Enhanced Uplink, also known as High-Speed Uplink Packet Access (HSUPA), has been introduced in WCDMA Release 6. It provides improvements in WCDMA uplink capabilities and performance in terms of higher data rates, reduced latency, and improved system capacity, and is therefore a natural complement to HSDPA.

At the core of Enhanced Uplink are two basic technologies used also for HSDPA: fast scheduling and fast hybrid ARQ with soft combining. For similar reasons as for HSDPA, Enhanced Uplink also introduces a short 2 ms uplink TTI. These enhancements are implemented in WCDMA through a new transport channel, the Enhanced Dedicated Channel (E-DCH). Although the same technologies are used both for HSDPA and Enhanced Uplink, there are fundamental differences between them, which have affected the detailed implementation of the features:

- In the downlink, the shared resource is transmission power and the code space, both of which are located in one central node, the NodeB. In the uplink, the shared resource is the amount of allowed uplink interference, which depends on the transmission power of multiple distributed nodes, the UEs.
- The scheduler and the transmission buffers are located in the same node in the downlink, while in the uplink the scheduler is located in the NodeB while the data buffers are distributed in the UEs. Hence, the UEs need to signal buffer status information to the scheduler.
- The WCDMA uplink, also with Enhanced Uplink, is inherently non-orthogonal, and subject to interference between uplink transmissions within the same cell. This is in contrast to the downlink, where different transmitted channels are orthogonal. Fast power control is therefore essential for the uplink to handle the near-far problem. The E-DCH is transmitted with a power offset relative to the power-controlled uplink control channel and by adjusting the maximum allowed power offset, the scheduler can control the E-DCH data rate. This is in contrast to HSDPA, where a (more or less) constant transmission power with rate adaptation is used.
- Soft handover is supported by the E-DCH. Receiving data from a terminal in multiple cells is fundamentally beneficial as it provides diversity, while transmission from multiple cells in case of HSDPA is cumbersome and with questionable benefits as discussed in the previous chapter. Soft handover also implies power control by multiple cells, which is necessary to limit the amount of interference generated in neighboring cells and to maintain backward compatibility and coexistence with UE not using the E-DCH for data transmission.
- In the downlink, higher-order modulation, which trades power efficiency for bandwidth efficiency, is useful to provide high data rates in some situations, for example when the scheduler has assigned a small number of channelization codes for a transmission but the amount of available transmission power is relatively high. The situation in the uplink is different; there is no need to share channelization codes between users and the channel coding rates are therefore typically lower than for the downlink. Hence, unlike the downlink, higher order modulation is less useful in the uplink macro-cells and therefore not part of the first release of enhanced uplink.

HSUPA Architecture

For efficient operation, the scheduler should be able to exploit rapid variations in the interference level and the channel conditions. Hybrid ARQ with soft combining also benefits from rapid retransmissions as this reduces the cost of retransmissions. These two functions should therefore reside close to the radio-interface. As a result, and for similar reasons as for HSDPA, the scheduling and hybrid ARQ functionalities of Enhanced Uplink are located in the NodeB. Furthermore, also similar to the HSDPA design, it is preferable to keep all radio-interface layers above MAC intact. Hence, ciphering, admission control, etc., is still under the control of the RNC. This also allows for a smooth introduction of Enhanced Uplink in selected areas; in cells not supporting E-DCH transmissions, channel switching can be used to map the user's data flow onto the DCH instead.

Following the HSDPA design philosophy, a new MAC entity, the MAC-e , is introduced in the UE and NodeB. In the NodeB, the MAC-e is responsible for support of fast hybrid ARQ retransmissions and scheduling, while in the UE, the MAC-e is responsible for selecting the data rate within the limits set by the scheduler in the NodeB MAC-e.

When the UE is in soft handover with multiple NodeBs, different transport blocks may be successfully decoded in different NodeBs. Consequently, one transport block may be successfully received in one NodeB while another NodeB is still involved in retransmissions of an earlier transport block. Therefore, to ensure in-sequence delivery of data blocks to the RLC protocol, a reordering functionality is required in the RNC in the form of a new MAC entity, the MAC-es. In soft handover, multiple MAC-e entities are used per UE as the data is received in multiple cells. However, the MAC-e in the serving cell has the main responsibility for the scheduling; the MAC-e in a non-serving cell is mainly handling the hybrid ARQ protocol (Figure 16.7)

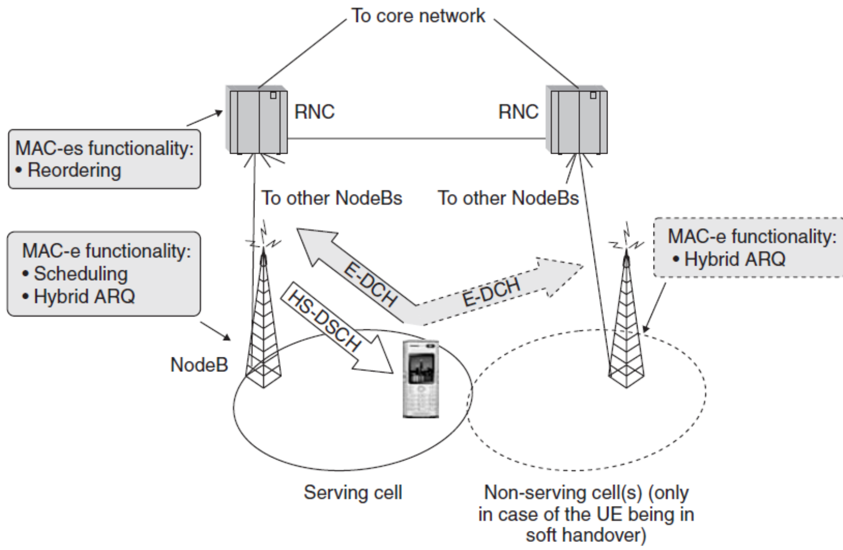


Figure 16.7: The architecture with E-DCH (and HS-DSCH) configured.

Key Features of Enhanced Uplink

Scheduling: For Enhanced Uplink, the scheduler is a key element, controlling when and at what data rate the UE is allowed to transmit. The higher the data rate a terminal is using, the higher the terminal's received power at the NodeB must be to maintain the E_b/N_0 required for successful demodulation. By increasing the transmission power, the UE can transmit at a higher data rate. However, due to the non-orthogonal uplink, the received power from one UE represents interference for other terminals. Hence, the shared resource for Enhanced Uplink is the amount of tolerable interference in the cell.

If the interference level is too high, some transmissions in the cell, control channels and non-scheduled uplink transmissions, may not be received properly. On the other hand, a too low interference level may indicate that UEs are artificially throttled and the full system capacity not exploited. Therefore, Enhanced Uplink relies on a scheduler to give users with data to transmit permission to use an as high data rate as possible without exceeding the maximum tolerable interference level in the cell.

Unlike HSDPA, where the scheduler and the transmission buffers both are located in the NodeB, the data to be transmitted resides in the UEs for the uplink case. At the same time, the scheduler is located in the NodeB to coordinate different UEs transmission activities in the cell. Hence, a mechanism for communicating the scheduling decisions to the UEs and to provide buffer information from the UEs to the scheduler is required.

The scheduling framework for Enhanced Uplink is based on scheduling grants sent by the NodeB scheduler to control the UE transmission activity and scheduling requests sent by the UEs to request resources. The scheduling grants control the maximum allowed E-DCH-to-pilot power ratio the terminal may use; a larger grant implies the terminal may use a higher data rate but also contributes more to the interference level in the cell. Based on measurements of the (instantaneous) interference level, the scheduler controls the scheduling grant in each terminal to maintain the interference level in the cell at a desired target (Figure 16.8). In HSDPA, typically a single user is addressed in each TTI. For Enhanced Uplink, the implementation-specific uplink scheduling strategy in most cases schedules multiple users in parallel. The reason is the significantly smaller transmit power of a terminal compared to a NodeB: a single terminal typically cannot utilize the full cell capacity on its own.

Inter-cell interference also needs to be controlled. Even if the scheduler has allowed a UE to transmit at a high data rate based on an acceptable intra-cell interference level, this may cause non-acceptable interference to neighboring cells. Therefore, in soft handover, the serving cell has the main responsibility for the scheduling operation, but the UE monitors scheduling information from all cells with which the UE is in soft handover. The non-serving cells can request all its non-served users to lower their E-DCH data rate by transmitting an overload indicator in the downlink. This mechanism ensures a stable network operation.

Fast scheduling allows for a more relaxed connection admission strategy. A larger number of bursty high-rate packet-data users can be admitted to the system as the scheduling mechanism can handle the situation when multiple users need to transmit in parallel. If this creates an unacceptably high interference level in the system, the scheduler can rapidly react and restrict the data rates they may use. Without fast scheduling, the admission control would have to be more conservative and reserve a margin in the system in case of multiple users transmitting simultaneously.

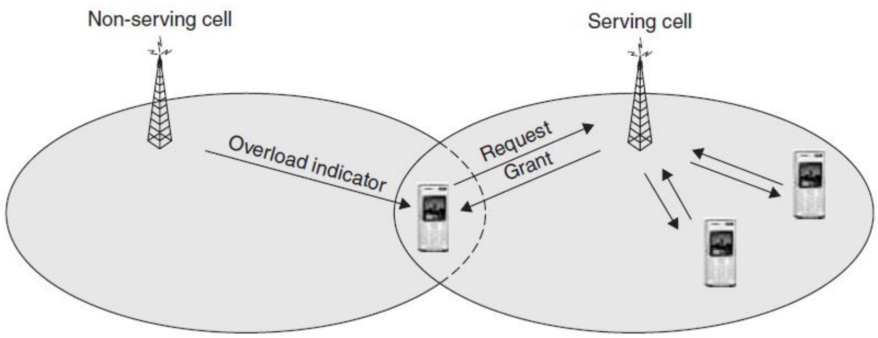


Figure 16.8: Enhanced Uplink scheduling framework.

Hybrid ARQ with soft combining: Fast hybrid ARQ with soft combining is used by Enhanced Uplink for basically the same reason as for HSDPA: to provide robustness against occasional transmission errors. A similar scheme as for HSDPA is used. For each transport block received in the uplink, a single bit is transmitted from the NodeB to the UE to indicate successful decoding (ACK) or to request a retransmission of the erroneously received transport block (NAK).

One main difference compared to HSDPA stems from the use of soft handover in the uplink. When the UE is in soft handover, this implies that the hybrid ARQ protocol is terminated in multiple cells. Consequently, in many cases, the transmitted data may be successfully received in some NodeBs but not in others. From a UE perspective, it is sufficient if at least one NodeB successfully receives the data. Therefore, in soft handover, all involved NodeBs attempt to decode the data and transmits an ACK or a NAK. If the UE receives an ACK from at least one of the NodeBs, the UE considers the data to be successfully received.

Hybrid ARQ with soft combining can be exploited not only to provide robustness against unpredictable interference, but also to improve the link efficiency to increase capacity and/or coverage. One possibility to provide a data rate of x Mbit/s is to transmit at x Mbit/s and set the transmission power to target a low error probability (in the order of a few percent) in the first transmission attempt.

Alternatively, the same resulting data rate can be provided by transmitting using n times higher data rate at an unchanged transmission power and use multiple hybrid ARQ retransmissions. From the discussion in Chapter 7, this approach on average results in a lower cost per bit (a lower E_b/N_0) than the first approach. The reason is that, on average, less than n transmissions will be used. This is sometimes known as early termination gain and can be seen as implicit rate adaptation.

Additional coded bits are only transmitted when necessary. Thus, the code rate after retransmissions is determined by what was needed by the instantaneous channel conditions. This is exactly what rate adaptation also tries to achieve, the main difference being that rate adaptation tries to find the correct code rate prior to transmission. The same principle of implicit rate adaptation can also be used for HS-DSCH in the downlink to improve the link efficiency.

Multimedia Broadcast Multicast Services (MBMS): In the past, cellular systems have mostly focused on transmission of data intended for a single user and not on broadcast services. Broadcast networks, exemplified by the radio and TV broadcasting networks, have on the other hand focused on covering very large areas and have offered no or limited possibilities for transmission of data intended for a single user. Multimedia Broadcast and Multicast Services, (MBMS), introduced for WCDMA in Release 6, supports multicast/ broadcast services in a cellular system, thereby combining multicast and unicast transmissions within a single network. With MBMS, the same content is transmitted to multiple users located in a specific area, the MBMS service area, in a unidirectional fashion. The MBMS service area typically covers multiple cells, although it can be made as small as a single cell.

Broadcast and multicast describe different, although closely related scenarios:

In broadcast, a point-to-multipoint radio resource is set up in each cell being part of the MBMS broadcast area and all users subscribing to the broadcast service simultaneously receive the same transmitted signal. No tracking of users' movement in the radio access network is performed and users can receive the content without notifying the network. Mobile TV is an example of a service that could be provided through MBMS broadcast.

In multicast, users request to join a multicast group prior to receiving any data. The user movements are tracked and the radio resources are configured to match the number of users in the cell. Each cell in the MBMS multicast area may be configured for point-to-point or point-to-multipoint transmission.

In sparsely populated cells with only one or a few users subscribing to the MBMS service, point-to-point transmission may be appropriate, while in cells with a larger number of users, point-to-multipoint transmission is better suited. Multicast therefore allows the network to optimize the transmission type in each cell.

To a large extent, MBMS affects mainly the nodes above the radio-access network. A new node, the Broadcast Multicast Service Center (BM-SC), illustrated in Figure 16.9, is introduced.

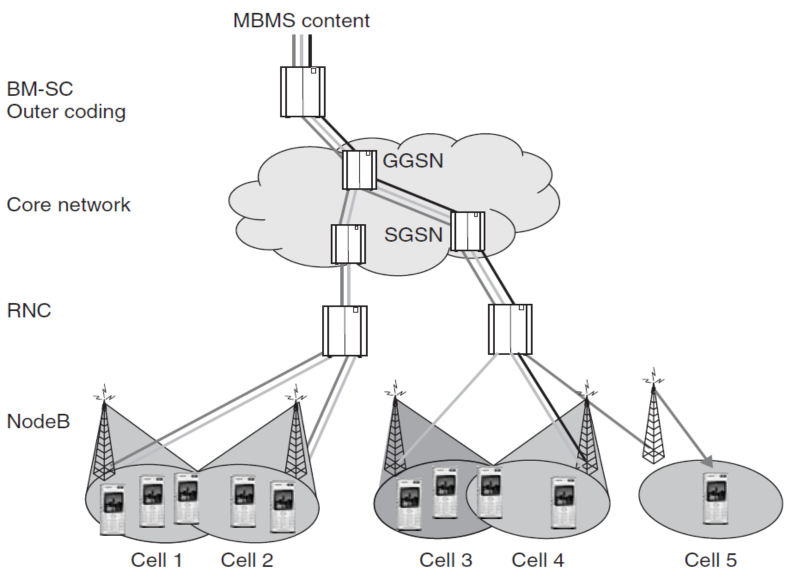


Figure 16.9: Example of MBMS services. Different services are provided in different areas using broadcast in cells 1–4. In cell 5, unicast is used as there is only single user subscribing to the MBMS service.

The BM-SC is responsible for authorization and authentication of content provider, charging, and the overall configuration of the data flow through the core network. It is also responsible for application-level coding. In Figure 16.10, typical phases during an MBMS session are illustrated. First, the service is announced. In case of broadcast, there are no further actions required by the user; the user simply 'tunes' to the channel of interest. In case of multicast, a request to join the session has to be sent to become member of the corresponding MBMS service group and, as such, receive the data. Before the MBMS transmission can start, the BM-SC sends a session-start request to the core network, which allocates the necessary internal resources and request the appropriate radio resources from the radio access network. All terminals of the corresponding MBMS service group are

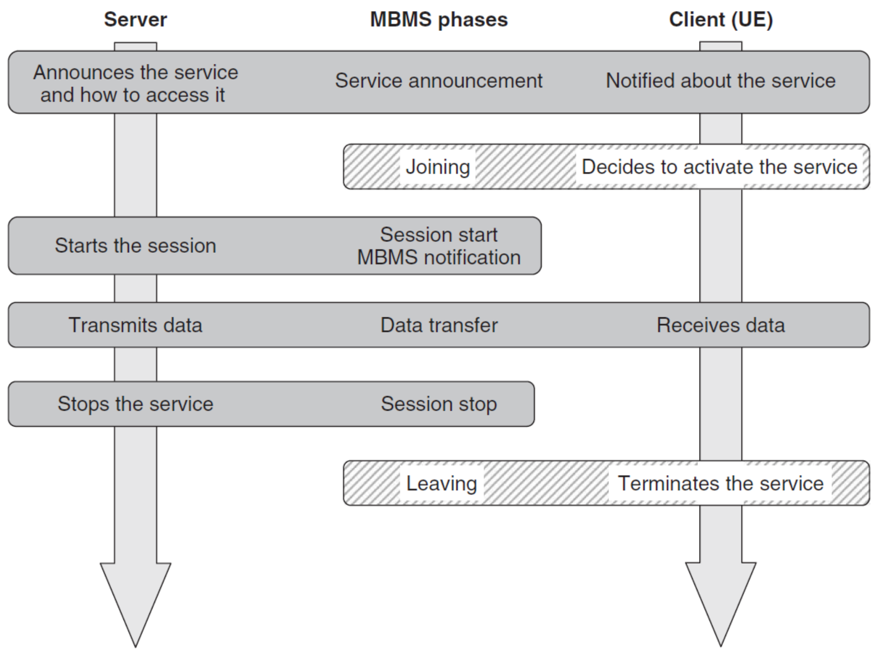


Figure 16.10: Example of typical phases during an MBMS session. The dashed phases are only used in case of multicast and not for broadcast.

also notified that content delivery from the service will start.

Data will then be transmitted from the content server to the end users. When the data transmission stops, the server will send a session-stop notification. Also, users who want to leave an MBMS multicast service can request to be removed from the MBMS service group. One of the main benefits brought by MBMS is the resource savings in the network as a single stream of data may serve multiple users. This is seen in Figure 16.9, where three different services are offered in different areas. From the BMSC, data streams are fed to each of the NodeBs involved in providing the MBMS services. As seen in the figure, the data stream intended for multiple users is not split until necessary. For example, there is only a single stream of data sent to all the users in cell 3. This is in contrast to previous releases of UTRAN, where one stream per user has to be configured throughout both the core network and the radio access network.

16.6 HSPA EVOLUTION

Although HSPA as defined in Release 6 is a significant enhancement to the packet-data functionality in WCDMA, the performance is further enhanced in Release 7 and beyond. Herein, this is referred to as ‘ HSPA Evolution ’ and consists of both the introduction of new major features, such as MIMO, and many smaller improvements to existing structures which, when taken together as a package, represents a major increase in performance and capabilities. In the following, some of these enhancements are discussed.

MIMO: MIMO is one of the major new features in Release 7, introduced to increase the peak data rates through multi-stream transmission. Strictly speaking, MIMO, Multiple Input Multiple Output, in its general interpretation denotes the use of multiple antennas at both transmitter and receiver. This can be used to obtain a diversity gain and thereby increase the carrier-to-interference ratio at the receiver. However, the term is commonly used to denote the transmission of multiple layers or multiple streams as a mean to increase the data rate possible in a given channel. Hence, MIMO, or spatial multiplexing, should mainly be seen as a tool to improve the end-user throughput by acting as a ‘data-rate booster’. Naturally, an improved end-user throughput will to some extent also result in an increased system throughput.

The scheme used for HSDPA-MIMO is sometimes referred to as dual-stream transmit adaptive arrays (D-TxAA), which is a multi-codeword scheme with rank adaptation and pre-coding. Each stream is subject to the same physical-layer processing in terms of coding, spreading and modulation as the corresponding single-layer HSDPA case. After coding, spreading, and modulation, linear pre-coding is used before the result is mapped to the two transmit antennas. Mainly the physical-layer processing is affected by the introduction of MIMO; the impact to the protocol layer is small and MIMO is to a large extent only visible as a higher data rate.

HSDPA-MIMO data transmission: To support dual-stream transmission, the HS-DSCH is modified to support up to two transport blocks per TTI. Each transport block represents one stream. A CRC is attached to each of the transport blocks and each transport block is individually coded. This is illustrated in Figure 16.11. Since two transport blocks are used in case of multi-stream transmission, HSDPA-MIMO is a multi-codeword scheme and allows for a successive interference- cancellation receiver in the UE. The physical-layer processing for each stream is identical to the single-stream case up to, and including spreading.

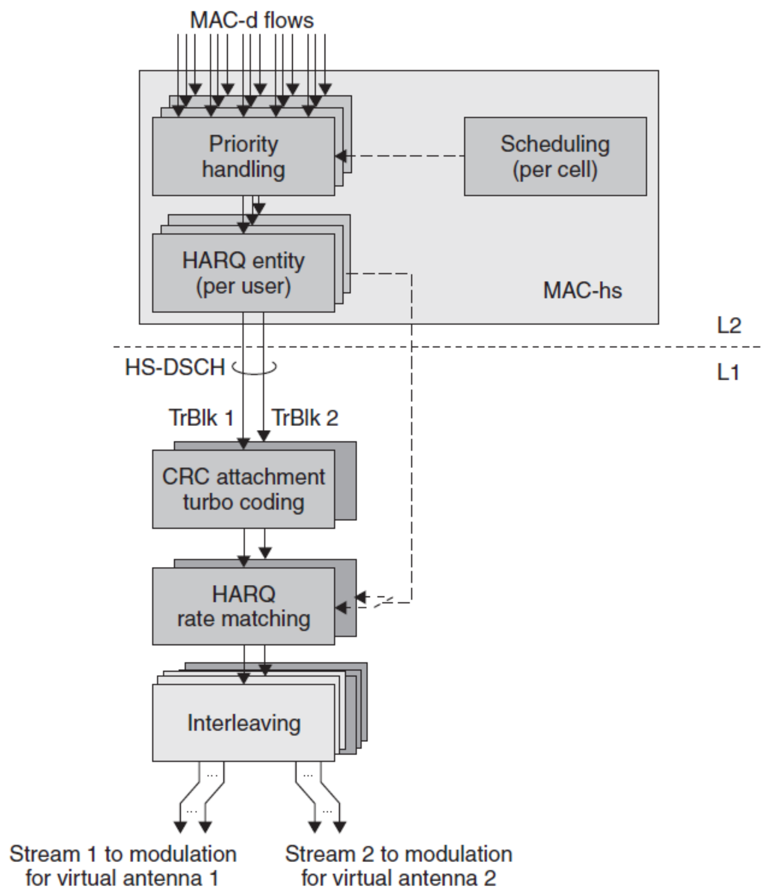


Figure 16.11: HS-DSCH processing in case of MIMO transmission.

To avoid wasting channelization-code resources, the same set of channelization codes should be used for the two streams. At the receiver, the two streams are separated by the appropriate receiver processing, for example by interference cancellation.

Rate control for HSDPA-MIMO: Rate control for each stream is similar to the single-stream case. However, the rate-control mechanism also needs to determine the number of streams to transmit and the pre-coding matrix to use. Hence, for each TTI, the number of streams to transmit, the transport-block sizes for each of the streams, the number of channelization codes, the modulation scheme, and the pre-coding matrix is determined by the rate-control mechanism. This information is provided to the UE on the HS-SCCH, similar to the non-MIMO case. As the scheduler controls the size of the two transport blocks in case of multi-stream transmission, the data rate of the two streams can be individually controlled.

Hybrid-ARQ with soft combining for HSDPA-MIMO: For each stream, the physical-layer hybrid-ARQ processing and the use of multiple hybrid-ARQ processes are identical to the single-stream case. However, as the multiple streams are transmitted over different antennas, one stream may be correctly received while another stream require retransmission of the payload. Therefore, one hybrid-ARQ acknowledgement per stream is sent from the UE to the NodeB.

Control signaling for HSDPA-MIMO: To support MIMO, the out-band control signaling is modified accordingly. No changes to the in-band control signaling in the form of the MAC-hs header are required as reordering and priority queue selection is not affected by the introduction of MIMO. However, to efficiently support the higher data rates provided by MIMO, the MAC, and RLC layers are updated with flexible segmentation.

UE capabilities: To allow for a wide range of different UE implementations, MIMO support is not mandatory for all UEs. Furthermore, as multi-stream transmission mainly is a tool for increasing the supported peak data rates, MIMO is mainly relevant for the high-end UE categories. Therefore, UEs supporting MIMO are always capable of receiving 15 channelization codes. Furthermore, some capabilities may differ depending on whether the terminal is configured in MIMO mode
or not; as an example some categories of MIMO-capable terminals may support 64QAM only when operating in non-MIMO mode.

Higher-order modulation: The introduction of MIMO allows for a substantial increase in peak data rate. This increase is achieved by exploiting certain propagation conditions in the channel through multi-stream transmission. However, in some situations the UE can experience a high carrier-to-interference ratio at the same time as the channel does not support multi-stream transmission, for example in case of line-of-sight propagation. Higher-order modulation is useful in such cases.

