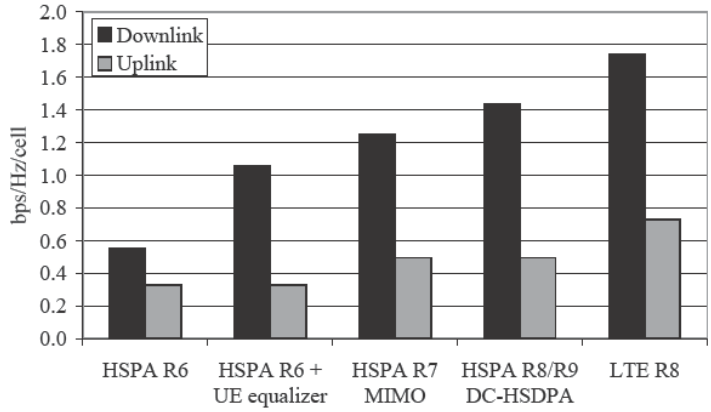


Figure 9.25 Spectral efficiency of HSPA and LTE

applications are, for example, voice, real time gaming and other interactive applications. The latency can be measured by the time it takes for a small IP packet to travel from the terminal through the network to the internet server, and back. That measure is called round trip time and is illustrated in Figure 9.26.

The end-to-end delay budget is calculated in Table 9.21 and illustrated in Figure 9.27. The 1-ms frame size allows a very low transmission time. On average, the packet needs to wait for 0.5 ms for the start of the next frame. The retransmissions take 8 ms at best and the assumed retransmission probability is 10%. The average delay for sending the scheduling request is

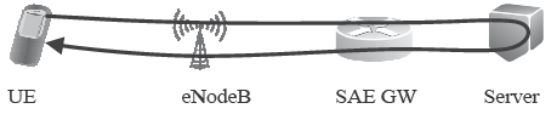


Figure 9.26 Round trip time measurement

Table 9.21 Latency components

Delay component	Delay value
Transmission time uplink + downlink	2 ms
Buffering time (0.5 × transmission time)	2 × 0.5 × 1 ms = 1 ms
Retransmissions 10%	2 × 0.1 × 8 ms = 1.6 ms
Uplink scheduling request	0.5 × 5 ms = 2.5 ms
Uplink scheduling grant	4 ms
UE delay estimated	4 ms
eNodeB delay estimated	4 ms
Core network	1 ms
Total delay with pre-allocated resources	13.6 ms
Total delay with scheduling	20.1 ms

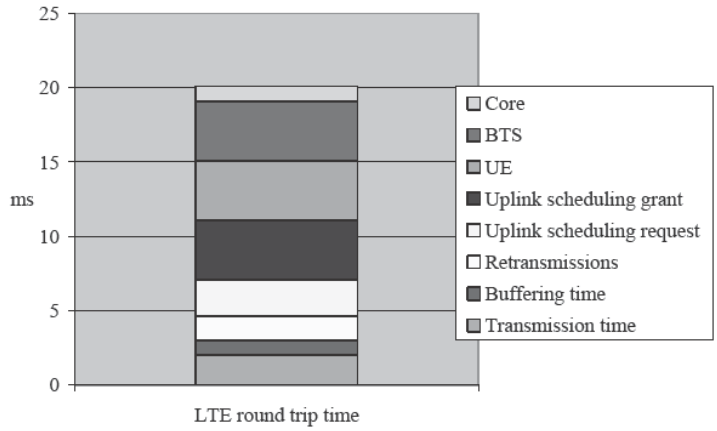


Figure 9.27 End-to-end round trip time including scheduling latency

2.5 ms and the scheduling grant 4 ms. We further assume a UE processing delay of 4 ms, an eNodeB processing delay of 4 ms and a core network delay of 1 ms.

The average round trip including retransmission can be clearly below 15 ms if there are pre-allocated resources. If the scheduling delay is included, the delay round trip time will be approximately 20 ms. These round trip time values are low enough even for applications with very tough delay requirements. The practical round trip time in the field may be higher if the transport delay is longer, or if the server is far away from the core network. Often the end-to-end round trip time can be dominated by non-radio delays, e.g. by the distance and by the other elements in the internet. The propagation time of 5000 km is more than 20 ms.

9.8 LTE Refarming to GSM Spectrum

LTE will be deployed in the existing GSM spectrum like 900 MHz or 1800 MHz. The flexible LTE bandwidth makes refarming easier than with WCDMA because LTE can start with 1.4 MHz or 3.0 MHz bandwidths and then grow later when the GSM traffic has decreased. The required separation of the LTE carrier to the closest GSM carrier is shown in Table 9.22. The required total spectrum for LTE can be calculated based on the carrier spacing. The coordinated

Table 9.22 Spectrum requirements for LTE refarming

	LTE–GSM carrier spacing		LTE total spectrum requirement	
	Coordinated	Uncoordinated	Coordinated	Uncoordinated
5 MHz LTE (25 RBs)	2.5 MHz	2.7 MHz	4.8 MHz	5.2 MHz
3 MHz LTE (125 RBs)	1.6 MHz	1.7 MHz	3.0 MHz	3.2 MHz
1.4 MHz LTE (6 RBs)	0.8 MHz	0.9 MHz	1.4 MHz	1.6 MHz

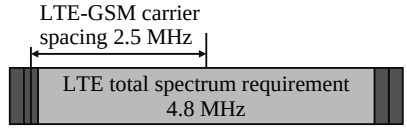


Figure 9.28 LTE 5-MHz refarming example

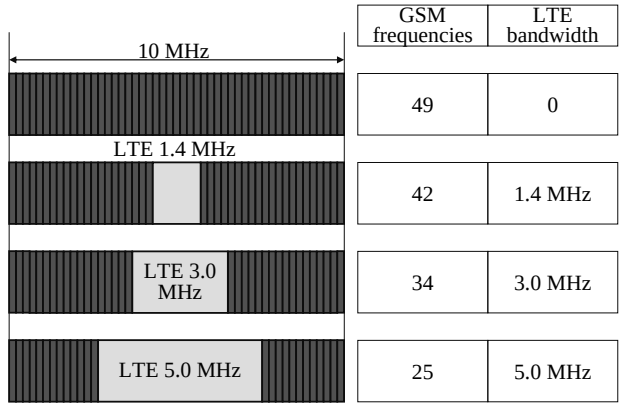


Figure 9.29 LTE refarming to GSM spectrum

case assumes that LTE and GSM use the same sites while the uncoordinated case assumes that different sites are used for LTE and GSM. The uncoordinated case causes larger power differences between the systems and leads to a larger guard band requirement. The coordinated case values are based on the GSM UE emissions and the uncoordinated values on LTE UE blocking requirements. It may be possible to push the LTE spectrum requirements down further for coordinated deployment depending on the GSM UE power levels and the allowed LTE uplink interference levels. The limiting factor is the maximum allowed interference to the PUCCH RBs that are located at the edge of the carrier.

The carrier spacing definition is illustrated in Figure 9.28. Figure 9.29 shows the expansion of the LTE carrier bandwidth when the GSM traffic decreases. Only seven GSM carriers need to be removed to make room for LTE 1.4 MHz and 15 GSM carriers for LTE 3.0 MHz