Higher-order modulation is also useful as it allows for high data rates in cases when the UE or NodeB are not equipped with multiple antennas. Therefore, Release 7 increases the highest modulation order to 64QAM in the downlink and 16QAM in the uplink and Release 8 takes this one step further by providing simultaneous support for 64QAM and MIMO in the downlink. The peak rates with higher order modulation and MIMO are listed in Table 16.1.

Supporting 64QAM in the downlink and 16QAM in the uplink is based on the same principles as already specified for Release 6. However, to fulfill the transmitter RF requirements, a somewhat larger power amplifier back-off may be required for these modulation schemes.

Table 16.1: Peak rates in downlink and uplink with higher order modulation and MIMO.

Downlink peak rate, Mbit/s				Uplink peak rate, Mbit/s	
Non-MIMO		MIMO			
16QAM	64QAM	16QAM	64QAM	BPSK/QPSK	16QAM
14	21	28	42	5.7	11

Continuous packet connectivity: Packet-data traffic is often highly bursty with occasional periods of transmission activity. Clearly, from a user performance perspective, it is advantageous to keep the HS-DSCH and E-DCH configured to rapidly be able to transmit any user data. At the same time, maintaining the connection in uplink and downlink comes at a cost. From a network perspective, there is a cost in uplink interference from the DPCCH transmission even in absence of data transmission.

From a UE perspective, power consumption is the main concern; even when no data is received the UE needs to transmit the DPCCH and monitor the HS-SCCH. To improve the packet-data support in HSPA, a set of features known as Continuous Packet Connectivity (CPC), is introduced in Release 7. CPC consists of three building blocks:

- Discontinuous transmission (DTX), to reduce the uplink interference and thereby increase the uplink capacity, as well as to save battery power.
- Discontinuous reception (DRX), to allow the UE to periodically switch off the receiver circuitry and save battery power.
- HS-SCCH-less operation to reduce the control signaling overhead for small amounts of data, as will be the case for services such as VoIP.

The intention with these features is to provide an 'always-on' experience for the end user by keeping the UE in CELL˙DCH for a longer time and avoiding frequent state changes to the low-activity states, as well as improving the capacity for services, such as VoIP. Since they mainly relate to packet-data support, they are only supported in combination with HSPA; thus if a DCH is configured, the CPC features cannot be used. In the following, the three building blocks and the interaction between them is described.

Enhanced CELL˙FACH operation: The purpose of continuous packet connectivity is, as discussed in the previous sections, to provide an ‘ always-on ’ user experience by keeping the UE in the active state (known as CELL˙DCH in WCDMA) while still providing mechanisms for reduced power consumption. However, eventually the UE will be switched to CELL˙FACH if there has been no transmission activity for a certain period of time. Once the UE is in to CELL˙FACH, signaling on the Forward Access Channel (FACH), a low-rate common downlink transport channel, is required to move the UE to CELL˙DCH prior to any data exchange on HS-DSCH and E-DCH can take place. The physical resources to which the FACH is mapped is semi-statically configured by the RNC and, to maximize the resources available for HS-DSCH and other downlink channels, the amount of resources (and thus the FACH data rate) is typically kept small, in the order of a few tens of kbit/s.

To reduce the latency associated with state changes, Release 7 improves the performance by allowing HS-DSCH to be used also in the CELL˙FACH state. This is often referred to as Enhanced CELL˙FACH operation. Using the HS-DSCH also in CELL˙FACH allows for a significant reduction in the delays associated with switching to CELL˙DCH state. Instead of using a low-rate FACH, the signaling from the network to the UE can be carried on the high-rate HS-DSCH. This can result in a significant reduction in call-setup delay and a corresponding improvement in the user perception.

Layer 2 Protocol Enhancements: To fully benefit from the high data rates supported by HS-DSCH, especially in combination with 64QAM and MIMO, Release 7 introduces enhancements to the RLC and MAC-hs protocols in additions to the physical-layer enhancements. In releases prior to Release 7, the RLC PDU size is semi-statically configured.

This is appropriate for the low-to-medium data rates, but for the high data rates targeted by HSPA Evolution, the RLC PDU size, the RLC roundtrip time, and the RLC window size may limit the peak data rates and cause the RLC protocol to stall. One possibility to avoid this is to increase the RLC PDU size, but for Release 7 a somewhat more advanced solution has been adopted, flexible RLC.

Advanced receivers: There are many ways to enhance performance in terms of, for example, data throughput and coverage without modifications to the specifications. Many of these enhancements are based on more advanced receiver algorithms and are thus implemented in software in the baseband processing. Other enhancements require more ‘hardware’ in terms of antennas and RF components, for example, receiver antenna diversity and beam-forming techniques.

Advanced receivers are possible for both base stations and mobile devices (UEs). For the single receiver, the enhancement is manifested by a decrease in the signal-to-noise ratio (Eb / N_0) required for a specific quality of service. The improved receiver performance enables improved quality of service in terms of, for example, end-user data rates. If a large number of the user devices have receiver enhancements, it will lead to improved system performance in terms of, for example, system wide data throughput.

The standards developed in 3GPP do in principle not specify the receiver structure to be used. The specifications define performance requirements for demodulation of the different physical channels. What type of receiver implementation that is used to meet those requirements is not specified, there is full freedom for a UE vendor to use any implementation, as long as the 3GPP requirements are met. It is for this reason not possible to mandate use of certain receivers through the 3GPP specifications, if the freedom of implementation is to be kept.

Most performance requirements are however developed with a baseline receiver in mind. The performance of the baseline receiver is simulated and an agreed ‘implementation margin’ is added to the results to model (additional) receiver imperfections not included in the simulations. Once the agreed performance limit is entered into the specification, it is to be fulfilled regardless of what receiver has been implemented.

MBSFN Operation: The basic principle behind MBMS is to transmit the same signal from multiple base stations. To improve the performance of MBMS, true single-frequency operation, also referred to as Multicast-Broadcast Single Frequency Network (MBSFN) is advantageous. In MBSFN, the cells are time synchronized and identical copies of the signal are transmitted from all the cells. Combined with advanced interference- suppressing receivers in the terminal, very high signal-to-noise ratios can be obtained. To exploit this, Release 7 adds support for 16QAM on FACH and time-multiplexed pilots.

16.7 SUMMARY

HSPA, consisting of improved packet-data support in the downlink and uplink, by adapting the transmission parameters to rapid variations in the radio-channel quality, as well as traffic variations, a significant performance gain in terms of higher peak data rates, reduced latencies, and improved system capacity can be achieved.

Technologies such as channel dependent scheduling, rate adaptation, and hybrid ARQ with soft combining are used to rapidly adjust the transmission parameters.

The broadcast performance of WCDMA has also been considerably improved through the introduction of MBMS and its enhancements. Transmissions from multiple cells are combined to obtain diversity, which is of key importance for efficient broadcast performance. Furthermore, the use of application-level coding provides additional diversity as the terminal is able to reconstruct the source data even if some packets are missing.

HSPA is also subject to a continuous evolution. Through the use of MIMO, multiple antennas at both the NodeB and the UE can be use to further increase the peak data rates. The use of Continuous Packet Connectivity provides an 'always-on' experience for the end user. These enhancements all build upon the basic WCDMA structure defined by Release 99.

Terminals using a later release of the specifications are able to coexist with terminals from previous releases on the same carrier. Naturally, this sets some constraints on what is possible to introduce, but also offers the significant benefit for an operator to gradually improve the network capacity.

CHAPTER 17
LTE AND LTE-ADVANCED

17.1 PREFACE

The work on LTE was initiated in late 2004 with the overall aim of providing a new radio-access technology focusing on packet-switched data only. The first phase of the 3GPP work on LTE was to define a set of performance and capability targets for LTE [10]. This included targets on peak data rates, user/system throughput, spectral efficiency, and control/user-plane latency. In addition, requirements were also set on spectrum flexibility, as well as on interaction/compatibility with other 3GPP radio-access technologies (GSM, WCDMA/HSPA, and TD-SCDMA).

The first release of the LTE specifications, release 8, was completed in the spring of 2008 and commercial network operation began in late 2009. Release 8 has then been followed by additional LTE releases, introducing additional functionality and capabilities in different areas, as illustrated in Figure 17.1.

In parallel to the development of LTE, there has also been an evolution of the overall 3GPP network architecture, termed System Architecture Evolution (SAE), including both the radio-access network and the core network. Requirements were also set on the architecture evolution, leading to a new flat radio-access-network architecture with a single type of node, the eNodeB1, as well as new core-network architecture. The most important technologies used by LTE release 8 include transmission schemes, scheduling, multi-antenna support, and spectrum flexibility as well as the additional features introduced in LTE releases 9 and 10 will be addressed.

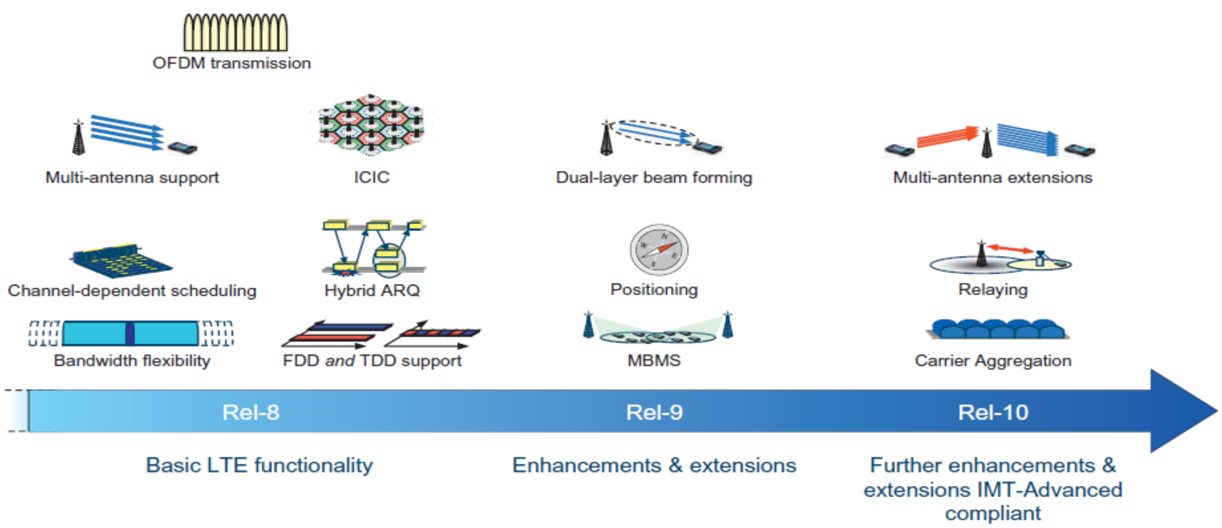


Figure 17.1: LTE and its evolution.

17.1 LTE RELEASE 8

Basic Principles

The main principles behind LTE are described in the following.

Transmission Scheme

The LTE downlink transmission scheme is based on conventional OFDM since LTE for which a wide bandwidth and support for advanced multi-antenna transmission are the key requirements. Due to the relatively long OFDM symbol time in combination with a cyclic prefix, OFDM provides a high degree of robustness against channel frequency selectivity in the downlink. Also, the LTE uplink is based on OFDM transmission. For uplink data transmission, the OFDM modulator is preceded by a DFT precoder, leading to DFT-spread OFDM (DFTS-OFDM). Often, the term DFTS-OFDM is used to describe the LTE uplink transmission scheme in general. Hence, LTE uplink transmission scheme is described as OFDM with different techniques, including DFT precoding for data transmission. The use of DFTS-OFDM on the LTE uplink allows for orthogonal separation of uplink transmissions also in the frequency domain.

Orthogonal separation is in many cases beneficial as it avoids interference between uplink transmissions from different terminals within the cell (intra-cell interference). Allocating a very large instantaneous bandwidth for transmission from a single terminal is not an efficient strategy in situations where the data rate is mainly limited by the available terminal transmit-power rather than the bandwidth. In such situations, a terminal can instead be allocated only a part of the total available bandwidth and other terminals within the cell can be scheduled to transmit in parallel on the remaining part of the spectrum.

In other words, the LTE uplink transmission scheme allows for both time division (TDMA) and frequency division (FDMA) between users.

Channel-Dependent Scheduling and Rate Adaptation

At the core of the LTE transmission scheme is the use of shared-channel transmission with the overall time–frequency resource dynamically shared between users. The use of shared-channel transmission is well matched to the rapidly varying resource requirements posed by packet-data communication and also enables several of the other key technologies on which LTE is based. The scheduler controls, for each time instant, to which users the different parts of the shared resource should be assigned. The scheduler also determines the data rate to be used for each transmission. Thus, rate adaptation can be seen as a part of the scheduling functionality. The scheduler is thus a key element and to a large extent determines the overall system performance, especially in a highly loaded network.

Both downlink and uplink transmissions are subject to tight scheduling in LTE. It is well known that a substantial gain in system capacity can be achieved if the channel conditions are taken into account in the scheduling decision, so-called channel-dependent scheduling. Due to the use of OFDM in both the downlink and uplink transmission directions, the scheduler has access to both the time and frequency domains. In other words, the scheduler can, for each time instant and frequency region, select the user with the best channel conditions, as illustrated in Figure 17.2.

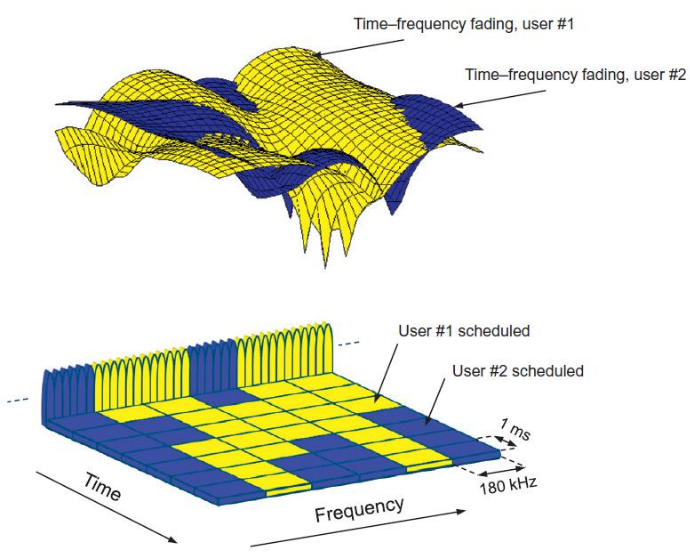


Figure 17.2: Downlink channel-dependent scheduling in the time and frequency domains.

For LTE, scheduling decisions can be taken as often as once every 1 ms and the granularity in the frequency domain is 180 kHz. This allows for relatively rapid channel variations in both the time and frequency domains to be tracked and utilized by the scheduler. To support downlink scheduling, a terminal may provide the network with channel-state reports indicating the instantaneous downlink channel quality in both the time and frequency domains. The channel state is typically obtained by measuring on reference signals transmitted in the downlink.

Based on the channel-state reports, also referred to as channel-state information (CSI), the downlink scheduler can assign resources for downlink transmission to different terminals, taking the channel quality into account in the scheduling decision. In principle, a scheduled terminal can be assigned an arbitrary combination of 180 kHz wide resource blocks in each 1 ms scheduling interval.

The LTE uplink is based on orthogonal separation of different uplink transmissions and it is the task of the uplink scheduler to assign resources in both the time and frequency domains to different terminals. Scheduling decisions, taken once per 1 ms, control what set of terminals are allowed to transmit within a cell during a given time interval and, for each terminal, on what frequency resources the transmission is to take place and what transmission parameters, including the data rate, to use. However, obtaining information about the uplink channel conditions may not be feasible or desirable in all situations. Therefore, different means to obtain uplink diversity are important as a complement in situations where uplink channel-dependent scheduling is not suitable.

Inter-Cell Interference Coordination

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

The basic aim of such inter-cell interference coordination (ICIC) is to, if possible, avoid scheduling transmissions to/from terminals at the cell border simultaneously in neighboring cells, thereby avoiding the worst-case interference situations.

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

An even more complicated interference situation may occur in so-called heterogeneous network deployments consisting of overlapping cell layers with large differences in the cell output power. This is as part of the discussion on LTE release-10 features.

Hybrid ARQ with Soft Combining

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

Multi-Antenna Transmission

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

As dual receive antennas are the baseline for all LTE terminals, the downlink performance is also improved. The simplest way of using multiple receive antennas is classical receive diversity to collect additional energy and suppress fading, but additional gains can be achieved in interference-limited scenarios if the antennas are used not only to provide diversity, but also to suppress interference.

Multiple transmit antennas at the base station can be used for transmit diversity and different types of beam-forming. The main goal of beam-forming is to improve the received SINR and, eventually, improve system capacity and coverage.

Spatial multiplexing, sometimes referred to as MIMO, using multiple antennas at both the transmitter and receiver is supported by LTE. Spatial multiplexing results in an increased data rate, channel conditions permitting, in bandwidth-limited scenarios by creating several parallel "channels". Alternatively, by combining the spatial properties with the appropriate interference suppressing receiver processing, multiple terminals can transmit on the same time–frequency resource in order to improve the overall cell capacity. This is sometimes referred to as multi-user MIMO. Up to four layers can be spatially multiplexed in release 8. Later releases further enhance the multi-antenna support, as described later.

Spectrum Flexibility

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

Flexibility in Duplex Arrangement: One important part of the LTE requirements in terms of spectrum flexibility is the possibility to deploy LTE-based radio access in both paired and unpaired spectrum. Therefore, LTE supports both frequency- and time-division-based duplex arrangements. Frequency-Division Duplex (FDD), as illustrated on the left in Figure PII 3.3, implies that downlink and uplink transmission take place in different, sufficiently separated, frequency bands. Time-Division Duplex (TDD), as illustrated on the right in Figure PII 3.3, implies that downlink and uplink transmission take place in different, non-overlapping time slots. Thus, TDD can operate in unpaired spectrum, whereas FDD requires paired spectrum.

Bandwidth Flexibility: An important characteristic of LTE is the possibility for different transmission bandwidths on both downlink and uplink. The main reason for this is that the amount of spectrum available for LTE deployment may vary significantly between different frequency bands and also depending on the exact situation of the operator.

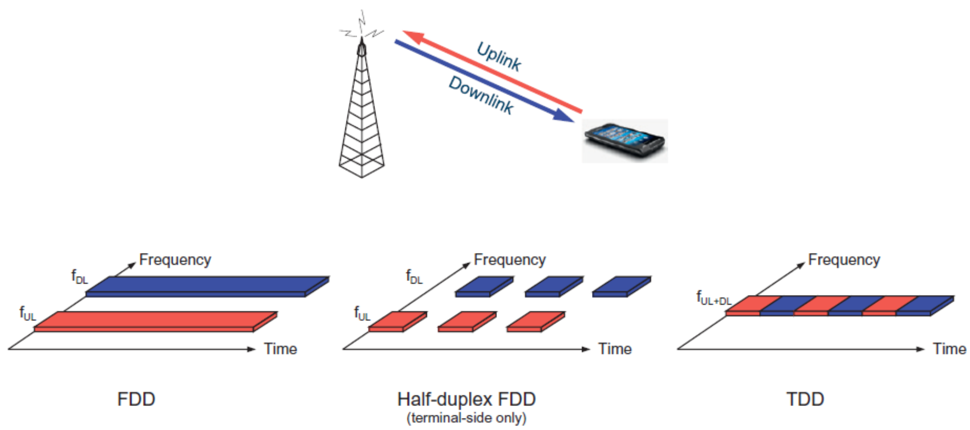


Figure 17.3: Frequency- and time-division duplex.

Furthermore, the possibility of operating in different spectrum allocations gives the possibility for gradual migration of spectrum from other radio-access technologies to LTE. LTE supports operation in a wide range of spectrum allocations, achieved by a flexible transmission bandwidth being part of the LTE specifications. A sufficiently large amount of spectrum may not always be available, either due to the band of operation or due to a gradual migration from another radio-access technology, in which case LTE can be operated with a more narrow transmission bandwidth. The LTE physical-layer specifications are bandwidth agnostic and do not make any particular assumption on the supported transmission bandwidths beyond a minimum value.

The basic radio-access specification, including the physical-layer and protocol specifications, allows for any transmission bandwidth ranging from roughly 1 MHz up to around 20 MHz. However, in practice the LTE radio-access technology supports a limited set of transmission bandwidths since at an initial stage radio-frequency requirements are only specified for a limited subset of transmission bandwidths, corresponding to what is predicted to be relevant spectrum-allocation sizes and relevant migration scenarios. But, additional transmission bandwidths can easily be introduced by updating only the RF specifications.

17.2 LTE RELEASE 9

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

Multicast and Broadcast Support

Multi-cell broadcast implies transmission of the same information from multiple cells. By exploiting this at the terminal, effectively using signal power from multiple cell sites at the detection, a substantial improvement in coverage (or higher broadcast data rates) can be achieved.

By transmitting not only identical signals from multiple cell sites (with identical coding and modulation), but also synchronizing the transmission timing between the cells, the signal at the terminal will appear exactly as a signal transmitted from a single cell site and subject to multi-path propagation. Due to the OFDM robustness to multi-path propagation, such multi-cell transmission, in 3GPP also referred to as Multicast/ Broadcast Single-Frequency Network (MBSFN) transmission, will then not only improve the received signal strength, but also eliminate the inter-cell interference.

Thus, with OFDM, multi-cell broadcast/multicast throughput may eventually be limited by noise only and can then, in the case of small cells, reach extremely high values. Note that the use of MBSFN transmission for multi-cell broadcast/multicast assumes the use of tight synchronization and time alignment of the signals transmitted from different cell sites.

Positioning

Positioning, as the name implies, refers to functionality in the radio-access network to determine the location of individual terminals. Determining the position of a terminal can, in principle, be done by including a GPS receiver in the terminal. Although this is a quite common feature, not all terminals include the necessary GPS receiver and there may also be cases when the GPS service is not available.

LTE release 9 therefore introduces positioning support inherent in the radio-access network. By measuring on special reference signals transmitted regularly from different cell sites, the location of the terminal can be determined.

Dual-Layer Beam-Forming

Release 9 enhances the support for combining spatial multiplexing with beam-forming. Although the combination of beam-forming and spatial multiplexing was already possible in release 8, this was then restricted to so-called codebook-based precoding.

In release 9, spatial multiplexing can be combined with so-called non-codebook-based precoding, thereby improving the flexibility in deploying various multi-antenna schemes. The enhancements can also be seen as the basis for a further improvement in the area in LTE release10.

17.3 LTE RELEASE 10 AND IMT-ADVANCED

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

One of the main targets of LTE release 10 was to ensure that the LTE radio-access technology would be fully compliant with the IMT-Advanced requirements, thus the name LTE-Advanced is often used for LTE release 10. However, in addition to the ITU requirements, 3GPP also defined its own targets and requirements for LTE release 10 (LTE-Advanced). These targets / requirements extended the ITU requirements both in terms of being more aggressive as well as including additional requirements.

One important requirement was backwards compatibility. Essentially this means that an earlier-release LTE terminal should always be able to access a carrier supporting LTE release-10 functionality, although obviously not being able to utilize all the release-10 features of that carrier. LTE release 10 was completed in late 2010 and enhances LTE spectrum flexibility through carrier aggregation, further extends multi-antenna transmission, introduces support for relaying, and provides improvements in the area of inter-cell interference coordination in heterogeneous network deployments.

Carrier Aggregation

In LTE release 10 the transmission bandwidth can be further extended by means of so-called carrier aggregation (CA), where multiple component carriers are aggregated and jointly used for transmission to/from a single terminal. Up to five component carriers, possibly each of different bandwidth, can be aggregated, allowing for transmission bandwidths up to 100 MHz.

Backwards compatibility is catered for as each component carrier uses the release-8 structure. Hence, to a release-8/9 terminal each component carrier will appear as an LTE release-8 carrier, while a carrier-aggregation capable terminal can exploit the total aggregated bandwidth, enabling higher data rates. In the general case, a different number of component carriers can be aggregated for the downlink and uplink.

Component carriers do not have to be contiguous in frequency, which enables exploitation of fragmented spectrum; operators with a fragmented spectrum can provide high-data-rate services based on the availability of a wide overall bandwidth even though they do not pose a single wideband spectrum allocation. From a baseband perspective, there is no difference between the cases in Figure 17.4 and they are all supported by LTE release 10.

However, the RF-implementation complexity is vastly different with the first case being the least complex. Thus, although spectrum aggregation is supported by the basic specifications, the actual release-10 RF requirements will be strongly constrained, including specification of only a limited number of aggregation scenarios and including support of inter-band aggregation only for the most advanced terminals, but excluding non-contiguous intra-band aggregation.

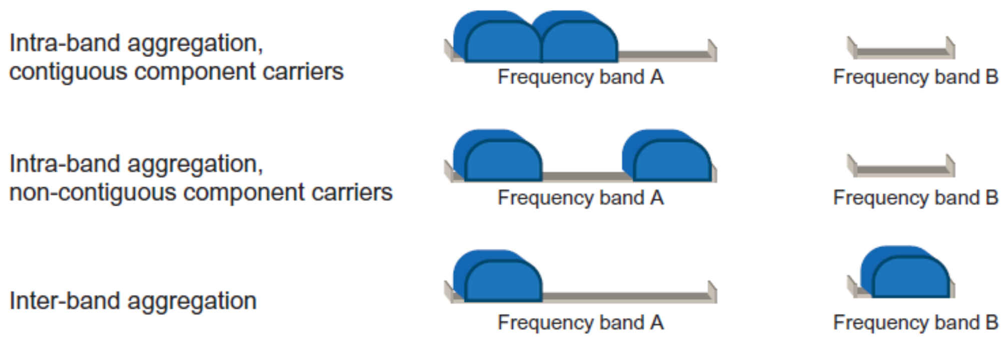


Figure 17.4: Carrier aggregation.

Extended Multi-Antenna Transmission

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

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

Relaying

Relaying implies that the terminal communicates with the network via a relay node that is wirelessly connected to a donor cell using the LTE radio-interface technology (Figure 17.5). From a terminal point of view, the relay node will appear as an ordinary cell. This has the important advantage of simplifying the terminal implementation and making the relay node backwards compatible – that is, also accessible to LTE release-8/9 terminals.

In essence, the relay is a low-power base station wirelessly connected to the remaining part of the network. One of the attractive features of a relay is the LTE-based wireless backhaul, as this could provide a simple way of improving coverage, for example in indoor environments, by simply placing relays at the problematic locations. At a later stage, if motivated by the traffic situation, the wireless donor-relay link could be replaced by an optical fiber in order to use the precious radio resources in the donor cell for terminal communication instead of serving the relay.

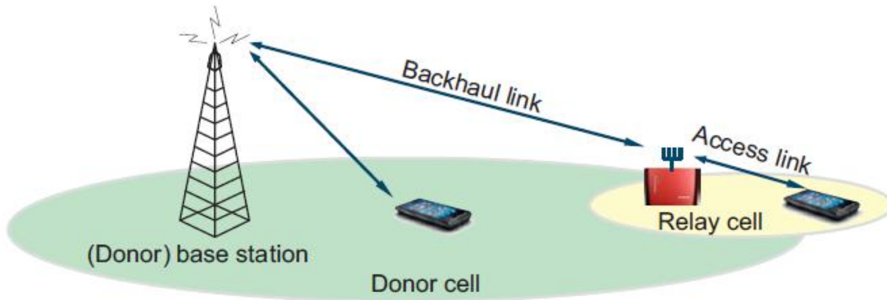


Figure 17.5: Example of relaying.

Heterogeneous Deployments

Heterogeneous deployments refer to deployments with a mixture of cells with different downlink transmission power, operating on (partially) the same set of frequencies and with overlapping geographical coverage (Figure 17.6). A typical example is a pico cell placed within the coverage area of a macro cell. Although such deployments were already supported in release 8, release 10 introduced improved inter-cell interference handling focusing on scenarios with large power differences between overlapping cells.

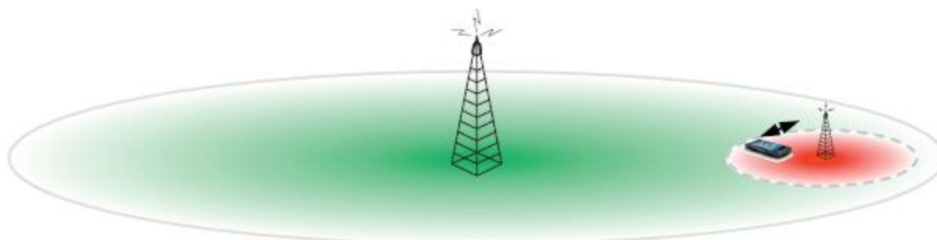


Figure 17.6: Example of heterogeneous deployment with a pico cell inside a macro cell.

Terminal Capabilities

In total five different UE categories have been specified for LTE release 8/9, ranging from the low-end category 1 not supporting spatial multiplexing to the high-end category 5 supporting the full set of features in the release-8/9 physical layer specifications. The categories are summarized in Table 17.1. Note that, regardless of the category, a terminal is always capable of receiving transmissions from up to four antenna ports.

This is necessary as the system information can be transmitted on up to four antenna ports.

In LTE release 10, features such as carrier aggregation and uplink spatial multiplexing are introduced, which calls for additional capability signaling compared to release 8/9, either in the form of additional UE categories or as separate capabilities.

Defining new categories for each foreseen combination of the maximum number of component carriers and maximum degree of spatial multiplexing could be done in principle, although the number of categories might become very large and which categories a terminal support may be frequency-band dependent. Therefore, in release 10, three additional

UE categories were defined as seen in Table 17.1, and the maximum number of component carriers and degree of spatial multiplexing supported, both in uplink and downlink, are signaled separately from the category number. A release-10 terminal may therefore declare itself as, for example, category 4 but capable of uplink spatial multiplexing. Hence, categories 1–5 may have a slightly different meaning for a release-8/9 and a release-10 terminal, depending on the value of the separately declared capabilities. Furthermore, in order to be able to operate in release-8/9 networks, a release-10.

UE has to be able to declare both release-8/9 and release-10 categories. In addition to the capabilities mentioned in the different UE categories, there are some capabilities specified outside the categories. The duplexing schemes supported is one such example, the support of UE-specific reference signals for FDD in release 8 is another. Whether the terminal supports other radio-access technologies, for example GSM and WCDMA, is also declared separately.

Table 17.1: UE Categories.

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

17.4 LTE RADIO-INTERFACE ARCHITECTURE

Overall System Architecture

In parallel to the work on the LTE radio-access technology in 3GPP, the overall system architecture of both the Radio-Access Network (RAN) and the Core Network (CN) was revisited, including the split of functionality between the two network parts. This work was known as the System Architecture Evolution (SAE) and resulted in a flat RAN architecture, as well as a new core network architecture referred to as the Evolved Packet Core (EPC). Together, the LTE RAN and the EPC can be referred to as the Evolved Packet System (EPS). The RAN is responsible for all radio-related functionality of the overall network including, for example, scheduling, radio-resource handling, retransmission protocols, coding and various multi-antenna schemes.

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

Core Network

The EPC is a radical evolution from the GSM/GPRS core network used for GSM and WCDMA/ HSPA. EPC supports access to the packet-switched domain only, with no access to the circuit switched domain. It consists of several different types of nodes, some of which are briefly described below and illustrated in Figure PII 3.6.

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

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

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

In addition, the EPC also contains other types of nodes such as Policy and Charging Rules Function (PCRF) responsible for quality-of-service (QoS) handling and charging, and the Home Subscriber Service

(HSS) node, a database containing subscriber information. There are also some additional nodes present as regards network support of Multimedia Broadcast Multicast Services (MBMS).

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

Radio-Access Network

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

As can be seen in Figure17.7, the eNodeB is connected to the EPC by means of the S1 interface, more specifically to the S-GW by means of the S1 user-plane part, S1-u, and to the MME by means of the S1 control-plane part, S1-c. One eNodeB can be connected to multiple MMEs/S-GWs for the purpose of load sharing and redundancy.

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

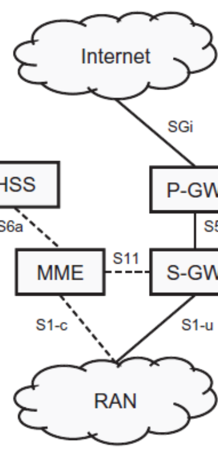


Figure 17.6: Core-network (EPC) architecture.

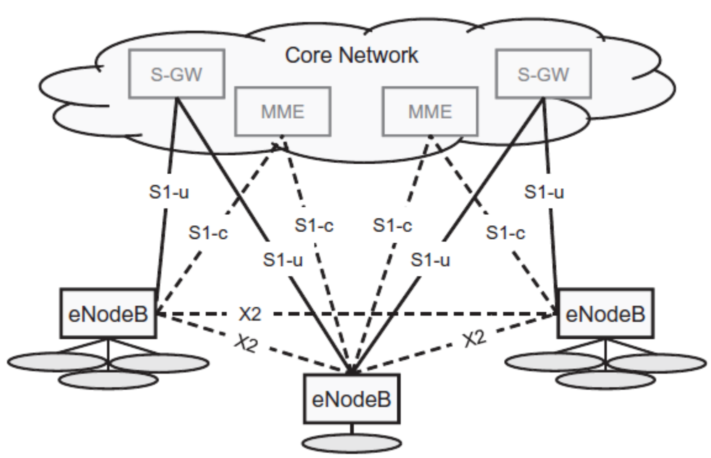


Figure 17.7: Radio-access-network interfaces.

With the overall network architecture in mind, the RAN protocol architecture for the user as well as the control planes can be discussed. Figure 17.8 illustrates the RAN protocol architecture. As seen in the figure, many of the protocol entities are common to the user and control planes. Therefore, although this section mainly describes the protocol architecture from a user-plane perspective, the description is in many respects also applicable to the control plane.

The LTE radio-access network provides one or more Radio Bearers to which IP packets are mapped according to their Quality-of-Service requirements. A general overview of the LTE (user plane) protocol architecture for the downlink is illustrated in Figure 17.9. Neither MAC scheduling nor hybrid ARQ with soft combining is used for broadcast of the basic system information. The LTE protocol structure related to uplink transmissions is similar to the downlink structure in Figure PII 3.9, although there are some differences with respect to, for example, transport-format selection. The different protocol entities of the radio-access network are summarized below.

Packet Data Convergence Protocol (PDCP) performs IP header compression to reduce the number of bits to transmit over the radio interface. The header-compression mechanism is based on Robust Header Compression (ROHC), a standardized header-compression algorithm also used for several mobile-communication technologies. PDCP is also responsible for ciphering and, for the control plane, integrity protection of the transmitted data, as well as in-sequence delivery and duplicate removal for handover. At the receiver side, the PDCP protocol performs the corresponding deciphering and decompression operations. There is one PDCP entity per radio bearer configured for a terminal.

Radio-Link Control (RLC) is responsible for segmentation/concatenation, retransmission handling, duplicate detection, and in-sequence delivery to higher layers. The RLC provides services to the PDCP in the form of radio bearers. There is one RLC entity per radio bearer configured for a terminal.

Medium-Access Control (MAC) handles multiplexing of logical channels, hybrid-ARQ retransmissions, and uplink and downlink scheduling. The scheduling functionality is located in the eNodeB for both uplink and downlink. The hybrid-ARQ protocol part is present in both the transmitting and receiving ends of the MAC protocol. The MAC provides services to the RLC in the form of logical channels.

Physical Layer (PHY) handles coding/ decoding, modulation/demodulation, multi-antenna mapping, and other typical physical-layer functions. The physical layer offers services to the MAC layer in the form of transport channels.

The flow of downlink data through all the protocol layers is summarized in an example illustration for a case with three IP packets, two on one radio bearer and one on another radio bearer and is given in Figure 17.10. The data flow in the case of uplink transmission is similar. The PDCP performs (optional) IP-header compression, followed by ciphering. A PDCP header is added, carrying information required for deciphering in the terminal. The output from the PDCP is forwarded to the RLC.

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

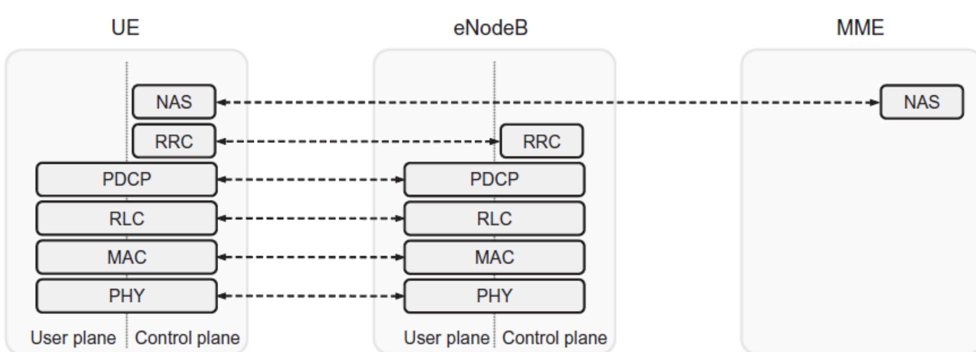


Figure 17.8: Overall RAN protocol architecture.

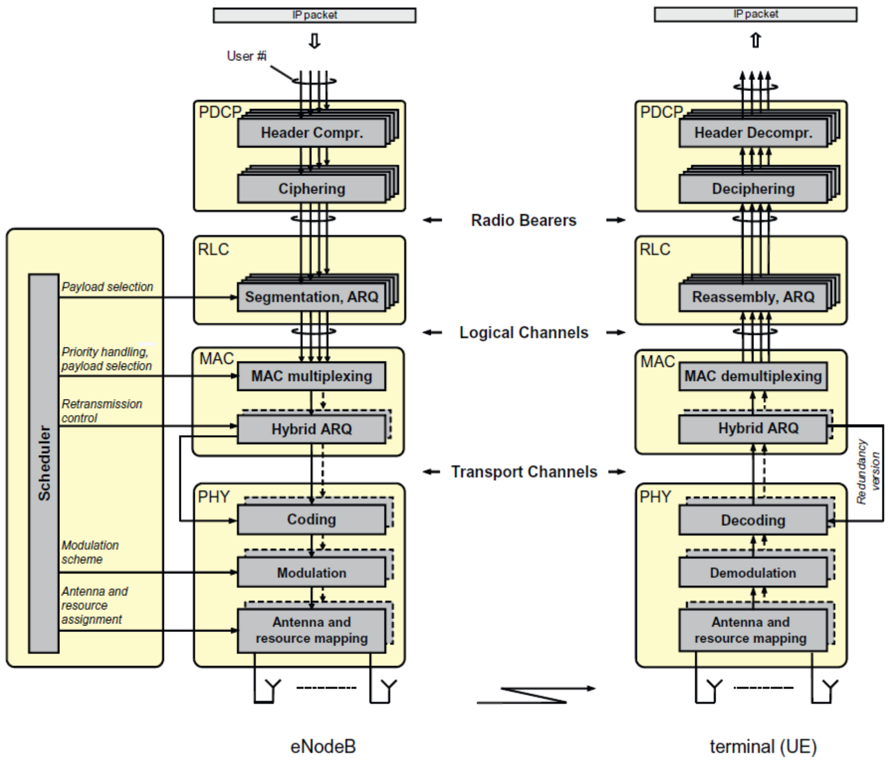


Figure 17.9: LTE protocol architecture (downlink).

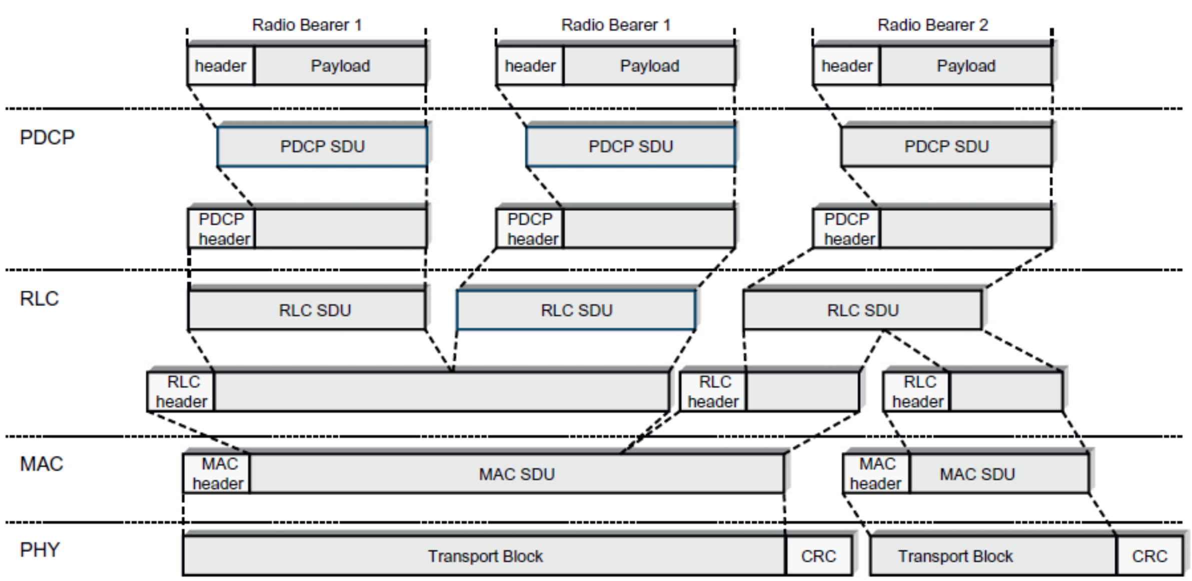


Figure 17.10: Example of LTE data flow.

17.6 PHYSICAL LAYER

The physical layer is responsible for coding, physical-layer hybrid-ARQ processing, modulation, multi-antenna processing, and mapping of the signal to the appropriate physical time–frequency resources. It also handles mapping of transport channels to physical channels, as shown in Figures 17.11 and 17.12. The physical layer provides services to the MAC layer in the form of transport channels. Data transmission in downlink and uplink use the DL-SCH and UL-SCH transport-channel types respectively. There is at most one or, in the case of spatial multiplexing, two transport blocks per TTI on a DL-SCH or UL-SCH. In the case of carrier aggregation, there is one DL-SCH (or UL-SCH) per component carrier.

A physical channel corresponds to the set of time–frequency resources used for transmission of a particular transport channel and each transport channel is mapped to a corresponding physical channel, as shown in Figures 17.11 and 17.12. In addition to the physical channels with a corresponding transport channel, there are also physical channels without a corresponding transport channel.

These channels, known as L1/L2 control channels, are used for downlink control information (DCI), providing the terminal with the necessary information for proper reception and decoding of the downlink data transmission, and uplink control information (UCI) used for providing the scheduler and the hybrid-ARQ protocol with information about the situation at the terminal. The physical-channel types defined in LTE include the following:

The Physical Downlink Shared Channel (PDSCH) is the main physical channel used for unicast data transmission, but also for transmission of paging information.

The Physical Broadcast Channel (PBCH) carries part of the system information, required by the terminal in order to access the network.

The Physical Multicast Channel (PMCH) is used for MBSFN operation.

The Physical Downlink Control Channel (PDCCH) is used for downlink control information, mainly scheduling decisions, required for reception of PDSCH, and for scheduling grants enabling transmission on the PUSCH.

The Physical Hybrid-ARQ Indicator Channel (PHICH) carries the hybrid-ARQ acknowledgement to indicate to the terminal whether a transport block should be retransmitted or not.

The Physical Control Format Indicator Channel (PCFICH) is a channel providing the terminals with information necessary to decode the set of PDCCHs. There is only one PCFICH per component carrier.

The Physical Uplink Shared Channel (PUSCH) is the uplink counterpart to the PDSCH. There is at most one PUSCH per uplink component carrier per terminal.

The Physical Uplink Control Channel (PUCCH) is used by the terminal to send hybrid-ARQ acknowledgements, indicating to the eNodeB whether the downlink transport block(s) was successfully received or not, to send channel-state reports aiding downlink channel-dependent scheduling, and for requesting resources to transmit uplink data upon. There is at most one PUCCH per terminal.

The Physical Random-Access Channel (PRACH) is used for random access.

Note that some of the physical channels, more specifically the channels used for downlink control information (PCFICH, PDCCH, PHICH) and uplink control information (PUCCH), do not have a corresponding transport channel. The remaining downlink transport channels are based on the same general physical-layer processing as the DL-SCH, although with some restrictions in the set of features used. This is especially true for PCH and MCH transport channels.

For the broadcast of system information on the BCH, a terminal must be able to receive this information channel as one of the first steps prior to accessing the system. Consequently, the transmission format must be known to the terminals a priori and there is no dynamic control of any of the transmission parameters from the MAC layer in this case. The BCH is also mapped to the physical resource (the OFDM time–frequency grid) in a different way.

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

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

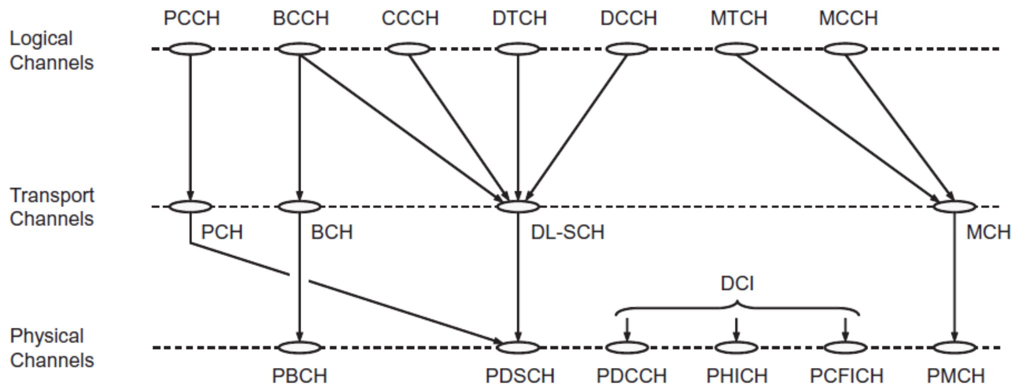


Figure 17.11: Downlink channel mapping.

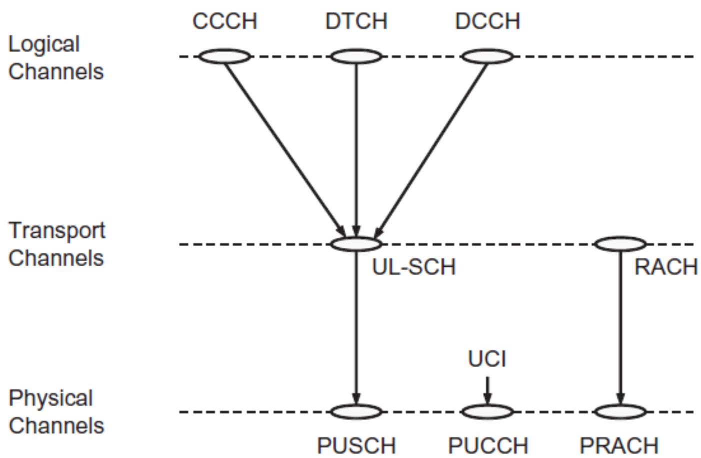


Figure 17.12: Uplink channel mapping.

17.7 OVERALL TIME–FREQUENCY STRUCTURE

OFDM is the basic transmission scheme for both the downlink and uplink transmission directions in LTE although, for the uplink, specific means are taken to reduce the cubic metric of the transmitted signal, thereby allowing for improved efficiency for the terminal transmitter power amplifier. Thus, for data transmission, DFT precoding is applied before OFDM modulation, leading to DFT-spread OFDM or DFTS-OFDM. It should be noted though that DFTS-OFDM is only applied to uplink data transmission – that is, for the transmission of the UL-SCH transport channel.

For other uplink transmissions, such as the transmission of L1/L2 control signaling and different types of reference-signal transmissions, other means are taken to limit the cubic metric of the transmitted signal. The LTE OFDM subcarrier spacing equals 15 kHz for both downlink and uplink. As discussed in Chapter 3, the selection of the subcarrier spacing in an OFDM-based system needs to carefully balance overhead from the cyclic prefix against sensitivity to Doppler spread/shift and other types of frequency errors and inaccuracies. The choice of 15 kHz for the LTE subcarrier spacing was found to offer a good balance between these two constraints.

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

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

In the time domain, LTE transmissions are organized into (radio) frames of length 10 ms, each of which is divided into ten equally sized subframes of length 1 ms, as illustrated in Figure PII 3.13. Each subframe consists of two equally sized slots of length $T_{slot} = 0.5$ ms, with each slot consisting of a number of OFDM symbols including cyclic prefix.

To provide consistent and exact timing definitions, different time intervals within the LTE specifications are defined as multiples of a basic time unit $T_s = 1/(15000 \cdot 2048)$. The basic time unit Ts can thus be seen as the sampling time of an FFT-based transmitter/receiver implementation with an FFT size equal to 2048. The time intervals outlined in Figure PII 3.13 can thus also be expressed as $T_{frame} = 307\,200 \cdot Ts$, $T_{subframe} = 30720 \cdot T_s$, and $T_{slot} = 15\,360 \cdot Ts$ for the frame, subframe, and slot durations respectively.

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

SFN repeats itself after 1024 frames or roughly 10 seconds. The 15 kHz LTE subcarrier spacing corresponds to a useful symbol time $T_u = 2048 \cdot T_s$ or approximately 66.7 μs. The overall OFDM symbol time is then the sum of the useful symbol time and the cyclic-prefix length T_{CP}.

As illustrated in Figure 17.13, LTE defines two cyclic-prefix lengths, the normal cyclic prefix and an extended cyclic prefix, corresponding to seven and six OFDM symbols per slot respectively. The exact cyclic-prefix lengths, expressed in the basic time unit T_s, are given in Figure 17.13. It can be noted that, in the case of the normal cyclic prefix, the cyclic-prefix length for the first OFDM symbol of a slot is somewhat larger compared to the remaining OFDM symbols.

The reason for this is simply to fill the entire 0.5 ms slot, as the number of basic time units T_s per slot (15360) is not divisible by seven. It should be noted that different cyclic-prefix lengths may be used for different subframes within a frame. As an example MBSFN-based multicast/broadcast transmission is always confined to a limited set of subframes, in which case the use of the extended cyclic prefix, with its associated additional cyclic-prefix overhead, may only be applied to these subframes.

A resource element, consisting of one subcarrier during one OFDM symbol, is the smallest physical resource in LTE. Furthermore, as illustrated in Figure 17.14, resource elements are grouped into resource blocks, where each resource block consists of 12 consecutive subcarriers in the frequency domain and one 0.5 ms slot in the time domain. Each resource block thus consists of $7 \cdot 12 = 84$ resource elements in the case of a normal cyclic prefix and $6 \cdot 12 = 72$ resource elements in the case of an extended cyclic prefix.

Although resource blocks are defined over one slot, the basic time-domain unit for dynamic scheduling in LTE is one subframe, consisting of two consecutive slots. The reason for defining the resource blocks over one slot is that distributed downlink transmission and uplink frequency hopping are defined on a slot or resource-block basis.

The minimum scheduling unit, consisting of two time-consecutive resource blocks within one subframes (one resource block per slot), can be referred to as a resource-block pair. The LTE physical-layer specification allows for a carrier to consist of any number of resource blocks in the frequency domain, ranging from a minimum of six resource blocks up to a maximum of 110 resource blocks.

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

The resource-block definition above applies to both the downlink and uplink transmission directions. However, there is a minor difference between the downlink and uplink in terms of where the carrier center frequency is located in relation to the subcarriers. In the downlink (upper part of Figure 17.15), there is an unused DC-subcarrier that coincides with the carrier center frequency. The reason why the DC-subcarrier is not used for downlink transmission is that it may be subject to disproportionately high interference due, for example, to local-oscillator leakage.

On the other hand, in the uplink (lower part of Figure 17.15), no unused DC-subcarrier is defined and the center frequency of an uplink carrier is located between two uplink subcarriers. The presence of an unused DC-carrier in the center of the spectrum would have prevented the assignment of the entire cell bandwidth to a single terminal and still retain the assumption of mapping to consecutive inputs of the OFDM modulator, something that is needed to retain the low-cubic-metric property of the DFTS-OFDM modulation used for uplink data transmission.

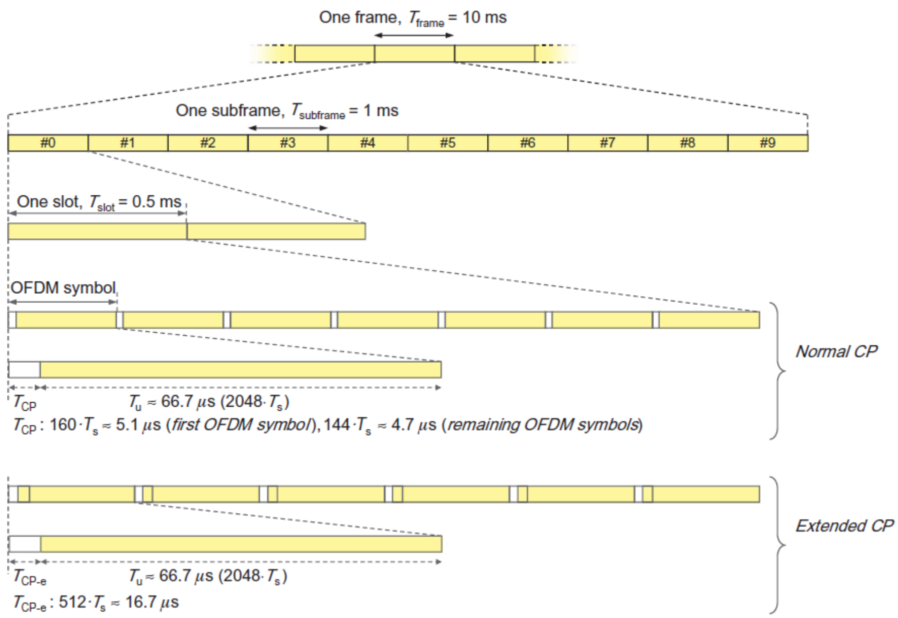


Figure 17.13: LTE time-domain structure.

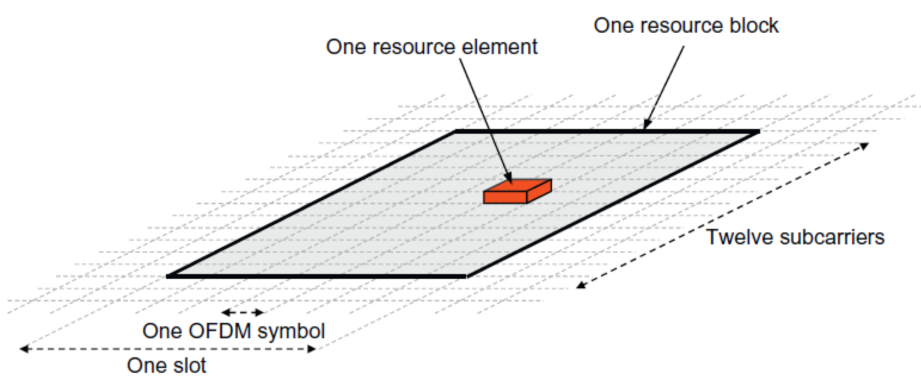


Figure 17.14: The LTE physical time–frequency resource.

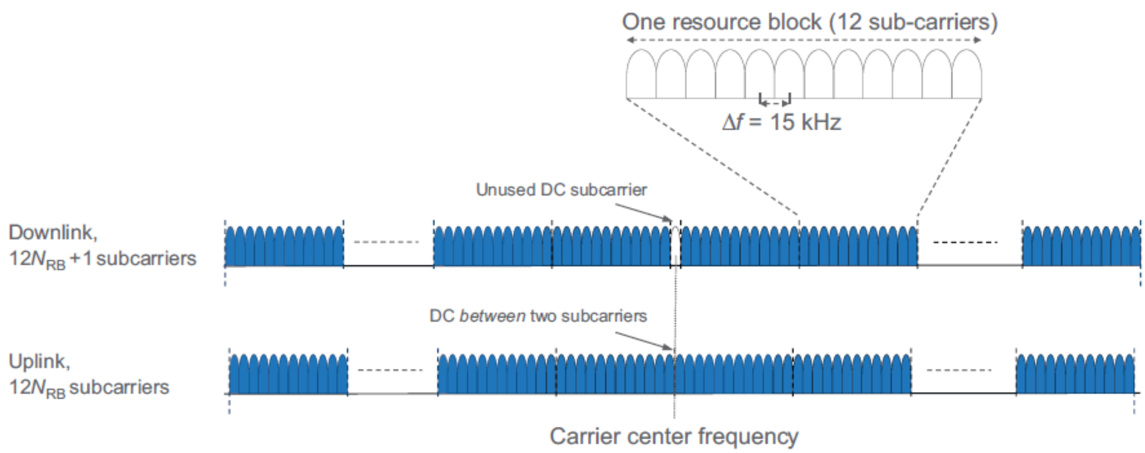


Figure 17.15: Frequency-domain structure for LTE.

17.8 POWER CONTROL, SCHEDULING, AND INTERFERENCE HANDLING

Uplink Power Control

Uplink power control for LTE is the set of algorithms and tools by which the transmit power for different uplink physical channels and signals are controlled to ensure that they, if possible, are received at the cell site with appropriate power. This means that the transmission should be received with sufficient power to allow for proper demodulation of the corresponding information. At the same time, the transmit power should not be unnecessarily high as that would cause unnecessary interference to other cells.

The transmit power will thus depend on the channel properties, including the channel attenuation and the noise and interference level at the receiver side. Furthermore, in the case of DL-SCH transmission on PDSCH, if the received power is too low one can either increase the transmit power or reduce the data rate by use of rate control. Thus, in this case there is an intimate relation between power control and rate control.

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

In principle, each physical channel is separately and independently power controlled. However, in the case of multiple physical channels to be transmitted in parallel from the same terminal, the total power to be transmitted for all physical channels may, in some cases, exceed the maximum terminal output power P_{TMAX} corresponding to the terminal power class.

The basic strategy is then to first ensure that transmission of any L1/L2 control signaling is assigned the power assumed to be needed for reliable transmission. The remaining available power is then assigned to the remaining physical channels. For each uplink component carrier configured for a terminal there is also an associated and explicitly configured maximum per-carrier transmit power $P_{CMAX,c}$, which may be different for different component carriers (indicated by the index c).

Furthermore, although it obviously does not make sense for $P_{CMAX,c}$ to exceed the maximum terminal output power P_{TMAX}, the sum of $P_{CMAX,c}$ for all configured component carriers may very well, and typically will, exceed P_{TMAX}. The reason is that, in many cases, the terminal will not be scheduled for uplink transmission on all its configured component carriers and the terminal should also in that case be able to transmit with its maximum output power.

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

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

If PUCCH is not transmitted in the subframe but L1/L2 control signaling is multiplexed on to PUSCH, the PUSCH carrying the L1/L2 control signaling is first assigned the power determined by its corresponding power-control algorithm, before any power is assigned to any other PUSCH to be transmitted in parallel. Once again, this ensures that L1/L2 control signaling is assigned the power assumed to be needed before any power is assigned for other PUSCH transmissions only carrying UL-SCH. Note that, in the case of transmission of multiple PUSCH in parallel (carrier aggregation), at most one PUSCH may include L1/L2 control signaling. Also, there cannot be PUCCH

transmission and L1/L2 control signaling multiplexed on to PUSCH in the same subframe. Thus, there will never be any conflict between the above rules. If the remaining available transmit power is not sufficient to fulfill the power requirements of any remaining PUSCH to be transmitted, the powers of these remaining physical channels, which only carry UL-SCH, are scaled so that the total power for all physical channels to be transmitted does not exceed the maximum terminal output power.

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

$$\sum_{i} w_c \cdot P_{\mathrm{PUSCH},c} \leq P_{\mathrm{TMAX}} - P_{\mathrm{PUCCH}},$$

where $P_{\mathrm{PUSCH},c}$ is the transmit power for PUSCH on carrier c as determined by the power-control algorithm (before power scaling but including the per-carrier limitation $P_{\mathrm{CMAX},c}$), PPUCCH is the transmit power for PUCCH (which is zero if there is no PUCCH transmission in the subframe), and wc is the power-scaling factor for PUSCH on carrier c ($w_c \leq 1$). For any PUSCH carrying L1/L2 control signaling the scaling factor wc should be set to 1.

For the remaining PUSCH, some scaling factors may be set to zero by decision of the terminal, in practice implying that the PUSCH, as well as the corresponding UL-SCH mapped to the PUSCH, are not transmitted. For the remaining PUSCH the scaling factors w_c are set to the same value less than or equal to 1 to ensure that the above inequality is fulfilled. Thus, all PUSCH that are actually transmitted are power scaled by the same factor.

Power Control for PUCCH

For PUCCH, the appropriate received power is simply the power needed to achieve a desired – that is, a sufficiently low – error rate in the decoding of the L1/L2 control information transmitted on the PUCCH. Overall, power-control for PUCCH can be described by the following expression:

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

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

The parameter $P_{0,\mathrm{PUCCH}}$ in expression is a cell-specific parameter that is broadcast as part of the cell system information. Considering only the part $P_{0,\mathrm{PUCCH}}$. P_{LDL} in the PUCCH power-control expression and assuming that the (estimated) downlink path loss accurately reflects the true uplink path loss, it is obvious that $P_{0,\mathrm{PUCCH}}$ can be seen as the desired or target received power.

The required received power will depend on the uplink noise/interference level. From this point of view, the value of $P_{0,\mathrm{PUCCH}}$ should take the interference level into account and thus vary in time as the interference level varies. However, in practice it is not feasible to have $P_{0,\mathrm{PUCCH}}$ varying with the instantaneous interference level. One simple reason is that the terminal does not read the system information continuously and thus the terminal would anyway not have access to a fully up-to-date $P_{0,\mathrm{PUCCH}}$ value. Another reason is that the uplink path-loss estimates derived from downlink measurements will anyway not be fully accurate, for example due to differences between the instantaneous downlink and uplink path loss, as well as due to measurement inaccuracies.

Thus, in practice, $P_{0,\mathrm{PUCCH}}$ may reflect the average interference level, or perhaps only the relatively constant noise level. More rapid interference variations can then be taken care of by closedloop power control, see below. Finally, it is possible for the network to directly adjust the PUCCH transmit power by providing the terminal with explicit power-control commands that adjust the term δ in the power-control expression above. These power-control commands are accumulative – that is, each received power-control command increases or decreases the term δ by a certain amount.

Power Control for PUSCH

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

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

where M indicates the instantaneous PUSCH bandwidth measured in number of resource blocks and the term ΔMCS is similar to the term ΔFormat in the expression for PUCCH power control – that is, it reflects the fact that different SINR is required for different modulation schemes and coding rates used for the PUSCH transmission. The above expression is clearly similar to the power-control expression for PUCCH transmission, with some key differences:

The use of "$P_{CMAX,c}$. $P_{0,PUCCH}$" reflects the fact that the transmit power available for PUSCH on a carrier is the maximum allowed per-carrier transmit power after power has been assigned to any PUCCH transmission on that carrier. This ensures priority of L1/L2 signaling on PUCCH over data transmission on PUSCH in the power assignment.

The term 10. log10(M) reflects the fact that what is fundamentally controlled by the parameter $P_{0,PUSCH}$ is the power per resource block. For a larger resource assignment, a correspondingly higher received power and thus a correspondingly higher transmit power is needed.2 The parameter α, which can take a value smaller than or equal to 1, allows for so-called partial path-loss compensation, as described below. In general, the parameters $P_{0,PUSCH}$, α, and ΔMCS can be different for the different component carriers configured for a terminal.

Power Control for SRS

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

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

where the parameters P0,PUSCH, α, and δ are the same as for PUSCH power control, as discussed in Section 13.1.3. Furthermore, MSRS is the bandwidth, expressed as number of resource blocks, of the SRS transmission and PSRS is a configurable offset.

Power Headroom

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

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

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

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

where the values for M and ΔMCS correspond to the resource assignment and modulation-and-coding scheme used in the subframe to which the power-headroom report corresponds. It can be noted that the power headroom is not a measure of the difference between the maximum per-carrier transmit power and the actual carrier transmit power. Rather, comparing with expression (13.3) it can be seen that the power headroom is a measure of the difference between $P_{CMAX,c}$ and the transmit power that would have been used assuming that there would have been no upper limit on the transmit power. Thus, the power headroom can very well be negative.

More exactly, a negative power headroom indicates that the per-carrier transmit power was limited by $P_{CMAX,c}$ at the time of the power headroom reporting. As the network knows what modulation-and-coding scheme and resource size the terminal used for transmission in the subframe to which the power-headroom report corresponds, it can determine what are the valid combinations of modulation-and-coding scheme and resource size M, assuming that the downlink path loss P_{LDL} and the term δ have not changed substantially.

17.9 SCHEDULING AND RATE ADAPTATION

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

For carrier aggregation, each component carrier is independently scheduled with individual scheduling assignments/grants and one DL-SCH/UL-SCH per scheduled component carrier. Semipersistent scheduling is only supported on the primary component carriers, motivated by the fact that the main usage is for small payloads not requiring multiple component carriers. The downlink scheduler is responsible for dynamically controlling the terminal(s) to transmit to and, for each of these terminals, the set of resource blocks upon which the terminal's DL-SCH (or DL-SCHs in the case of carrier aggregation) is transmitted. Transport-format selection (selection of transport-bock size, modulation-and-coding scheme, resource-block allocation, and antenna mapping) for each component carrier and logical channel multiplexing for downlink transmissions are controlled by the eNodeB, as illustrated in the left part of Figure 17.16.

The uplink scheduler serves a similar purpose, namely to dynamically control which terminals are to transmit on their UL-SCH (or UL-SCHs in the case of carrier aggregation) and on which uplink resources.

The uplink scheduler is in complete control of the transport format the terminal will use, whereas the logical-channel multiplexing is controlled by the terminal according to a set of rules. Thus, uplink scheduling is per terminal and not per radio bearer. This is illustrated in the right part of Figure PII 3.16, where the scheduler controls the transport format and the terminal controls the logical-channel multiplexing.

Downlink Scheduling

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

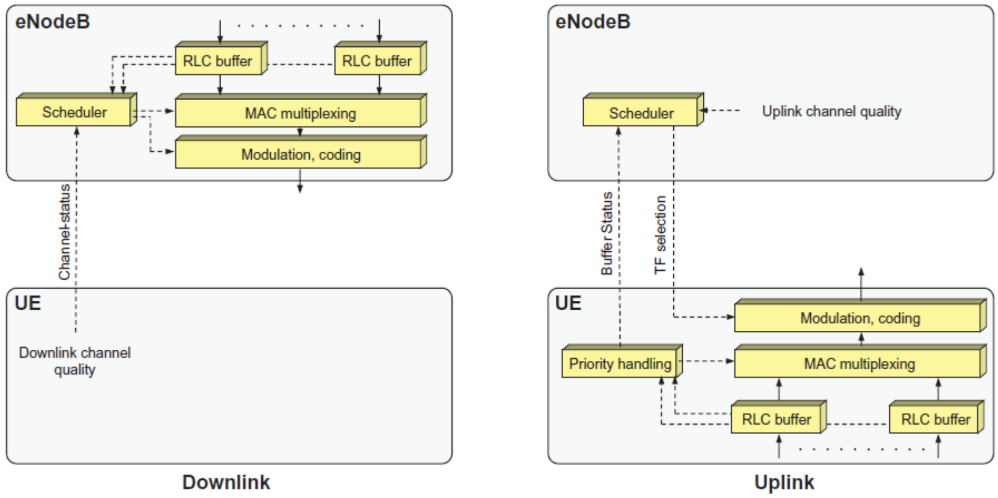


Figure 17.16: Transport-format selection in downlink (left) and uplink (right).

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

RLC. Segmentation/concatenation of RLC SDUs is directly related to the instantaneous data rate. For low data rates, it may only be possible to deliver a part of an RLC SDU in a TTI, in which case segmentation is needed. Similarly, for high data rates, multiple RLC SDUs may need to be concatenated to form a sufficiently large transport block.

MAC. Multiplexing of logical channels depends on the priorities between different streams. For example, radio resource control signaling, such as handover commands, typically has a higher priority than streaming data, which in turn has higher priority than a background file transfer. Thus, depending on the data rate and the amount of traffic of different priorities, the multiplexing of different logical channels is affected. Hybrid-ARQ retransmissions also need to be accounted for.

L1. Coding, modulation and, if applicable, the number of transmission layers and the associated precoding matrix are obviously affected by the scheduling decision. The choices of these parameters are mainly determined by the radio conditions and the selected data rate – that is, the transport block size.

The scheduling decision is communicated to each of the scheduled terminals through the downlink L1/L2 control signaling as described in Chapter 10, using one PDCCH per downlink assignment. Each terminal monitors a set of PDCCHs for downlink scheduling assignments. A scheduling assignment is transmitted in the same subframe as the data. If a valid assignment matching the identity of the terminal is found, then the terminal receives and processes the transmitted signal as indicated in the assignment. Once the transport block is successfully decoded, the terminal will demultiplex the received data into the appropriate logical channels.

In the case of carrier aggregation, there is one PDCCH per component carrier. Furthermore, if cross-carrier scheduling is configured, the downlink assignment does not have to be transmitted on the same component carrier as the associated data, as information about the component carrier containing the associated data is included in the scheduling assignment in this case. The scheduling strategy is implementation specific and not part of the 3GPP specifications. In principle. However, the overall goal of most schedulers is to take advantage of the channel variations between terminals and preferably to schedule transmissions to a terminal when the channel conditions are advantageous. Most scheduling strategies therefore need information about:

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

In addition to the channel-state information, the scheduler should take buffer status and priority levels into account. Obviously it does not make sense to schedule a terminal with empty transmission buffers. Priorities of the different types of traffic may also vary; RRC signaling may be prioritized over user data.

Furthermore, RLC and hybrid-ARQ retransmissions, which are in no way different from other types of data from a scheduler perspective, are typically also given priority over initial transmissions.

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

Uplink Scheduling

The basic function of the uplink scheduler is similar to its downlink counterpart, namely to dynamically determine, for each 1 ms interval, which terminals are to transmit and on which uplink resources. As discussed before, the LTE uplink is primarily based on maintaining Orthogonality between different uplink transmissions and the shared resource controlled by the eNodeB scheduler is time–frequency resource units. In addition to assigning the time–frequency resources to the terminal, the eNodeB scheduler is also responsible for controlling the transport format the terminal will use for each of the uplink component carriers. As the scheduler knows the transport format the terminal will use when it is transmitting, there is no need for outband control signaling from the terminal to the eNodeB.

This is beneficial from a coverage perspective; taking into account that the cost per bit of transmitting outbands control information can be significantly higher than the cost of data transmission, as the control signaling needs to be received with higher reliability. It also allows the scheduler to tightly control the uplink activity to maximize the resource usage compared to schemes where the terminal autonomously selects the data rate, as autonomous schemes typically require some margin in the scheduling decisions.

A consequence of the scheduler being responsible for selection of the transport format is that accurate and detailed knowledge about the terminal situation with respect to buffer status and power availability is more accentuated in LTE compared to systems where the terminal autonomously controls the transmission parameters.

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

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

Similarly to the downlink case, the uplink scheduler can exploit information about channel conditions, buffer status, and priorities of the different data flows, and, if some form of interference coordination is employed, the interference situation in neighboring cells. Channel-dependent scheduling, which typically is used for the downlink, can be used for the uplink as well. In the uplink, estimates of the channel quality can be obtained from the use of uplink channel sounding.

Semi-Persistent Scheduling

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

With semi-persistent scheduling, the terminal is provided with the scheduling decision on the PDCCH, together with an indication that this applies to every nth subframe until further notice. Hence, control signaling is only used once and the overhead is reduced, as illustrated in Figure PII 3.17. The periodicity of semi-persistently scheduled transmissions – that is, the value of n – is configured by RRC signaling in advance, while activation (and deactivation) is done using the PDCCH using the semi-persistent C-RNTI.3 For example, for voice-over IP the scheduler can configure a periodicity of 20 ms for semi-persistent scheduling and, once a talk spurt starts, the semi-persistent pattern is triggered by the PDCCH.

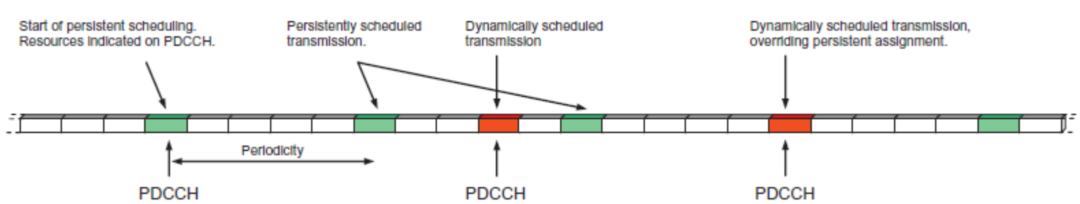
After enabling semi-persistent scheduling, the terminal continues to monitor the PDCCH for uplink and downlink scheduling commands. When a dynamic scheduling command is detected, it takes precedence over the semi-persistent scheduling in that particular subframe, which is useful if the semi-persistently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

For the downlink, only initial transmissions use semi-persistent scheduling. Retransmissions are explicitly scheduled using a PDCCH assignment. This follows directly from the use of an asynchronous hybrid-ARQ protocol in the downlink. Uplink retransmissions, in contrast, can either follow the semi-persistently allocated subframes or be dynamically scheduled. Semi-persistent scheduling is only supported on the primary component carrier and any transmission on a secondary component carrier must be dynamically scheduled. This is reasonable as semi-persistent scheduling is intended for low-rate services for which a single component carrier is sufficient. Figure PII 3.16: Example of semi-persistent scheduling.

17.10 CHANNEL-STATE REPORTING

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

Rank indication (RI), providing a recommendation on the transmission rank to use or, expressed differently, the number of layers that should preferably be used for downlink transmission to the terminal. RI only needs to be reported by terminals that are configured to be in one of the spatialmultiplexing transmission modes. There is at most one RI reported, valid across the full bandwidth – that is, the RI is frequency non-selective. Frequency-dependent transmission rank would be impossible to utilize since all layers are transmitted on the same set of resource blocks in LTE.



Precoder matrix indication (PMI), indicating which of the precoder matrices should preferably be used for the downlink transmission. The reported precoder matrix is determined assuming the number of layers indicated by the RI. The precoder recommendation may be frequency selective, implying that the terminal may recommend different precoders for different parts of the downlink spectrum. Furthermore, the network can restrict the set of matrices from which the terminal should select the recommended precoder, so-called codebook subset restriction, to avoid reporting precoders that are not useful in the antenna setup used.

Channel-quality indication (CQI), representing the highest modulation-and-coding scheme that, if used, would mean PDSCCH transmissions (using the recommended RI and PMI) were received with a block-error rate of at most 10%. The reason to use CQI as a feedback quantity instead of, for example, the signal-to-noise ratio, is to account for different receiver implementation in the terminal. Also, basing the feedback reports on CQI instead of signal-to-noise ratio also simplifies the testing of terminals; a terminal delivering data with more than 10% block-error probability when using the modulation-and-coding scheme indicated by the CQI would fail the test. As will be discussed further below, multiple CQI reports, each representing the channel quality in a certain part of the downlink spectrum, can be part of a channel-state report.

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

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

- The network may follow the latest terminal recommendation, in which case the eNodeB only has to confirm (a one-bit indicator in the downlink scheduling assignment) that the precoder configuration recommended by the terminal is used for the downlink transmission. On receiving such a confirmation, the terminal will use its recommended configuration when demodulating and decoding the corresponding DL-SCH transmission. Since the PMI computed in the terminal can be frequency selective, an eNodeB following the precoding matrix recommended by the terminal may have to apply different precoding matrices for different (sets of) resource blocks.

- The network may select a different precoder, information about which then needs to be explicitly included in the downlink scheduling assignment. The terminal then uses this configuration when demodulating and decoding the DL-SCH. To reduce the amount of downlink signaling, only a single precoding matrix can be signaled in the scheduling assignment, implying that, if the network overrides the recommendation, then the precoding is frequency non-selective. The network may also choose to override the transmission rank only, in which case the terminal assumes that a subset of the columns in each of the recommended precoder matrices is used.

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

Aperiodic channel-state reports are delivered when explicitly requested by the network by means of the channel-state-request flag included in uplink scheduling grants. An aperiodic channel-state report is always delivered using the PUSCH – that is, on a dynamically assigned resource.

Periodic channel-state reports are configured by the network to be delivered with a certain periodicity, possibly as often as once every 2 ms, on a semi-statically configured PUCCH resource. However, similar to hybrid-ARQ acknowledgements normally delivered on PUCCH, channelstate reports are "re-routed" to the PUSCH4 if the terminal has a valid uplink grant and is anyway to transmit on the PUSCH.

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

17.11 INTER-CELL INTERFERENCE COORDINATION

Like all modern mobile-communication technologies, LTE can be deployed with one-cell frequency reuse. Fundamentally, this implies that the LTE transmission structure has been designed so that reliable transmission is also possible for the low signal-to-interference ratios (SIR) that may occur in a reuse-one deployment when the same time–frequency resource issued in neighboring cells (SIR as low as -5 dB or even somewhat lower in the worst case6). This is especially true for transmission of critical information such as system information and L1/L2 control signaling. It should be noted that data transmission on DL-SCH can always be made sufficiently reliable by selecting a sufficiently low instantaneous data rate in combination with hybrid-ARQ retransmissions. Still, a one-cell frequency reuse typically implies relatively large SIR variations over the cell area.

As a consequence, the data rates that can be offered to the end-user may also vary substantially, with nly relatively low data rates being available at the “cell edge”. If one was only interested in maximizing the data rates that could be offered to users at the cell edge – that is, maximizing the “worst-case-user” quality – a reuse larger than one could actually be preferred. For cell-edge users receiving high interference from neighboring cells, the negative impact of reduced bandwidth availability due to frequency reuse larger than one could be more than compensated for by the higher SIR and corresponding higher achievable data rate per MHz, leading to overall higher achievable data rates. However, the overall system efficiency would be degraded as the majority of users are not at the cell edge and thus have a relatively good SIR even with one-cell reuse.

For such users, any further SIR improvement due to larger reuse would typically not be able to compensate for the reduced bandwidth availability for at least two reasons:

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

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

Thus, the basic mode of operation should be one-cell reuse, giving each cell access to the overall available spectrum. However, system performance, and especially the quality for cell-edge users, could be further enhanced if one could at least partly coordinate the scheduling between neighboring cells. The basic principle of such inter-cell interference coordination (ICIC) would be to, if possible, avoid high-power transmission on time–frequency resources on which cell-edge users are scheduled in neighboring cells, users that would otherwise experience high interference and correspondingly low data rates. This kind of “selective” interference avoidance would benefit the cell-edge-user quality and could also enhance overall system performance. 295

The above discussion implicitly assumed downlink transmission with terminals at the cell edge being interfered by downlink transmissions from other cells. The concept of inter-cell interference coordination is equally applicable to the uplink, although the interference situation in this case is somewhat different. For the uplink, the interference level experienced by a certain link does not depend on where the transmitting terminal is located, but rather on the location of the interfering terminals, with interfering terminals closer to the cell border causing more interference to neighboring cells.

The location of the transmitting terminal is still important though, as a terminal closer to the cell site can raise its transmission power to compensate for high interference coming from terminals in neighboring cells, something that may not be possible for cell-edge terminals. Thus, the fundamental goal of uplink inter-cell interference is the same as for the downlink – that is, to coordinate the, in this case, uplink scheduling between cells to avoid simultaneous transmissions from terminals at the cell border in neighboring cells causing severe interference to each other.

In the case of scheduling located at a higher-level node above the eNodeB, coordinated scheduling between cells of different eNodeB would, at least conceptually, be straightforward. However, in LTE there is no higher-level node and scheduling is carried out locally at the eNodeB – that is, in practice at the cell site.7 Thus, the best that can be done in the radio-access specifications is to introduce messages that convey information about the scheduling strategy between neighboring eNodeBs using the Figure 17.18: Illustration of uplink ICIC based on HII and OI X2 signaling.

X2 interface. An eNodeB can then use the information provided by a neighboring eNodeB as input to its own scheduling process. For LTE, a number of such ICIC-related X2 messages have been defined. To assist uplink interference coordination, two messages are defined, the High Interference Indicator (HII) and the Overload Indicator (OI), see also Figure 17.18. The High Interference Indicator provides information about the set of resource blocks within which an eNodeB is likely to schedule transmissions from cell-edge terminals – that is, resource blocks on which a neighboring cell can expect higher interference. Although nothing is explicitly specified on how an eNodeB should react to the HII (or any other ICIC-related X2 signaling) received from a neighboring eNodeB, a reasonable action for the receiving eNodeB would be to try to avoid scheduling its own cell-edge terminals on the same resource blocks, thereby reducing the uplink interference to cell-edge transmissions in its own cell as well as in the cell from which the HII was received.

The HII can thus be seen as a proactive tool for ICIC, trying to prevent the occurrence of too-low-SIR situations.

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

For the downlink, the Relative Narrowband Transmit Power (RNTP) is defined to support ICIC operation (see Figure 17.19). The RNTP is similar to the HII in the sense that it provides information, for each resource block, whether or not the relative transmit power of that resource block is to exceed a certain level. Similar to the HII, a neighboring cell can use the information provided by the received RNTP when scheduling its own terminals, especially terminals on the cell edge that are more likely to be interfered by the neighboring cell.

One kind of deployment that has recently received some interest is the Centralized RAN (C-RAN, where the baseband processing of the eNodeBs is located in a central office, geographically separate from the actual cell sites. In such a scenario it is more straightforward to coordinate the scheduling between multiple geographically separate cells, either by introducing inter-eNodeB coordination within the central office or by simply deploying a massive eNodeB able to handle all the cells connected to the central office. As an eNodeB is anyway just a logical node, in terms of physical realization these two approaches are essentially the same thing.

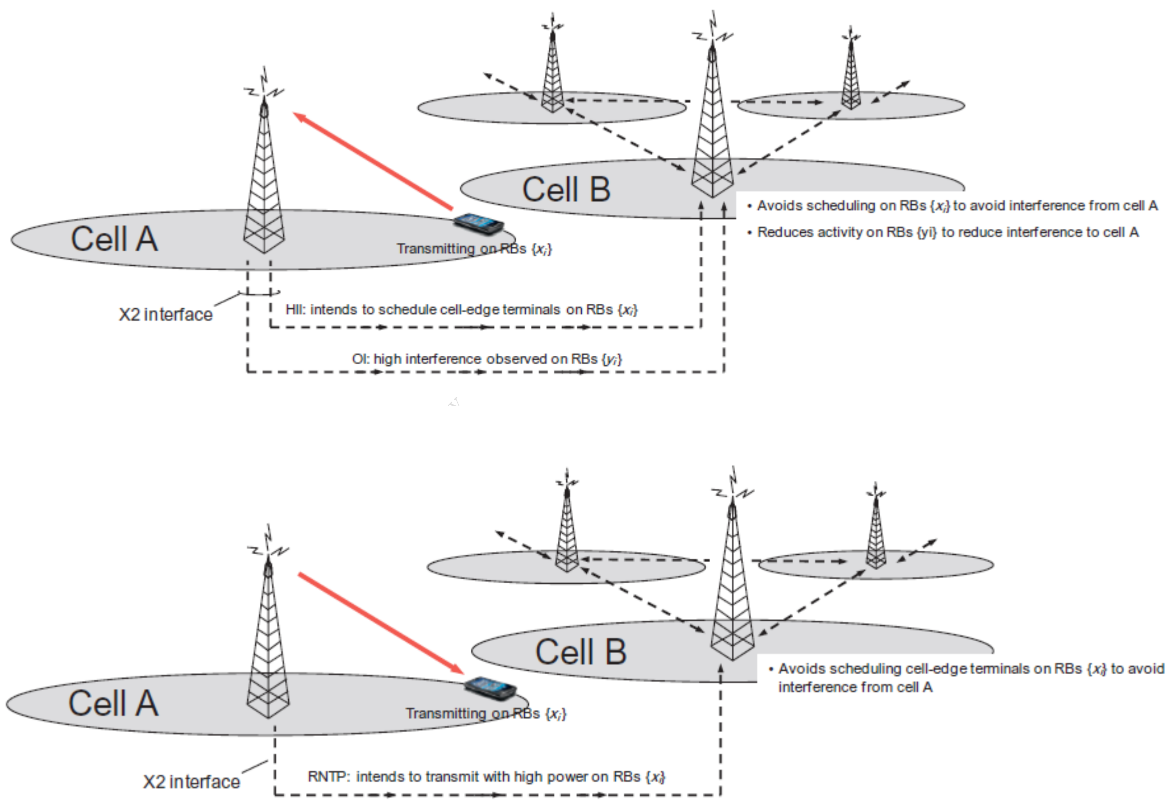


Figure 17.19: Illustration of downlink ICIC based on RNTP X2 signaling.

17.12 HETEROGENEOUS NETWORK DEPLOYMENTS

The continuous increase in traffic within mobile-broadband systems and an equally continuous increase in terms of the data rates requested by end-users will impact how cellular networks are deployed in the future. In general, providing very high system capacity (traffic per m2) and very high per-user data rates will require a densification of the radio-access network – that is, the deployment of additional network nodes. By increasing the number of cells, the traffic per m2 can be increased without requiring a corresponding increase in the traffic that needs to be supported per network node. Also, by increasing the number of network nodes, the base-station-to-terminal distances will, in general, be shorter, implying a link-budget improvement and a corresponding improvement in achievable data rates.

A general densification of the macro-cell layer – that is, reducing the coverage area of each cell and increasing the total number of macro-cell sites8 – as illustrated in the upper part of Figure 17.20, is a path that has already been taken by many operators. As an example, in many major cities the distance between macro-cell sites is often less than a few hundred meters in many cases.

An alternative or complement to a uniform densification of the macro-cell layer is to deploy additional lower-power nodes under the coverage area of a macro cell, as illustrated in the lower part of Figure PII 3.20. In such a heterogeneous or multi-layered network deployment, the underlaid picocell layer does not need to provide full-area coverage. Rather, pico sites can be deployed to increase capacity and achievable data rates where needed. Outside of the pico-layer coverage, terminals would access the network by means of the overlaid macro cell. Another example of heterogeneous network deployment is the complementary use of so-called home-eNodeBs, also often referred to as femto base stations. A home-eNodeB corresponds to a small low-power base station deployed by the end-user, typically within the home, and connecting to the operator network using the end-user's wireline broadband connection.

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

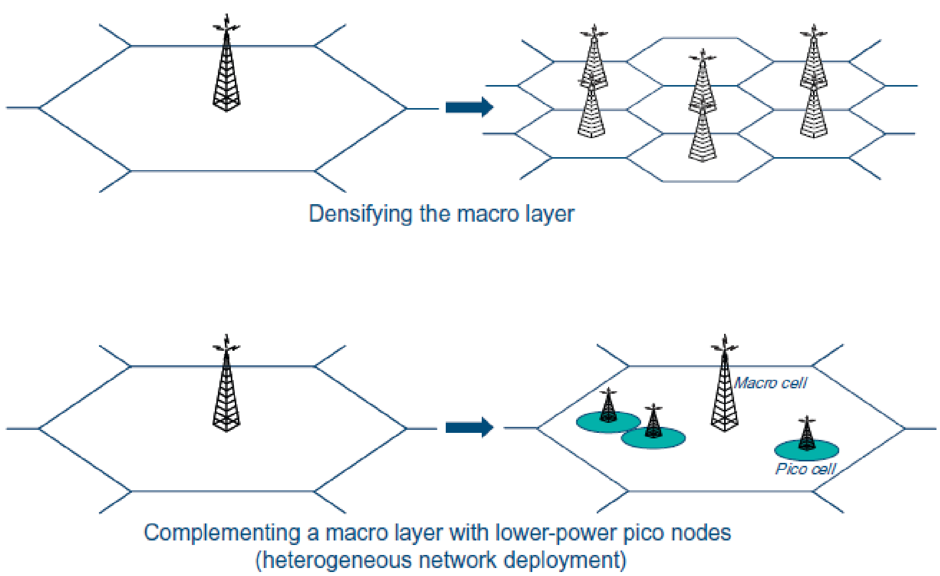


Figure 17.20: Network densification to enhance system capacity and support higher data rates.

Interference Coordination in the Case of Home-eNodeB

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

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

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

Multimedia Broadcast/Multicast Services the past, cellular systems have mostly focused on transmission of data intended for a single user and not on multicast/broadcast services. Broadcast networks, exemplified by the radio and TV broadcasting networks, have on the other hand focused on covering very large areas with the same content and have offered no or limited possibilities for transmission of data intended for a single user.

Multimedia Broadcast Multicast Services (MBMS) support multicast/broadcast services in a cellular system, thereby combining the provision of multicast/broadcast and unicast services within a single network. With MBMS, the same content is transmitted to multiple users located in a specific area, known as the MBMS service area and typically comprising multiple cells. In each cell participating in the transmission, a point-to-multipoint radio resource is configured and all users subscribing to the MBMS service simultaneously receive the same transmitted signal. No tracking of users' movement in the radio-access network is performed and users can receive the content without notifying the network.

When providing multicast/broadcast services for mobile devices there are several aspects to take into account, of which two deserve special attention and will be elaborated upon further below: good coverage and low terminal power consumption. The coverage, or more accurately the data rate possible to provide, is basically determined by the link quality of the worst-case user, as no user-specific adaptation of transmission parameters can be used in a multicast/broadcast system providing the same information to multiple users.

OFDM transmission provides specific benefits for provision of multi-cell multicast/ broadcast services. If the transmissions from the different cells are time synchronized, the resulting signal will, from a terminal point of view, appear as a transmission from a single point over a time dispersive channel. For LTE this kind of transmission is referred to as an MBMS Single-Frequency Network (MBSFN). MBSFN transmission provides several benefits:

- Increased received signal strength, especially at the border between cells involved in the MBSFN transmission, as the terminal can utilize the signal energy received from multiple cells.
- Reduced interference level, once again especially at the border between cells involved in the MBSFN transmission, as the signals received from neighboring cells will not appear as interference but as useful signals.
- Additional diversity against fading on the radio channel as the information is received from several, geographically separated locations, typically making the overall aggregated channel appear highly time-dispersive or, equivalently, highly frequency selective.

Altogether, this allows for significant improvements in the multicast/broadcast reception quality, especially at the border between cells involved in the MBSFN transmission, and, as a consequence, significant improvements in the achievable multicast/broadcast data rates. Providing for power-efficient reception in the terminal in essence implies that the structure of the overall transmission should be such that data for a service-of-interest is provided in short high-data rate bursts rather than longer low-data-rate bursts. This allows the terminal to occasionally wake up to receive data with long periods of DRX in between. In LTE, this is catered for by time-multiplexing unicast and broadcast transmissions, as well as by the scheduling of different MBMS services.

Multimedia Broadcast Multicast Services Architecture

An MBSFN area is a specific area where one or several cells transmit the same content. For example, in Figure 17.21, cells 8 and 9 both belong to MBSFN area C. Not only can an MBSFN area consist of multiple cells, a single cell can also be part of multiple, up to eight, MBSFN areas, as shown in Figure 17.21, where cells 4 and 5 are part of both MBSFN areas A and B. Note that, from an MBSFN reception point of view, the individual cells are invisible, although the terminal needs to be aware of the different cells for other purposes, such as reading system information and notification indicators.

The MBSFN areas are static and do not vary over time. The usage of MBSFN transmission obviously requires not only time synchronization among the cells participating in an MBSFN area, but also usage of the same set of radio resources in each of the cells for a particular service. This coordination is the responsibility of the Multi-cell/multicast Coordination Entity (MCE), which is a logical node in the radio-access network handling allocation of radio resources and transmission parameters (time–frequency resources and transport format) across the cells in the MBSFN area. As shown in Figure 17.22, the MCE1 can control multiple eNodeBs, each handling one or more cells.

The Broadcast Multicast Service Center (BM-SC), located in the core network, is responsible for authorization and authentication of content providers, charging, and the overall configuration of the data flow through the core network. The MBMS gateway (MBMS-GW) is a logical node handling multicast of IP packets from the BM-SC to all eNodeBs involved in transmission in the MBSFN area.

It also handles session control signaling via the MME. From the BM-SC, the MBMS data is forwarded using IP multicast, a method of sending an IP packet to multiple receiving network nodes in a single transmission, via the MBMS gateway to the cells from which the MBMS transmission is to be carried out. Hence, MBMS is not only efficient from a radio-interface perspective, but it also saves resources in the transport network by not having to send the same packet to multiple nodes individually unless necessary. This can lead to significant savings in the transport network.

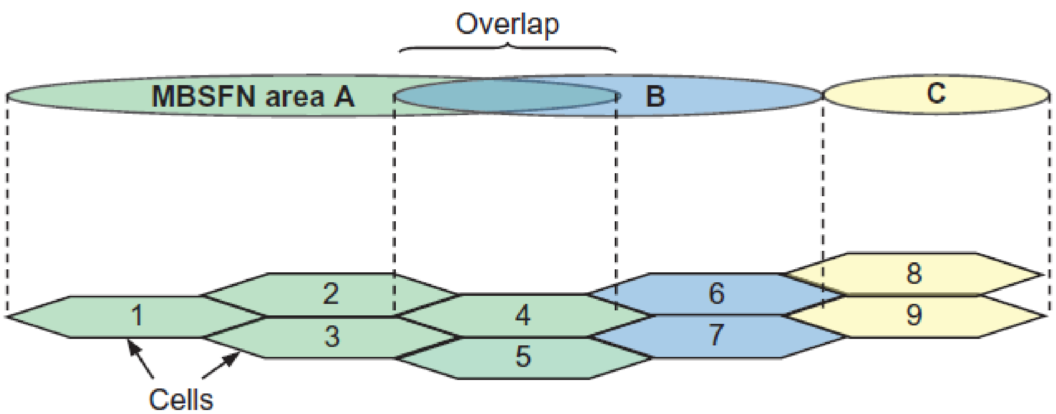


Figure17.20: Example of MBSFN areas.

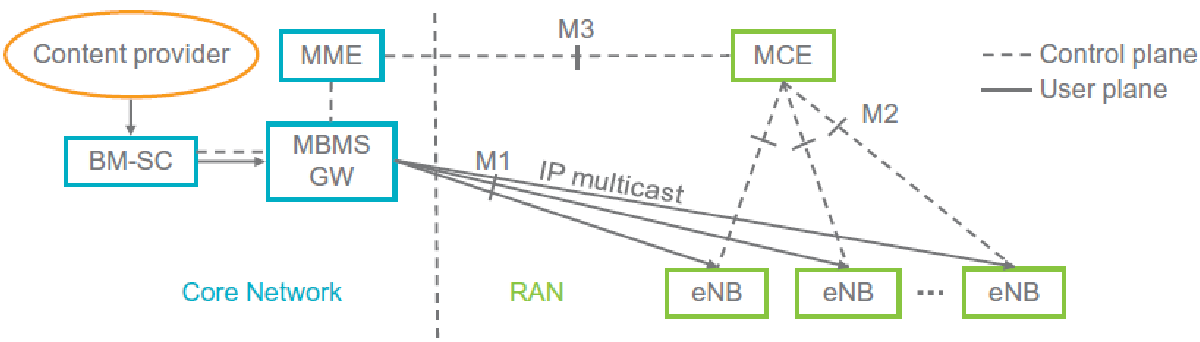


Figure PII 3.21: LTE MBMS architecture.

17.13 LTE RELAYING

The possibility of a terminal communicating with the network, and the data rate that can be used, depends on several factors, the path loss between the terminal and the base station being one. The link performance of LTE is already quite close to the Shannon limit and from a pure link-budget perspective, the highest data rates supported by LTE require a relatively high signal-to-noise ratio. Unless the link budget can be improved, for example with different types of beam-forming solutions, a denser infrastructure is required to reduce the terminal-to-base-station distance and thereby improve the link budget.

A denser infrastructure is mainly a deployment aspect, but in later releases of LTE, various tools enhancing the support for low-power base stations are included. One of these tools is relaying, which can be used to reduce the distance between the terminal and the infrastructure, resulting in an improved link budget and an increased possibility for high data rates. In principle this reduction in terminal-to-infrastructure distance could be achieved by deploying traditional base stations with a wired connection to the rest of the network. However, relays with a shorter deployment time can often be an attractive alternative, as there is no need to deploy a specific backhaul. A wide range of relay types can be envisioned, some of which could already be deployed in release 8.

Amplify-and-forward relays, commonly referred to as repeaters, simply amplify and forward the received analog signals and are, on some markets, relatively common as a tool for handling coverage holes.

Traditionally, once installed, repeaters continuously forward the received signal regardless of whether there is a terminal in their coverage area or not, although more advanced repeaters can be considered as well. Repeaters are transparent to both the terminal and the base station and can therefore be introduced in existing networks. The fact that the basic principle of a repeater is to amplify whatever it receives, including noise and interference as well as the useful signal, implies that repeaters are mainly useful in high-SNR environments. Expressed differently, the SNR at the output of the repeater can never be higher than at the input.

Decode-and-forward relays decode and re-encode the received signal prior to forwarding it to the served users. The decode-and-re-encode process results in this class of relays not amplifying noise and interference, as is the case with repeaters. They are therefore also useful in low-SNR environments. Furthermore, independent rate adaptation and scheduling for the base station–relay and relay– terminal links is possible. However, the decode-and-re-encode operation implies a larger delay than for an amplify-and-forward repeater, longer than the LTE subframe duration of 1 ms. As for repeaters, many different options exist depending on supported features (support of more than two hops, support for mesh structures, etc) and, depending on the details of those features, a decode-and-forward relay may or may not be transparent to the terminal.

Relays in LTE

LTE release 10 introduces support for a decode-and-forward relaying scheme (repeaters require no additional standardization support other than RF requirements and are available already in release 8). A basic requirement in the development of LTE relaying solutions was that the relay should be transparent to the terminal – that is, the terminal should not be aware of whether it is connected to a relay or to a conventional base station. This ensures that release-8/9 terminals can also be served by relays, despite relays being introduced in release 10. Therefore, so-called self-backhauling was taken as the basis for the LTE relaying solution. In essence, from a logical perspective, a relay is an eNodeB wirelessly connected to the rest of the radio-access network by using the LTE radio interface. It is important to note that, even though the relay from a terminal perspective is identical to an eNodeB, the physical implementation may differ significantly, from a traditional base station, for example in terms of output power.

In conjunction with relaying, the terms backhaul link and access link are often used to refer to the base station–relay connection and the relay–terminal connection respectively. The cell to which the relay is connected using the backhaul link is known as the donor cell and the donor cell may, in addition to one or several relays, also serve terminals not connected via a relay. This is illustrated in Figure17.22.

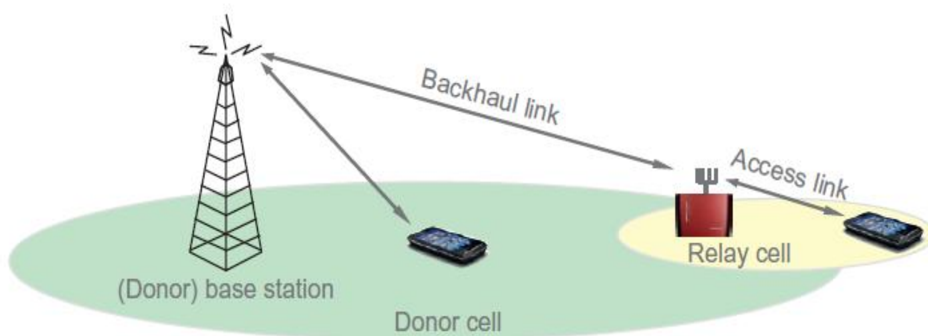


Figure 17.22: Access and backhaul links.

Since the relay communicates both with the donor cell and terminals served by the relay, interference between the access and backhaul links must be avoided. Otherwise, since the power difference between access-link transmissions and backhaul-link reception at the relay can easily be more than 100 dB, the possibility of receiving the backhaul link may be completely ruined. Similarly, transmissions on the backhaul link may cause significant interference to the reception of the access link. These two cases are illustrated in Figure 17.23. Therefore, isolation between the access and backhaul links is required, isolation that can be obtained in one or several of the frequency, time, and/or spatial domains.

Depending on the spectrum used for access and backhaul links, relaying can be classified into outband and inband types. Outband relaying implies that the backhaul operates in a spectrum separate from that of the access link, using the same radio interface as the access link. Provided that the frequency separation between the backhaul and access links is sufficiently large, interference between the backhaul and access links can be avoided and the necessary isolation is obtained in the frequency domain. Consequently, no enhancements to the release-8 radio interface are needed to operate an outband relay. There are no restrictions on the activity on the access and backhaul links and the relay can in principle operate with full duplex.

Inband relaying implies that the backhaul and access links operate in the same spectrum. Depending on the deployment and operation of the relay, this may, as the access and backhaul link share the same spectrum, require additional mechanisms to avoid interference between the access and backhaul links. Unless this interference can be handled by proper antenna arrangements, for example with the relay deployed in a tunnel with the backhaul antenna placed outside the tunnel, a mechanism to separate activity on the access and backhaul links in the time domain is required. Such a mechanism was introduced as part of release 10 and will be described in more detail in the following. Since the backhaul and access links are separated in the time domain, there is a dependency on the transmission activity and the two links cannot operate simultaneously.

Figure 17.23: Interference between access and backhaul links.

Relay Overall Architecture

From an architectural perspective, a relay can, on a high level, be thought of as having a "base-station side" and a "terminal side". Towards terminals, it behaves as a conventional eNodeB using the access link, and a terminal is not aware of whether it is communicating with a relay or a "traditional" base station. Relays are therefore transparent for the terminals and terminals from the first LTE release, release 8, can also benefit from relays. This is important from an operator's perspective, as it allows a gradual introduction of relays without affecting the existing terminal fleet.

Towards the donor cell, a relay initially operates as a terminal, using the LTE radio interface to connect to the donor cell. Once connection is established and the relay is configured, the relay uses a subset of the "terminal side" functionality for communication on the backhaul link. In this phase, the relay-specific enhancements described in this chapter may be used for the backhaul.

In release 10, the focus is on two-hop relaying and scenarios with a relay connected to the network via another relay are not considered. Furthermore, relays are stationary – that is, handover of a relay from one donor cell to another donor cell is not supported. The case for using mobile relays is not yet clear and therefore it was decided in release 10 not to undertake the relatively large task of adapting existing core-network procedures to handle cells that are moving over time, something that could have been a consequence of a mobile relay.

The overall LTE relaying architecture is illustrated in Figure 17.24. One key aspect of the architecture is that the donor eNodeB acts as a proxy between the core network and the relay. From a relay perspective, is appears as if it is connected directly to the core network as the donor eNodeB appears as an MME for the S1 interface and an eNodeB for X2 towards the relay.

From a core-network perspective, on the other hand, the relay cells appear as if they belong to the donor eNodeB. It is the task of the proxy in the donor eNodeB to connect these two views. The use of a proxy is motivated by the desire to minimize the impact to the core network from the introduction of relays, as well as to allow for features such as tight coordination of radio-resource management between the donor eNodeB and the relay.

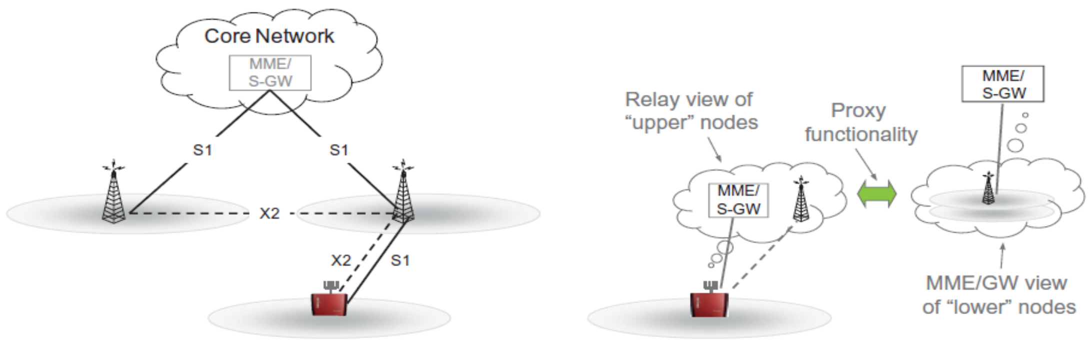


Figure 17.24: LTE relaying architecture.

17.14 LTE SPECTRUM

Spectrum flexibility is a key feature of LTE radio access and is set out in the LTE design targets. It consists of several components, including deployment in differentsized spectrum allocations and deployment in diverse frequency ranges, both in paired and unpaired frequency bands.

There are a number of frequency bands identified for mobile use and specifically for IMT today. Most of these bands were already defined for operation with WCDMA/HSPA, and LTE is the next technology to be deployed in those bands. Both paired and unpaired bands are included in the LTE specifications. The additional challenge with LTE operation in some bands is the possibility of using channel bandwidths up to 20 MHz with a single carrier and even beyond that with aggregated carriers.

The use of OFDM in LTE gives flexibility both in terms of the size of the spectrum allocation needed and in the instantaneous transmission bandwidth used. The OFDM physical layer also enables frequency-domain scheduling. Properties also impact the RF implementation in terms of filters, amplifiers, and all other RF components that are used to transmit and receive the signal. This means that the RF requirements for the receiver and transmitter will have to be expressed with flexibility in mind.

Spectrum for LTE

LTE can be deployed both in existing IMT bands and in future bands that may be identified. The possibility of operating radio-access technology in different frequency bands is, in itself, nothing new. For example, quad-band GSM terminals are common, capable of operating in the 850, 900, 1800, and 1900 MHz bands. From a radio-access functionality perspective, this has no or limited impact and the LTE physical-layer specifications do not assume any specific frequency band. What may differ, in terms of specification, between different bands are mainly the more specific RF requirements, such as the allowed maximum transmit power, requirements/limits on out-of-band (OOB) emission, and so on. One reason for this is that external constraints, imposed by regulatory bodies, may differ between different frequency bands.

Frequency Bands for LTE

The frequency bands where LTE will operate are in both paired and unpaired spectrum, requiring flexibility in the duplex arrangement. For this reason, LTE supports both FDD and TDD.

Release 8 of the 3GPP specifications for LTE includes 19 frequency bands for FDD and nine for TDD. The paired bands for FDD operation are numbered from 1 to 21, as shown in Table 17.3, while the unpaired bands for TDD operation are numbered from 33 to 41, as shown in Table 17.4. Note that the frequency bands for UTRA FDD use the same numbers as the paired LTE bands, but are labeled with Roman numerals. All bands for LTE are summarized in Figures 17.25 and 17.26, which also show the corresponding frequency allocation defined by the ITU.

Some of the frequency bands are partly or fully overlapping. In most cases this is explained by regional differences in how the bands defined by the ITU are implemented. At the same time, a high degree of commonality between the bands is desired to enable global roaming. The set of bands have first been specified as bands for UTRA, with each band originating in global, regional, and local spectrum developments. The complete set of UTRA bands was then transferred to the LTE specifications in release 8 and additional ones have been added in later releases.

Bands 1, 33, and 34 are the same paired and unpaired bands that were defined first for UTRA in release 99 of the 3GPPP specifications, also called the 2 GHz “core band”. Band 2 was added later for operation in the US PCS1900 band and Band 3 for 3 G operation in the GSM1800 band. The unpaired Bands 35, 36, and 37 are also defined for the PCS1900 frequency ranges, but are not deployed anywhere today. Band 39 is an extension of the unpaired Band 33 from 20 to 40 MHz for use in China.

Band 4 was introduced as a new band for the Americas following the addition of the 3 G bands at WRC-2000. Its downlink overlaps completely with the downlink of Band 1, which facilitates roaming and eases the design of dual Band 1 + 4 terminals. Band 10 is an extension of Band 4 from 2 x 45 to 2 x 60 MHz.

Band 9 overlaps with Band 3, but is intended only for Japan. The specifications are drafted in such a way that implementation of roaming dual Band 3 + 9 terminals is possible. The 1500 MHz frequency band is also identified in 3GPP for Japan as Bands 11 and 21. It is allocated globally to mobile service on a co-primary basis and was previously used for 2 G in Japan.

Table 17.3: Paired Frequency Bands Defined by 3GPP for LTE.

Band	Uplink Range (MHz)	Downlink Range (MHz)	Main Region(s)
1	1920–1980	2110–2170	Europe, Asia
2	1850–1910	1930–1990	Americas (Asia)
3	1710–1785	1805–1880	Europe, Asia (Americas)
4	1710–1755	2110–2155	Americas
5	824–849	869–894	Americas
6	830–840	875–885	Japan (only for UTRA)
7	2500–2570	2620–2690	Europe, Asia
8	880–915	925–960	Europe, Asia
9	1749.9–1784.9	1844.9–1879.9	Japan
10	1710–1770	2110–2170	Americas
11	1427.9–1447.9	1475.9–1495.9	Japan
12	698–716	728–746	USA
13	777–787	746–756	USA
14	788–798	758–768	USA
17	704–716	734–746	USA
18	815–830	860–875	Japan
19	830–845	875–890	Japan
20	832–862	791–821	Europe
21	1447.9–1462.9	1495.9–1510.9	Japan

Table 17.4: Unpaired Frequency Bands Defined by 3GPP for LTE.

Band	Frequency Range (MHz)	Main Region(s)
33	1900–1920	Europe, Asia (not Japan)
34	2010–2025	Europe, Asia
35	1850–1910	(Americas)
36	1930–1990	(Americas)
37	1910–1930	–
38	2570–2620	Europe
39	1880–1920	China
40	2300–2400	Europe, Asia
41	2496–2690	USA

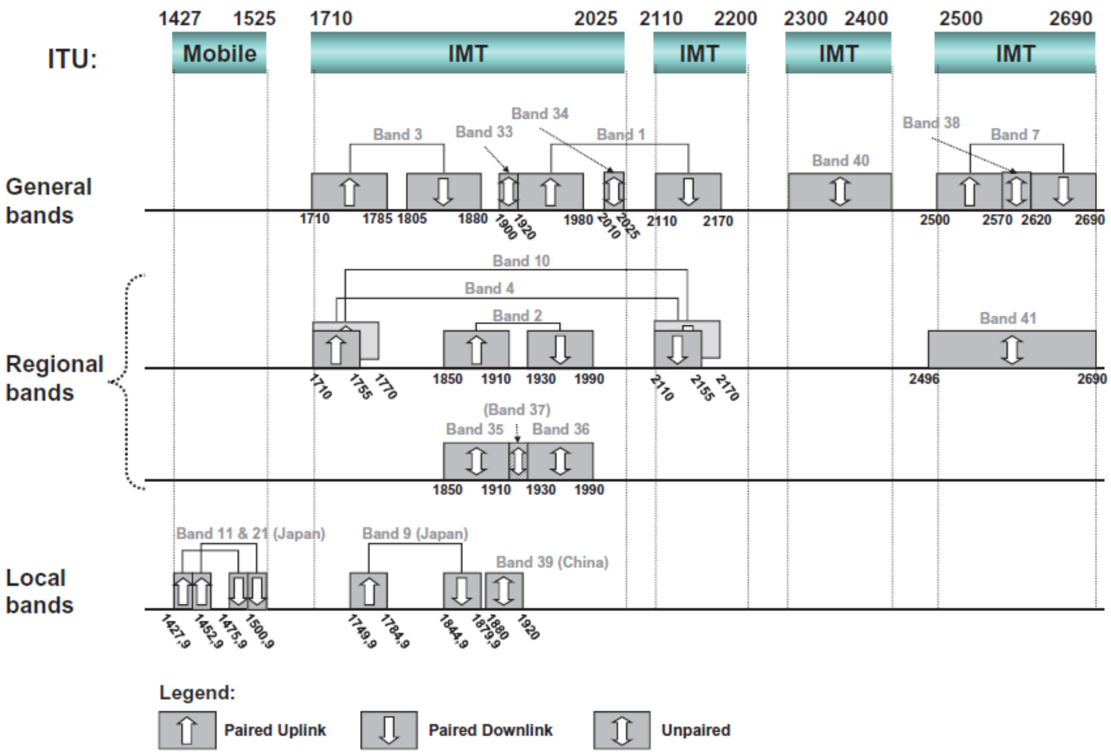


Figure 17.25: Operating bands specified for LTE in 3GPP above 1 GHz and the corresponding ITU allocation.

With WRC-2000, the band 2500–2690 MHz was identified for IMT-2000 and it is identified as Band 7 in 3GPP for FDD and Band 38 for TDD operation in the “center gap” of the FDD allocation. The band has a slightly different arrangement in North America, where a US-specific Band 41 is defined. Band 40 is an unpaired band specified for the new frequency range 2300–2400 MHz identified for IMT and has a widespread allocation globally.

WRC-2000 also identified the frequency range 806–960 MHz for IMT-2000, complemented by the frequency range 698–806 MHz in WRC’07. As shown in Figure PII 3.26, several bands are defined for FDD operation in this range. Band 8 uses the same band plan as GSM900. Bands 5, 18, and 19 overlap, but are intended for different regions. Band 5 is based on the US cellular band, while Bands 18 and 19 are restricted to Japan in the specifications. 2 G systems in Japan had a very specific band plan and Bands 18 and 19 are a way of partly aligning the Japanese spectrum plan in the 810–960 MHz range to that in other parts of the world. Note that Band 6 was originally defined in this frequency range for Japan, but it is not used for LTE.

Bands 12, 13, 14, and 17 make up the first set of bands defined for what is called the digital dividend – that is, for spectrum previously used for broadcasting. This spectrum is partly migrated to be used by other wireless technologies, since TV broadcasting is migrating from analog to more spectrum-efficient digital technologies. Another regional band for the digital dividend is Band 20 that is defined in Europe.

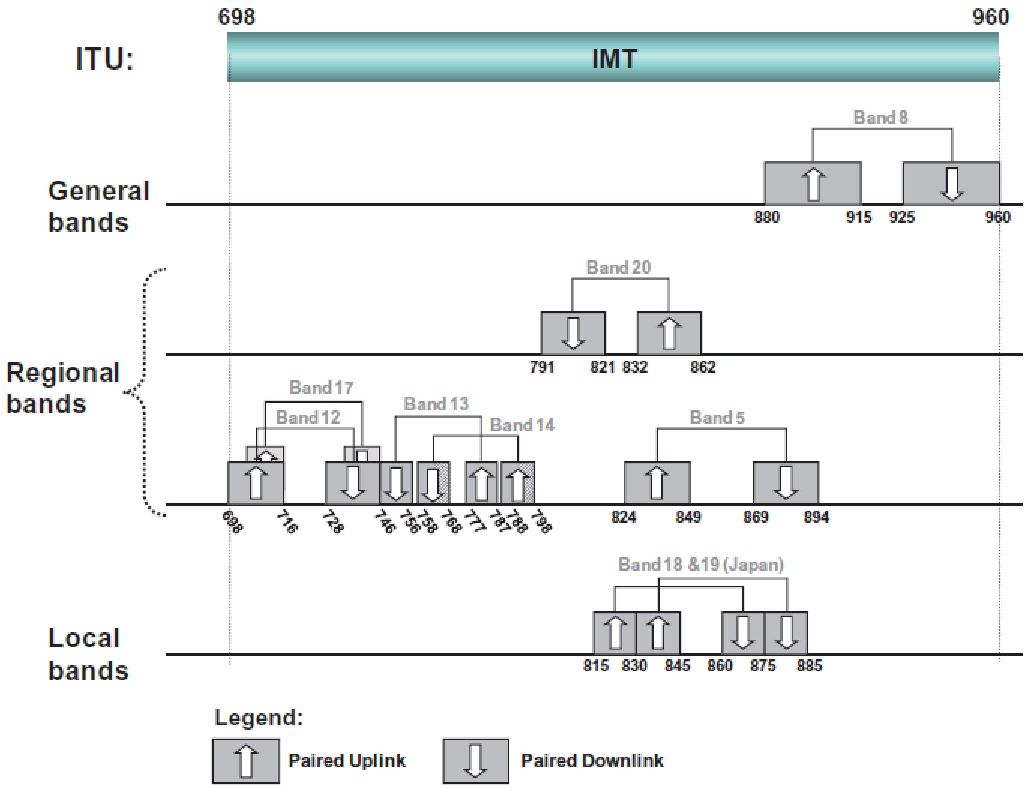


Figure 17.26: Operating bands specified for LTE in 3GPP below 1 GHz and the corresponding ITU allocation.

New Frequency Bands

Additional frequency bands are continuously specified for UTRA and LTE. WRC'07 identified additional frequency bands for IMT, which encompasses both IMT-2000 and IMT-Advanced. Several bands were defined by WRC'07 that will be available partly or fully for deployment on a global basis: 450–470 MHz was identified for IMT globally. It is already allocated to mobile service globally, but it is only 20 MHz wide. 698–806 MHz was allocated to mobile service and identified to IMT to some extent in all regions. Together with the band at 806–960 MHz identified at WRC-2000, it forms a wide frequency range from 698 to 960 MHz that is partly identified to IMT in all regions, with some variations. 2300–2400 MHz was identified for IMT on a worldwide basis in all three regions. 3400–3600 MHz was allocated to the mobile service on a primary basis in Europe and Asia and partly in some countries in the Americas. There is also satellite use in the bands today.

Flexible Spectrum Use

Most of the frequency bands identified above for deployment of LTE are existing IMT-2000 bands and some bands also have legacy systems deployed, including WCDMA/HSPA and GSM. Bands are also in some regions defined in a "technology neutral" manner, which means that coexistence between different technologies is a necessity.

The fundamental LTE requirement to operate in different frequency bands does not, in itself, impose any specific requirements on the radio-interface design. There are, however, implications for the RF requirements and how those are defined, in order to support the following:

Coexistence between operators in the same geographical area in the band. These other operators may deploy LTE or other IMT-2000 technologies, such as UMTS/HSPA or GSM/EDGE. There may also be non-IMT-2000 technologies. Such coexistence requirements are to a large extent developed within 3GPP, but there may also be regional requirements defined by regulatory bodies in some frequency bands.

Co-location of base station equipment between operators. There are in many cases limitations to where base-station equipment can be deployed. Often, sites must be shared between operators or an operator will deploy multiple technologies in one site. This puts additional requirements on both base-station receivers and transmitters.

Coexistence with services in adjacent frequency bands and across country borders. The use of the RF spectrum is regulated through complex international agreements, involving many interests. There will therefore be requirements for coordination between operators in different countries and for coexistence with services in adjacent frequency bands. Most of these are defined in different regulatory bodies. Sometimes the regulators request that 3GPP includes such coexistence limits in the 3GPP specifications.

Coexistence between operators of TDD systems in the same band is provided by inter-operator synchronization, in order to avoid interference between downlink and uplink transmissions of different operators. This means that all operators need to have the same downlink/uplink configurations and frame synchronization, not in itself an RF requirement, but it is implicitly assumed in the 3GPP specifications. RF requirements for unsynchronized systems become very strict.

Release-independent frequency-band principles. Frequency bands are defined regionally and new bands are added continuously. This means that every new release of 3GPP specifications will have new bands added. Through the "release independence" principle, it is possible to design terminals based on an early release of 3GPP specifications that support a frequency band added in a later release.

Flexible Channel Bandwidth Operation

The frequency allocations in 17.25 and 17.26 are up to 2 X 75 MHz, but the spectrum available for a single operator may be from 2 X 20 MHz down to 2 X 5 MHz for FDD and down to 1 X 5 MHz for TDD. Furthermore, the migration to LTE in frequency bands currently used for other radio-access technologies must often take place gradually to ensure that a sufficient amount of spectrum remains to support the existing users. Thus, the amount of spectrum that can initially be migrated to LTE can be relatively small, but may then gradually increase, as shown in Figure 17.27. The variation of 306

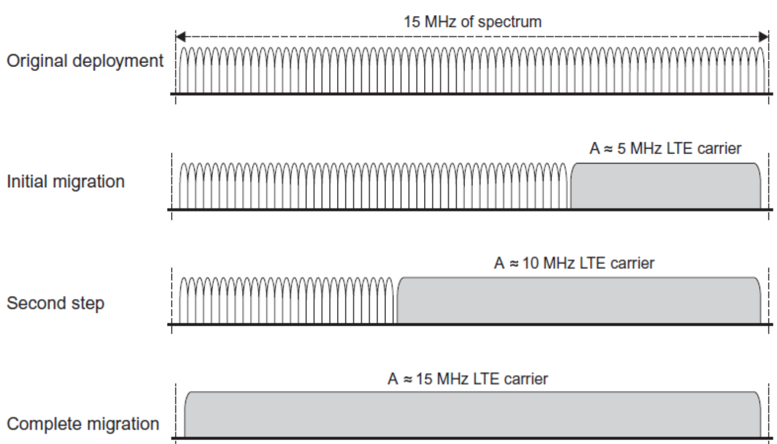


Figure 17.27: Example of how LTE can be migrated step-by-step into a spectrum allocation with an original GSM deployment.

Table 17.5: Channel Bandwidths Specified in LTE.

Channel Bandwidth, $BW_{channel}$ (MHz)	Number of Resource Blocks (N_{RB})
1.4	6
3	15
5	25
10	50
15	75
20	100

possible spectrum scenarios implies a requirement for spectrum flexibility for LTE in terms of the transmission bandwidths supported. The spectrum flexibility requirement points out the need for LTE to be scalable in the frequency domain. This flexibility requirement is stated in as a list of LTE spectrum allocations from 1.25 to 20 MHz. Note that the final channel bandwidths selected differ slightly from this initial assumption.

The frequency-domain structure of LTE is based on resource blocks consisting of 12 subcarriers with a total bandwidth of 12 X 15 kHz =180 kHz. The basic radio-access specification including the physical-layer and protocol specifications enables transmission bandwidth configurations from six up to 110 resource blocks on one LTE RF carrier. This allows for channel bandwidths ranging from 1.4 MHz up to beyond 20 MHz in steps of 180 kHz and is fundamental to providing the required spectrum flexibility.

In order to limit implementation complexity, only a limited set of bandwidths are defined in the RF specifications. Based on the frequency bands available for LTE deployment today and in the future, as described above, and considering the known migration and deployment scenarios in those bands, a limited set of six channel bandwidths is specified. The RF requirements for the base station and UE are defined only for those six channel bandwidths. The channel bandwidths range from 1.4 to 20 MHz, as shown in Table 17.5.

The lower bandwidths, 1.4 and 3 MHz, are chosen specifically to ease migration to LTE in spectrum where CDMA2000 is operated, and also to facilitate migration of GSM and TD-SCDMA to LTE. The specified bandwidths target relevant scenarios in different frequency bands. For this reason, the set of bandwidths available for a specific band is not necessarily the same as in other bands. At a later stage, if new frequency bands are made available that have other spectrum scenarios requiring additional channel bandwidths, the corresponding RF parameters and requirements can be added in the RF specifications, without actually having to update the physical layer specifications. The process of adding new channel bandwidths in this way is similar to adding new frequency bands.

Figure 17.27 illustrates in principle the relationship between the channel bandwidth and the number of resource blocks NRB for one RF carrier. Note that for all channel bandwidths except 1.4 MHz, the resource blocks in the transmission bandwidth configuration fill up 90% of the channel bandwidth.

The spectrum emissions shown in Figure 17.27 are for a pure OFDM signal, while the actual transmitted emissions will also depend on the transmitter RF chain and other components. The emissions outside the channel bandwidth are called unwanted emissions.

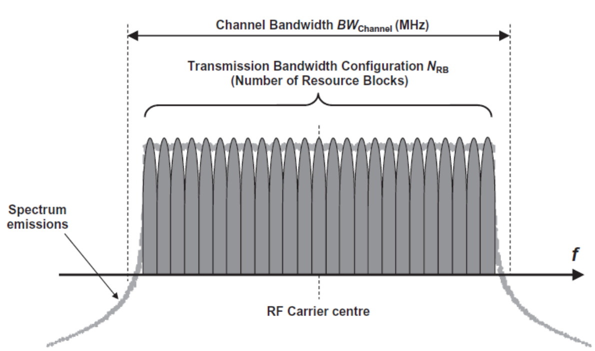


Figure 3.27: The channel bandwidth for one RF carrier and the corresponding transmission bandwidth configuration.

17.15 CARRIER AGGREGATION FOR LTE

The possibility in release 10 to aggregate two or more component carriers in order to support wider transmission bandwidths has several implications for the RF characteristics. The impact for the base station and UE RF characteristics are also quite different. Release 10 has some restrictions on carrier aggregation in the RF specification, compared to what has been specified for physical layer and signaling. There is, from an RF point of view, a substantial difference between the two types of Carrier Aggregation (CA) defined for LTE :

Intra-Band Carrier Aggregation implies that two or more carriers within the same operating band are aggregated (see also the first two examples in Figure 17.27). RF requirements are restricted in release 10 to contiguous intra-band aggregation and a maximum of two carriers. Since aggregated carriers from an RF perspective have similar RF properties as a corresponding wider carrier being transmitted and received, there are many implications for the RF requirements. This is especially true for the UE. For the base station, it corresponds in practice to a multicarrier configuration (non-aggregated) already supported in earlier releases, which also means that the impact is less than for the UE.

Inter-Band Carrier Aggregation implies that carriers in different operating bands are aggregated (see also the last example in Figure 17.27). Many RF properties within a band can, to a large extent, remain the same as for a single carrier case. There is, however, impact for the UE, due to the possibility for intermodulation and cross-modulation within the UE device when multiple transmitter and receiver chains are operated simultaneously. For the base station it has very little impact, since in practice it corresponds to a base station supporting multiple bands, which is a configuration not really treated in RF specifications.

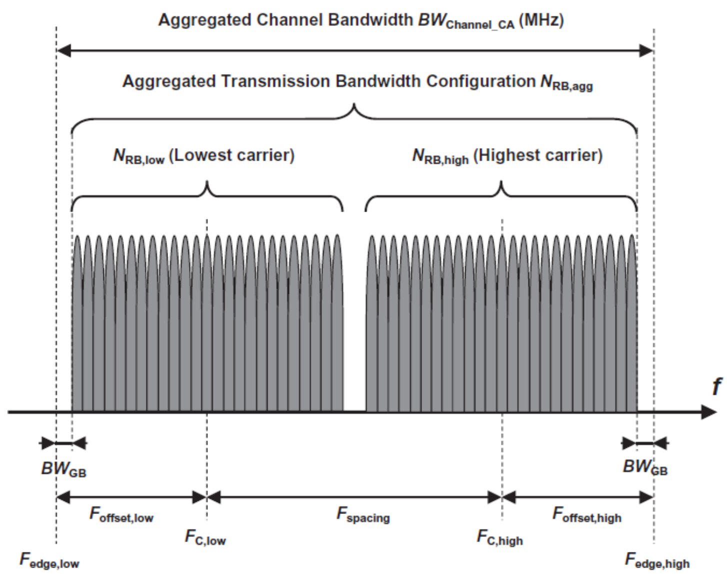


Figure 17.28: Definitions for Intra-Band Carrier Aggregation RF parameters, for an example with two aggregated carriers.

Intra-band carrier aggregation is limited to two component carriers and to one paired band (Band 1) and one unpaired (Band 40) band in release 10. Inter-band carrier aggregation is limited to the generic case of aggregating carriers between Bands 1 and 5. The next band pair for which a carrier aggregation capability is specified is a "European" scenario for Bands 3 and 7, which is planned for later inclusion in release 10. The band or set of bands over which carriers are aggregated is defined as a UE capability called E-UTRA CA Band. For the base station the band or set of bands defines what is called a Carrier Aggregation Configuration for the base station.

For intra-band carrier aggregation, the definitions of $BW_{channel}$ and NRB shown in Figure 17.27 still apply for each component carrier, while new definitions are needed for the Aggregated Channel Bandwidth ($BW_{Channel\ CA}$) and the Aggregated Transmission bandwidth Configuration (N_{RB},agg) shown in Figure PII 3.28. In connection with this, a new capability is defined for the UE called Carrier Aggregation Bandwidth Class. There are six classes, where each class corresponds to a range for N_{RB},agg and a maximum number of component carriers, as shown in Table 17.6. The classes corresponding to aggregation of more than two component carriers or consisting of more than 200 RBs are under study for later releases.

A fundamental parameter for intra-band CA is the channel spacing. A tighter channel spacing than the nominal spacing for any two single carriers could potentially lead to an increase in spectral efficiency, since there would be a smaller unused "gap" between carriers. On the other hand, there is also a requirement for the possibility to support legacy single-carrier terminals of earlier releases.

Table 17.6: UE Carrier Aggregation Bandwidth Classes.

Channel Aggregation Bandwidth Classes	Aggregated Transmission BW Configuration	Number of Component Carriers
A	≤100	1
B	≤100	2
C	101–200	2
D, E, F	Under study (201–500)	Under study

An additional complication is that the component carriers should be on the same 15 kHz subcarrier raster in order to allow reception of multiple adjacent component carriers using a single FFT instead of an FFT per subcarrier. This property, together with the fact that the frequency numbering scheme is on a 100 kHz raster, results in the spacing between two component carriers having to be a multiple of 300 kHz, which is the least common denominator of 15 and 100 kHz.

For the specification, RF requirements are based on a nominal channel spacing that is derived from the channel bandwidth of the two adjacent carriers BWChannel and BWChannel as follows:

$$F_{\text{Spacing,Nominal}} = \left\lfloor \frac{BW_{\text{Channel(1)}} + BW_{\text{Channel(2)}} - 0.1\left|BW_{\text{Channel(1)}} - BW_{\text{Channel(2)}}\right|}{2 \cdot 0.3} \right\rfloor 0.3.$$

In order to allow for a tighter packing of component carriers, the value of FSpacing can be adjusted to any multiple of 300 kHz that is smaller than the nominal spacing, as long as the carriers do not overlap. RF requirements for LTE are normally defined relative to the channel bandwidth edges. For intraband CA, this is generalized so that requirements are defined relative to the edges of the Aggregated Channel Bandwidth, identified in Figure 17.28 as $F_{\text{edge,low}}$ and $F_{\text{edge,high}}$. In this way many RF requirements can be reused, but with new reference points in the frequency domain. The aggregated channel bandwidth for both UE and base station is defined as: 309

$$BW_{\text{Channel\_CA}} = F_{\text{edge,high}} - F_{\text{edge,low}}$$

The location of the edges is defined relative to the carriers at the edges through a new parameter F_{offset} (see Figure 17.28) using the following relation to the carrier center positions FC of the lowest and highest carriers:

$$F_{\text{edge,low}} = F_{\text{C,low}} - F_{\text{offset,low}}$$

$$F_{\text{edge,high}} = F_{\text{C,high}} + F_{\text{offset,high}}$$

The value of Foffset for the edge carriers and the corresponding location of the edges are, however, not defined in the same way for UE and base station. For the base station, there are legacy scenarios where the base station receives and transmits adjacent independent carriers, supporting legacy terminals of earlier releases using single carriers. This scenario will also have to be supported for a configuration of aggregated carriers. In addition, for backward compatibility reasons, a fundamental parameter such as channel bandwidth and the corresponding reference points (the channel edge) for all RF requirements will have to remain the same. The implication is that the channel edges shown in Figure 17.27 for each component carrier will also remain as reference points when the carriers are aggregated. This results in the following base station definition of F_{offset}, for carrier aggregation, which is "inherited" from the single carrier scenario:

$$F_{\text{offset}} = \frac{BW_{\text{channel}}}{2} \quad \text{(for base station).}$$

Unlike the base station, the UE is not restricted by legacy operation, but rather from the nonlinear properties of the PA and the resulting unwanted emissions mask. At both edges of the aggregated channel bandwidth, a guard band BW_{GB} will be needed, in order for the emissions to reach a level where the out-of-band emissions limits in terms of an emission mask are applied. Whether a single wide carrier or multiple aggregated carriers of the same or different sizes are transmitted, the guard band needed will have to be the same at both edges, since the emission mask roll-off is the same. A problem with the backwards-compatible base station definition is that the resulting guard BW_{GB} is proportional to the channel BW and would therefore be different if carriers of different channel BW are aggregated.

For this reason, a different definition is used for the UE, based on a "symmetrical" guard band. For the edge carriers (low and high), F_{offset} is half of the transmission bandwidth configuration, plus a symmetrical guard band BW_{GB}:

$$F_{\text{offset}} = \frac{0.18\,\text{MHz} \cdot N_{\text{RB}}}{2} + BW_{\text{GB}} \quad \text{(for UE)},$$

where 0.18 MHz is the bandwidth of one resource block and BWGB is proportional to the channel BW of the largest component carrier. For the CA bandwidth classes defined in release 10 and where the edge carriers have the same channel bandwidth, F_{offset} will be the same for terminals and base stations and $BW_{Channel\_CA}$ will be the same.

It may look like an anomaly that the definitions may potentially lead to slightly different aggregated channel BW for the UE and the base station, but this is in fact not a problem. UE and base station requirements are defined separately and do not have to cover the same frequency ranges. The aggregated channel BW for both UE and base station do, however, have to be within an operator's license block in the operating band.

Once the frequency reference point is set, the actual RF requirements are to a large extent the same as for a single carrier configuration. Which requirements are affected is explained for each requirement in the discussion later in this chapter.

17.16 LTE PERFORMANCE

Computer simulation of mobile systems is a very powerful tool to assess the system performance. The "real-life" performance can, of course, be measured and evaluated in the field for an already deployed system and such values represent a valid example of performance for a certain system configuration. But there are several advantages with computer simulations:

- Evaluation can be made of system concepts that are not deployed or still under development.
- There is full control of the environment, including propagation parameters, traffic, system layout, etc., and full traceability of all parameters affecting the result.
- Well-controlled "experiments" comparing similar system concepts or parts of concepts can be one under repeatable conditions.

In spite of the advantages, simulation results obviously do not give a full picture of the performance of a system. It is impossible to model all aspects of the mobile environment and to properly model the behavior of all components in a system. Still, a very good picture of system performance can be obtained and it can often be used to find the potential limits for performance. Any simulated performance number should be viewed in the context that real radio network performance will depend on many parameters that are difficult to control or model, including:

- The mobile environment, including channel conditions, angular spreads, clutter type, terminal speeds, indoor/outdoor usage, and coverage holes.
- User-related behavior, such as voice activity, traffic distribution, and service distribution. l System tuning of service quality and network quality.
- Deployment aspects such as site types, antenna heights and types, and frequency planning.
- A number of additional parameters that are usually not modeled, such as signaling capacity and performance, and measurement quality.

There is no single universal measure of performance for a telecommunications system. Indeed, end-users (subscribers) and system operators define good performance quite differently. On the one hand, end-users want to experience the highest possible level of quality. On the other hand, operators want to derive maximum revenue, for example by squeezing as many users as possible into the system.

Performance-enhancing features can improve perceived quality of service (end-user viewpoint) or system performance (operator viewpoint). The good news, however, is that LTE and its evolution has the potential to do both. Compared to 3G systems, LTE yield better data rates and shorter delay. That is, LTE and its evolution can greatly improve both the service experience (end-user viewpoint) and the system capacity (operator viewpoint).

End-User Perspective of Performance

Users of circuit-switched services are assured of a fixed data rate. The quality of service in the context of speech or video telephony services is defined by perceived speech or video quality. Superior quality services have fewer bit errors in the received signal. By contrast, users who download a web page or movie clip via packet data describe quality of service in terms of the delay they experience from the time they start the download until the web page or movie clip is displayed. Best-effort services do not guarantee a fixed data rate. Instead, users are allocated whatever data rate is available under present conditions. This is a general property of packet-switched networks – that is, network resources are not reserved for each user. Given that the delay increases with the size of the object to be downloaded, absolute delay is not a fair measure of quality of service.

Performance of a packet data service in cellular systems can be characterized by several different measures depending on the perspective taken (see Figure 17.29). A single user in a radio network experiencing good radio conditions may enjoy the peak data rate of the radio interface. A user will, however, normally share radio resources with other users. If radio conditions are less than optimal or there is interference from other users, the radio-interface data rate will be less than the peak data rate.

In addition, some data packets might be lost, in which case the missing data must be retransmitted, further reducing the effective data rate as seen from higher protocol layers. Furthermore, the effective data rate diminishes even further as the distance from the cell increases (due to poorer radio conditions at cell edges). The data rate experienced above the MAC layer, after sharing the channel with other users, is denoted user throughput.

The Transmission Control Protocol (TCP) – the protocol at the transport layer – is commonly used together with IP traffic. However, due to the so-called slow-start algorithm, which is sensitive to latency in the network, it is especially prone to cause delay for small files. The slow-start algorithm is meant to ensure that the packet transmission rate from the source does not exceed the capability of network nodes and interfaces.

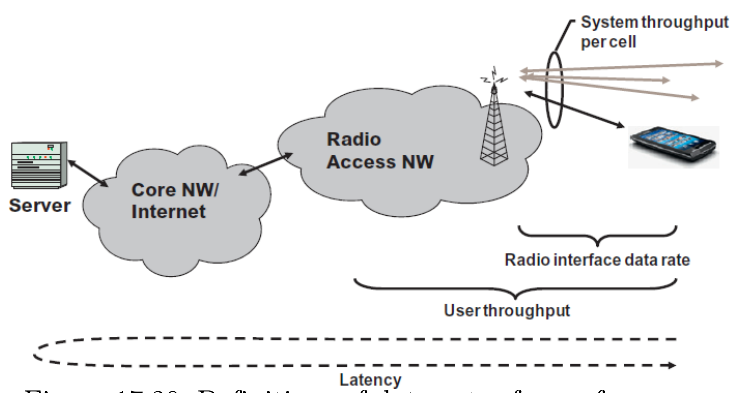


Figure 17.29: Definitions of data rates for performance.

Network latency, which in principle is a measure of the time it takes for a packet to travel from a client to a server and back again, has a direct impact on performance with TCP. Therefore, an important design objective for LTE has been to reduce network latency. One other quality-related criterion (end-user viewpoint) relates to the setup time for initiating, for example, a web-browsing session.

Operator Perspective

Radio resources need to be shared when multiple users are in the network. As a result, all data must be queued before it can be transmitted, which restricts the effective data rate to each user. Notwithstanding this fact, by scheduling radio resources, operators can improve system throughput or the total number of bits per second transmitted over the radio interface. A common measure of system performance is "spectral efficiency", which is the system throughput per MHz of spectrum in each cell of the system.

LTE employs intelligent scheduling methods to optimize performance, from both end-user and operator viewpoints. An important performance measure for operators is the number of active users who can be connected simultaneously. Given that system resources are limited, there will thus be a trade-off between number of active users and perceived quality of service in terms of user throughput.

Performance In Terms of Peak Data Rates and Latency

LTE has been developed in a process where design targets for performance parameters play an important role. One target is for the peak data rate over the radio interface. The original design targets for the first release of LTE are documented in 3GPP TR 25.913. The target capability when operating in a 20 MHz spectrum allocation is a peak data rate of 100 Mbit/s in the downlink and 50 Mbit/s in the uplink. The numbers assume two receive antennas in the terminal for the downlink capability and one transmit antenna for the uplink capability.

These target numbers are exceeded by a good margin by the peak data rate capability of the specified LTE standard. LTE release 8 supports peak data rates of 300 Mbit/s in the downlink and 75 Mbit/s in the uplink by using spatial multiplexing of four layers (4 X 4 MIMO) in the downlink and 64QAM in both downlink and uplink. With the assumptions in the design targets – that is, spatial multiplexing of two layers – the downlink peak data rate is 150 Mbit/s, which is still significantly higher than the target. The design targets for LTE release 10 ("LTE-Advanced") are documented in 3GPP TR 36.913, based on the targets set by ITU-R. There is no absolute peak data rate target expressed for LTE release 10; it is instead expressed relative to the channel bandwidth as a peak spectral efficiency, with targets of 15 bit/s/Hz for downlink and 6.75 bit/s/MHz for uplink. LTE release 10 exceeds those numbers by a good margin, as shown in Table 3.7.

The assumptions for deriving the peak spectral efficiency numbers is a deployment with 20 MHz channel bandwidth, 8 X 8 MIMO in the downlink, and 4X4 MIMO in the uplink. The ITU-R requirement for downlink peak spectral efficiency is in fact fulfilled already by LTE release 8, assuming 4 X4 MIMO in the downlink. The ITU-R target for latency in is set with a different definition than in Figure 17.29. Instead of a two-way latency, there is a 10 ms target for the one-way latency in both downlink and uplink.

The one-way latency is defined as the one-way transit time between a packet being available at the IP layer in the user base station and the availability of this packet at the IP layer in the user terminal, and vice versa. This is achieved with a good margin for LTE, where the latency achieved is 4 ms for LTE FDD and 4.9 ms for LTE TDD.

Table 17.7: LTE Peak Spectral Efficiency.

Peak Spectral Efficiency	ITU Requirement (bit/s/Hz)	LTE Fulfillment			
		Release 8		Release 10	
		FDD	TDD	FDD	TDD
Downlink	15	15.3	15.0	30.6	30.0
Uplink	6.75	4.2	4.0	16.8	16.0

17.17 PERFORMANCE EVALUATION OF LTE-ADVANCED

An essential part of the submission of LTE release 10 to ITU-R as a candidate for IMT-Advanced was the performance evaluation of the spectral efficiency. The technical requirements are set by ITU-R Report M.2134 and the detailed evaluation methodology is described in ITU-R Report M.2135. For the work on LTE release 10, 3GPP performed a large simulation campaign with input from several 3GPP members. The resulting performance numbers formed an essential part of the submission of LTE-Advanced as a candidate for IMT-Advanced and is reported in detail in 3GPP TR 36.912. In addition to the ITU-R evaluation criteria, an additional test environment with higher performance targets than the ones set by ITU-R was defined in 3GPP. Performance numbers for this additional environment are also reported in.

Table 17.2: Test Environment and Deployment Parameters for the Evaluation.

	Test Environment			
	Indoor	Microcellular	Base Coverage, Urban	High Speed
Deployment scenario	Indoor hotspot	Urban micro	Urban macro	Rural macro
Channel model	InH	UMi	UMa	RMa
Carrier frequency (GHz)	3.4	2.5	2.0	0.8
Inter-site distance (m)	60	200	500	1732
Terminal speed (km/h)	3	3	30	120
User distribution	100% indoor	50% indoor, 50% outdoor	100% outdoor	100% outdoor
BS antenna height (m)	6	10	25	35
BS antenna gain (dBi)	0	17	17	17
BS output power (dBm/20MHz)	21	44	49	49
UE output power (dBm)	21	24	24	24

Models and Assumptions

This section presents the models and assumptions used for the evaluation in. A summary of the test environments is given in Table 17.8 and the assumptions for the LTE-Advanced system characteristics are given in Table 17.9. The evaluation is performed in four deployment scenarios, each corresponding to a different test environment defined by ITU-R:

Indoor Hotspot. Deployment scenario for the indoor environment, having isolated office cells or hotspots for stationary or pedestrian users, with high user density and high user throughput.

Urban Micro. Deployment scenario for the microcellular environment, having small cells with outdoor and outdoor-to-indoor coverage for pedestrian and slow vehicular users, provided by below rooftop outdoor base stations. It has high traffic per unit area and will be interference limited.

Urban Macro. Deployment scenario for the base coverage urban environment, having large cells and continuous coverage for pedestrians up to fast vehicular users, provided by above rooftop outdoor base stations. It will be interference limited and has mostly non-line-of-sight propagation conditions.

Rural Macro. Deployment scenario for the high speed environment, having large cells and continuous coverage for high-speed vehicular and train users. It will be noise and/or interference limited.

Each deployment scenario described in corresponds to a set of deployment parameters, including detailed propagation models for line-of-sight, non-line-of-sight, and/or outdoor-to-indoor propagation. The channel model is based on the so-called WINNER II channel model, which is a Spatial Channel Model (SCM) applicable for MIMO simulations. Further details of the test environments and deployment parameters are given in Table PII 3.8.

The simulation methodology is "time dynamic", where a system of base stations and terminals is simulated over a limited time frame (20–100 seconds). This is repeated to create a number of samples to reach sufficient statistical confidence. For each simulation, terminals are randomly positioned over a model of a radio network and the radio channel between each base station and terminal antenna pair is simulated according to the propagation and fading models. A full buffer traffic model with an average of 10 users per cell is assumed. This results in the system operating at the maximum 100% system load. Based on the channel realizations and the active interferers, a signal-to-interference-and-noise ratio (SINR) is calculated for each terminal (or base-station) receive antenna. The SINR values are then mapped to block error probability for the modulation and coding scheme employed for each user. MIMO and the modulation-and-coding scheme according to the LTE standard are selected based on delayed feedback. Retransmissions are explicitly modeled and the user throughput for each active user i will be the number of correctly received bits χ_i (above the MAC layer) divided by the simulated time T. The distribution of user throughput between users is used as a basis for assessing end-user quality. The served traffic per cell is calculated from the total number of received bits χ_i for all users, averaged over all cells and divided by the simulated time T. Statistics are collected from each simulation run and then new terminals are randomly positioned for the next sample.

Table17.8: LTE-Advanced System Characteristics for the evaluation.

General Characteristics	
Duplex method assumptions	FDD
	TDD: Configuration 1, DwPTS/GP/UpPTS length set to 12/1/1 OFDM symbols (see Chapter 9)
Spectrum allocation	10 MHz DL + 10 MHz UL for FDD,
	10 MHz for TDD
	Double bandwidth for InH case
Antenna configuration at base station (BS)	Vertically co-polarized
	Antennas with 4 (InH) or 0.5 (UMi, UMa, RMa) wavelengths separation
Antenna configuration at terminal (UE)	Vertically co-polarized
	0.5 wavelengths separation
Network synchronization	Synchronized, not explicitly utilized other than for avoiding UE–UE and BS–BS interference for TDD
Detailed Radio-Interface Characteristics and Models	
Scheduler	DL: Proportional fair in time and frequency
	UL: Quality-based frequency-domain multiplexing
Downlink transmission scheme	InH: Transmission mode 4; closed-loop codebook-based precoded adaptive rank spatial multiplexing
	UMi, UMa, RMa: Transmission mode 5; coordinated beam-forming (within site) with MU-MIMO
Uplink transmission scheme	1 Tx, 4Rx antennas, no MU-MIMO
Receiver type	Minimum Mean Squared Error (MMSE) in DL and UL
Uplink power control	Open loop with fractional path-loss compensation, parameters chosen according to the deployment scenario. Effective noise rise below 10 dB
Hybrid-ARQ scheme	Incremental redundancy, synchronous, adaptive
Link adaptation	Non-ideal, based on delayed feedback
Channel estimation	Non-ideal channel estimation
	Non-ideal channel-state reports in downlink; CQI error per resource block is N(0,1) dB, error-free feedback of the reports, 6 ms reporting delay, 5 ms reporting periodicity
	Uplink quality estimated from PUSCH, 6 ms delay, 20 ms sounding period
Feedback channel errors	Error-free, but quantized and delayed
Control channel overhead	DL: 3 OFDM symbols per subframe
	UL: 4 resource blocks
	Overhead for common control channels (synchronization, broadcast and random access; ~1% for 10 MHz) has not been deducted

Evaluation Criteria

ITU-R defines two requirements related to the efficiency of the radio interface for evaluating the performance of the IMT-Advanced candidate radio-interface technologies (RITs). The first is cell spectral efficiency, defining the operator perspective, and the second is cell-edge spectral efficiency, defining the end-user perspective.

The cell spectral efficiency is the aggregated throughput over all users, averaged over all cells and divided by the channel bandwidth. The measure relates to the system throughput in Figure 17.29 and is a measure of the maximum total “capacity” available in the system to be shared between users; it is measured in bits/s/Hz/cell.

The cell spectral efficiency η is defined as:

$$\eta = \frac{\sum_{i=1}^{N} \chi_i}{T \cdot \omega \cdot M},$$

Table 17.9: ITU-R Requirements for IMT-Advanced Spectral Efficiency

Test Environment and Corresponding Deployment Scenario	Cell Spectral Efficiency (bit/s/Hz/cell)		Cell-Edge User Spectral Efficiency (bit/s/Hz)	
	Downlink	Uplink	Downlink	Uplink
Indoor (InH)	3	2.25	0.1	0.07
Microcellular (UMi)	2.6	1.8	0.075	0.05
Base coverage, urban (UMa)	2.2	1.4	0.06	0.03
High speed (RMa)	1.1	0.7	0.4	0.015

where χ_i denotes the number of correctly received bits for user i in a system with N users and M cells, ω is the channel bandwidth, and T is the time over which the data bits are received. The cell-edge user spectral efficiency is based on the distribution between users of the normalized user throughput (see also Figure 17..29), which is defined as the average user throughput over a certain period of time divided by the channel bandwidth, and is measured in bit/s/Hz. The cell-edge user spectral efficiency is defined as the 5% point of the cumulative distribution function (CDF) of the normalized user throughput. It is thus a measure of the end-user perceived "quality of service" for the 5% of the users with the lowest user throughput.
The normalized user throughput for user i is defined as:

$$\gamma_i = \frac{\chi_i}{T_i \cdot \omega},$$

where Ti is the active session time for user i. The requirements defined by ITU-R for cell spectral efficiency and cell-edge user spectral efficiency are listed in Table PII 3.9. The requirement levels are shown as dashed lines together with the simulated performance values in Figure 17.30 and 17.33. 315

Performance Numbers for FDD

Simulations for FDD and TDD and for all test environments were submitted for the ITU-R evaluation and are also presented here. The numbers here, however, are not identical to the numbers put forward to ITU-R, but will be seen as one sample of results from the evaluation.

For FDD, the cell spectral efficiency and cell-edge spectral efficiency are shown in Figure PII 3.30. All results exceed the ITU-R performance requirements in Table PII 3.9, in some cases by a very good margin. For the indoor scenario and the downlink, this is achieved using an uncorrelated antenna configuration together with single-user MIMO. For the other scenarios, a correlated antenna setup and intra-site coordinated beam-forming (CBF) with MU-MIMO are used in the downlink. A quite basic uplink configuration not utilizing MIMO is used for all environments. It should be noted that the cell-edge spectral efficiency is highly dependent on the system load and that the ITU-R simulations are performed with maximum load and a full buffer model.

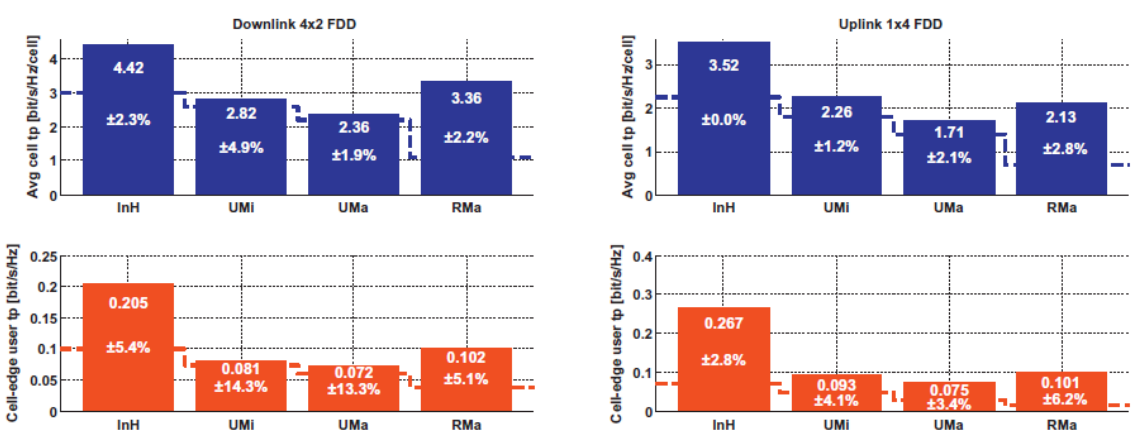


Figure 17.30: FDD cell spectral efficiency and cell-edge user spectral efficiency, compared with ITU-R requirements (downlink and uplink).

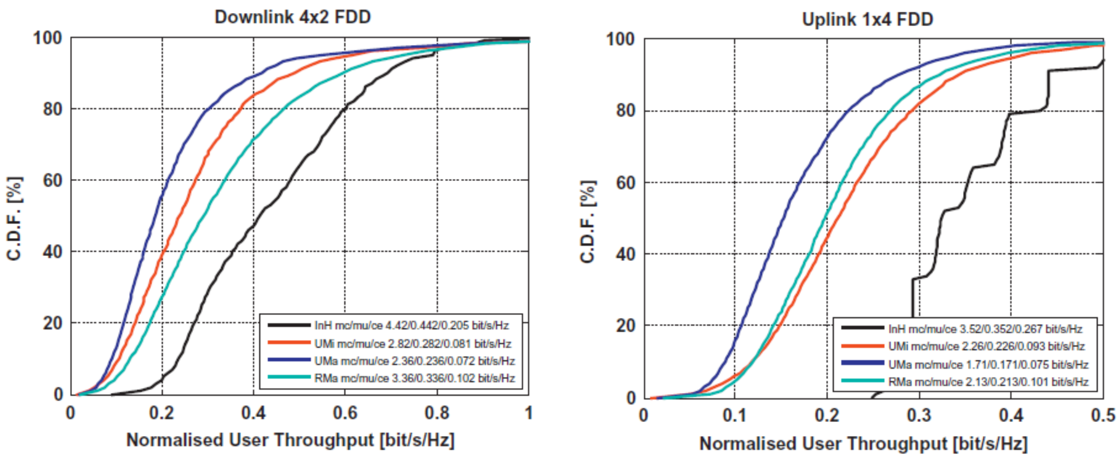


Figure 17.31: FDD normalized user throughput distributions (downlink and uplink).

The normalized user throughput distributions are given in Figure 17.30 and Signal-to-Noise-and-Interference Ratio (SINR) distributions are shown in 17.31. The figures demonstrate the impact of the very advantageous SINR distributions achieved in the indoor hotspot scenario and to some extent also for rural macro, resulting in higher spectral efficiency numbers in those environments.

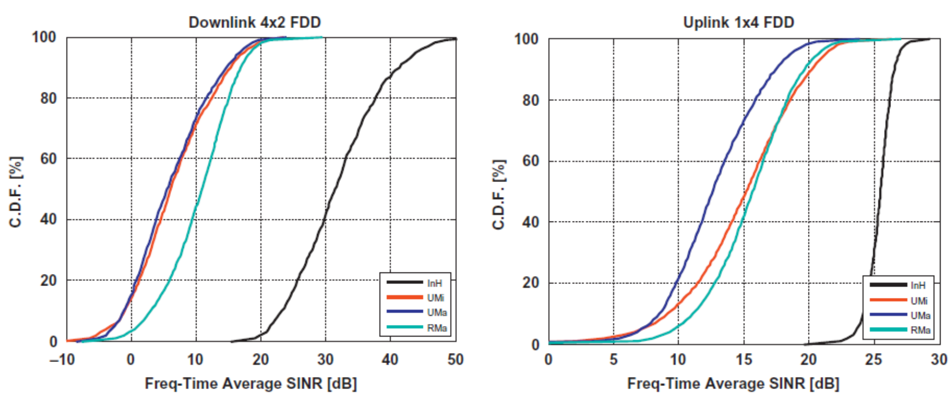


Figure 17.32: FDD SINR distributions (downlink and uplink).

Figure 17.33: TDD cell spectral efficiency and cell-edge user spectral efficiency, compared with ITU-R requirements (downlink and uplink). Spectral efficiency numbers compared to FDD, but all results exceed the ITU-R performance requirements. The difference compared to FDD is due to a higher overhead from the guard period between uplink and downlink. The average delay between making measurements in the terminal and receiving the measurement result at the base station is also increased compared to FDD, due to the TDD time domain structure. This has some impact on the performance of scheduling and link adaptation. The channel reciprocity is not utilized in the simulated TDD scheme.

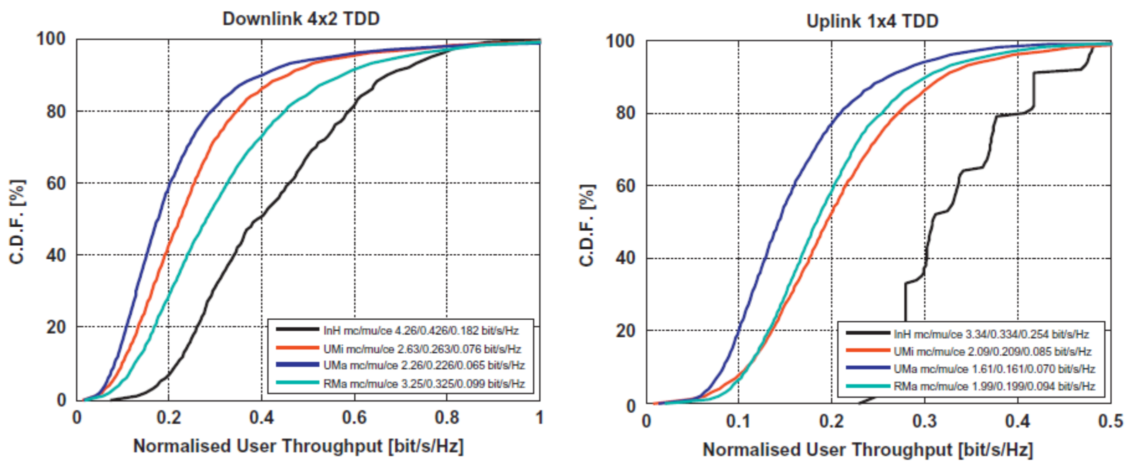
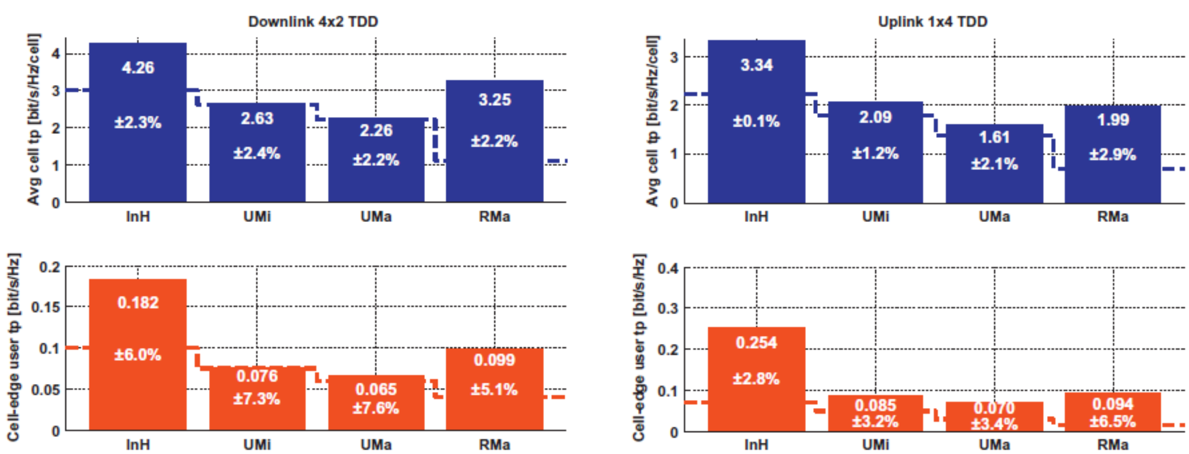


Figure 17.34: TDD normalized user throughput distributions (downlink and uplink).

The simulation results presented in this chapter demonstrate the high potential of LTE/LTE-Advanced in terms of both overall spectral efficiency to the benefit of operators and high cell-edge performance to the benefit of the end-user. The performance is assessed in four different test environments. The results for FDD and TDD in downlink and uplink all exceed the requirements set up by ITU-R for the evaluation of IMT-Advanced candidates. In addition, the peak radio-interface data rates and the latency achieved by LTE meet the ITU-R requirements.



EXAMPLE

Problem 01
Problem 39
A mobile wireless system is specified by the following.
Channel bandwidth =10 MHz; Received Signal-to-noise ratio = 20 dB. Note that for part a and b, we consider SISO antenna configuration.

a) i) Find the capacity of the channel.

ii) Find the channel bandwidth utilization as well as minimum required received $\frac{E_b}{N_0}$.

b) If the received signal-to-noise ratio is now reduced by half (in dB), find the required bandwidth to provide the same capacity (as you have found in part a (i)).
c) If 4 X 2 MIMO configuration is used in the system, find the channel capacity.
d) Compare the results that you have found in part a (i) with part c and draw a conclusion.

Problem 40
An IEEE 802.11a system is specified as follows.
300-MHz system bandwidth,
5 GHz band,
20-MHz channel bandwidth,
Length of information symbol = 64,
Length of cyclic prefix= 16,
48 sub channels for data transmission,
Error correction code rates: r = 1/2, and
Modulation formats: QPSK and 16-QAM.

Compute the followings.
a) Transmitted pulse rate
b) Overhead OFDM symbol period because of cyclic prefix and the corresponding bandwidth utilization.
c) Minimum and maximum data rates that can be achieved.

Problem 41
Consider a scenario where the same number of multiple mobile users uses a TDMA based network and a CDMA based network and a FDMA based network to transfer information to their respective receivers. Assume that the channel bandwidth of these networks is the same and is denoted by **W**. For multi-user access scenario, if simultaneously a maximum of 5 users access the networks, and all users are equally likely to channel bandwidth allocation by the respective network schedulers, what is the maximum and minimum time-bandwidth product that a TDMA user, a CDMA user and a FDMA user can be assigned to in a TTI (1 ms). Note that, we assume at least one user is active at any time instant to any network. Hints: you take help of the lecture note on fundamentals of CDMA mobile communications and apply the multiple access scheme principles there in.

Problem 42
The total spectrum =20 MHz
Consider guard band = 2 MHz
Measured CQI is post detection SINR
Each RB comprises of 12 subcarriers
Bandwidth of each subcarrier =15 kHz
measurement is done at each RB level
Transmission time interval (TTI)= 1 ms
Gaussian distributed CQI measurement error with zero mean and 1 dB standard deviation
The number of encoded bits for each quantized value = 7
Threshold window = 5 dB
Percentage number of sub-bands in the threshold window = 15 % of the total RBs.

Compute the followings.
a) The number of RBs in the system bandwidth
b) The number of measurement intervals.
c) Uplink signal overhead for wideband CQI scheme, full CQI scheme, and threshold based CQI scheme.
d) The number of claimed RBs when we consider threshold based CQI scheme.

Problem 43
Describe the typical approaches to increase the data rates in mobile communication systems. Note that your answer should not exceed more than one page written on both sides.

Problem 44

Explain in brief why Proportional Fair scheduling algorithm is more suitable in contrast to other scheduling algorithms.

Problem 45
Explain in brief an approach to address frequency selective channel property. Mention drawback if there is any, in using that approach.

Problem 46
Explain the terms power-limited region and bandwidth-limited region. Mention approaches that are typically used to increase data rates in these regions.

Problem 47
(a) Why does fast power control is necessary in WCDMA system?

(b) It is necessary to perform control signalling between MS and BS before the user generated information is sent via the network. Justify the statement and mention the control signals and their functions associated with this signalling procedure.
(c) It is not possible to track a particular user in MBMS area. Justify the statement.
(d) Explain why in HSDPA most physical layer operations for radio resource management are performed at the NodeB (BS).
(e) Mention advantages and disadvantages of amplify-and-forward relays over decode-and-forward relays.
(f) Justify why the buffer status should be taken into account in scheduling decision process.
(g) Explain why channel dependent scheduling is not suitable for high speed environments.
(h) Mention approaches to increase system capacity and to support high data rate in mobile communications.
(i) Write down the advantages and disadvantages of closed loop power control mechanism.
(j) High spectral efficiency is crucial to address high data rate. Justify the statement. Mention a network element that is mostly responsible to provide high spectral efficiency.

Problem 48
Consider an LTE-Advanced system. An LTE user can access only one component carrier at a time, whereas an LTE-Advanced user can access the whole bandwidth. Consider a system level scenario given by the followings.

System bandwidth is 80 MHz,
Component carrier (CC) bandwidth is 20 MHz,
Guard band for each CC is (10% of CC) bandwidth,
Non-contiguous carrier aggregation,
Single sector cell-site,
Number of users per cell is 20,
Simulation runtime is 1000 ms,
TTI is 1 ms,
One resource block (RB) size is 180 kHz,
Minimum frequency resource that can be allocated to any user at any TTI is one RB, and
Full buffer traffic.

Find the followings.
(a) Total number of RBs in the system.
(b) Maximum and minimum number of RBs that an LTE user can be allocated at any TTI.
(c) Maximum and minimum number of RBs that an LTE-Advanced user can be allocated at any TTI.
(e) Cell average spectral efficiency. Consider an average throughput of 80 bps per TTI per RB.
(f) Cell-edge user spectral efficiency. Consider a linear change in CDF (%) vs. normalized user throughput (bps/Hz) in the range of 0 to 20% of CDF for the corresponding 0 to 0.2 bps/Hz normalized user throughput.

Problem 49
Consider a contiguous Intra-Band carrier aggregation in an LTE-Advanced system with two component carriers (CCs), each has a bandwidth of 20 MHz (including guard bands). Find the following RF parameters. Assume 100 kHz channel raster between the edges of CCs. Note that the guard bands for both CCs are the same and is 1 MHz (approximate) on each edge of any CC.

(a) Nominal channel spacing between CCs.
(b) Offset frequency for the edge carriers, both in the case of base station and user terminal.
(c) Transmission bandwidth.
(d) Aggregated channel bandwidth.

Problem 50
Consider a cellular mobile system that operates at 890 MHz to 895 MHz. Assume, WCDMA, GSM, and CDMA (IS-95) technologies is considered by an operator for the feasibility study. Find the followings for each technology

(a) System bandwidth.
(b) Number of channels.
(c) Frequency reuse factor. Assume that all cells have equal amount of frequency allocated and is (5/12) MHz for reuse factor greater than unity.
(d) Number of TTIs required by the UE if at any TTI it is observed that uplink scheduler sends a control command to an arbitrary user equipment (UE) to reduce the transmit power by an amount of 1 dB. Assume that the UE is scheduled at each TTI for a total of 30 TTIs after the control command was sent by the BS to the UE. In addition, a power control command of (+/-) 0.5 dB is considered.

Problem 50
Consider a contiguous Intra-Band carrier aggregation in an LTE-Advanced system with two component carriers (CCs), each has a bandwidth of 20 MHz (including guard bands). Find the following RF parameters. Assume 100 kHz channel raster between the edges of CCs. Note that the guard bands for both CCs are the same and is 1 MHz (approximate) on each edge of any CC.

(a) Nominal channel spacing between CCs.
(c) Transmission bandwidth.
(d) Aggregated channel bandwidth.

Problem 51
Consider a cellular mobile system that operates at 890 MHz to 895 MHz. Assume, WCDMA, GSM, and CDMA (IS-95) technologies is considered by an operator for the feasibility study. Find the followings for each technology
(a) System bandwidth.
(b) Number of channels.
(c) Frequency reuse factor. Assume that all cells have equal amount of frequency allocated and is (5/7) MHz for reuse factor greater than unity.
(d) Number of TTIs required by the UE if at any TTI it is observed that uplink scheduler sends a control command to an arbitrary user equipment (UE) to reduce the transmit power by an amount of 2 dB. Assume that the UE is scheduled at each TTI for a total of 20 TTIs after the control command was sent by the BS to the UE. In addition, a power control command of (+/-) 0.5 is considered.

RFERENCES

[1] **A. Goldsmith**, 2005. "Wireless Communications", Cambridge University Press.

[2] **B. Sklar** (1997, September). "Rayleigh Fading Channels in Mobile Digital Communication Systems Part I: Characterization", IEEE Communications Magazine, pp. 136-146, September 1997..

[3] **S. Faruque**, (1996).Cellular Mobilesystems Engineering. Norwood, MA 02062. Artech House, INC.

[4] **D. J. Gibson**, (1996).The Communications Handbook.USA. CRC Press, Inc.

[5] **M. Hata**, (1998). Propagation Loss Prediction Models for Land Mobile Communications. IEEE, 15 - 18.

[6] **M. Hata**, (1980). Empirical Formula for Propagation Loss in Land Mobile Radio Services. IEEE Transactions On Vehicular Technology, 29 (3), 317-325.

[7] **K. M. Ahmed**, 2009. "AT 77.07: Cellular Mobile Systems". Lecture notes. Telecommunications Engineering, Asian Institute of Technology, Bangkok, Thailand.

[8] **D. Parsons**, (1994).The Radio Propagation Channel. London. PENTECH PRESS.

[9] **S. Ranvier**, (2004). Physical layer methods in wireless communication systems. (Lecture notes, Course no.S-72.333, SMARAD Centre of Excellence). Helsinki: Helsinki University of Technology.

[10] **K. T, Sarkar, et al.** (2003, June). A Survey of Various Propagation Models for Mobile Communication. IEEE Antennas and Propagation Magazine, 45(3), 51-82.

[11] **R. S. Saunders** (1999). Antenna And Propagation For Wireless Communication Systems. Chichester, England. John Willy & Sons Ltd.

[12] **T. S. Rappaport**, 2002. "Wireless Communications; Principles and Practice". Prentice-Hall Inc., NJ.

[13] **J. Walfisch & L. H. Bertoni**, (1988). A Theoretical Model of UHF Propagation in Urban Environments. IEEE Transactions On Antennas And Propagation, 36 (12), 7788 – 7796.

[14] **B. Ghribi & L. Logrippo** (2000). "Understanding GPRS: the GSM packet radio service", ELSEVIER, Computer Networks, vol. 34, pp. 763-779.

[15] H. **Granbohm & J. Wiklund** (1999). "GPRS-General packet radio service", Ericsson Review, vol. 2, pp. 82-88.

[16] **D. Yamaguchi, et al.**, "EDGE Introduction to High-Speed Data in GSM/GPRS Networks" whitepaper, microsite mobile.

[17] **Håkan Axelsson, et al.**, (2006). "GSM/EDGE continued evolution", Ericsson Review Vol. 1, pp. 20-29.

[18] **Ericsson** (Sep 2009). "The evolution of EDGE". White Paper.

[19] **Rysavy Resea**rch, (September 2007). "EDGE, HSPA, LTE: The Mobile Broadband Advantage". 3G Americas.

[20] **H. Holma and A. To**skala, "WCDMA for UMTS: Radio Access for Third Generation Mobile Communications", John Wiley & sons, Ltd., third edition, 2004.

[21] **C. Lang**ton (2002). "Intuitive Guide to Principles of Communications: CDMA Tutorial", retrieved from www.complextoreal.com.

[22] **K. Kettunen** (1997). "Code Selection for CDMA Systems", Signal Processing in Communications, FALL -97, VTT, Information Technology, FIN-02044 VTT.

[23] **E. Dahlman et al.** (2011), "4G LTE/LTE-Advanced for Mobile Broadband", Academic press, UK.

[24] **Chae et al.** (2010, May), "Adaptive MIMO Transmission Techniques for Broadband Wireless Communication Systems", IEEE Communications Magazine, pp. 112-118.

[25] **Stüber et al.** (2004, February), "Broadband MIMO-OFDM Wireless Communications", Proceedings of the IEEE, Vol. 92, No. 2, pp. 271-293.

[26] **Mietzner et al.** (2009), ''Multiple-antenna Techniques For Wireless Communications – A Comprehensive Literature Survey'', IEEE Communications Surveys & Tutorial, vol. 11, No. 2, second quarter, pp. 87-105.

[27] **K. M. Ahmed** (2010), ''Cellular Mobile Systems'', Lecture notes on AT 77.07, Asian Institute of Technology, Bangkok, Thailand.

[28] **P. Saengudomlert** (April 2011), "Fundamentals of Orthogonal Frequency Division Multiplexing (OFDM)". Telecommunications program, Asian Institute of Technology (AIT), Bangkok, Thailand.

[29] **R.K. Saha**, "Resource reuse and scheduling in femtocell deployed two-tier LTE-Advanced Systems", Ait master thesis-ict, asian institute of technology, bangkok, thailand, may 2011.

[30] **Wang, Y., et al.** (2010). Carrier Load Balancing and Packet Scheduling for Multi-Carrier Systems. IEEE Transactions on Wireless Communications, 9 (5), 780-1789.

[31] **Chen, L., et al**. (2009). Analysis and Simulation for Spectrum Aggregation in LTE-Advanced System. IEEE. Retrieved May 25, 2010, from http://ieeexplore.ieee.org/.

[32] **Pokhariyal, A., et al.** (2007). Frequency Domain Packet Scheduling Under Fractional Load for the UTRAN LTE Downlink. IEEE Online. Retrieved June 30, 2010, from http://ieeexplore.ieee.org/.

[33] **Pokhariyal, A., et al**. (2007). HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness for the UTRAN Long Term Evolution. IEEE Online. Retrieved June 30, 2010, from http://ieeexplore.ieee.org/.

[34] **S. Parkvall, E. Englund, M. Lundevall, and J. Torsner** (2006, Feb), "Evolving 3G Mobile Systems: Broadband and Broadcast Services in WCDMA", IEEE Communications Magazine, pp. 68-74.

[35] **E. Dahlman, S. Parkvall, J. Sköld, and P. Beming**, "3G Evolution: HSPA and LTE for Mobile Broadband", Second edition, Academic Press , Linacre House, Jordan Hill, Oxford, OX2 8DP, 2008.

[36] **M. Schwartz (2005), "**Mobile Wireless Communications", Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK.

[37]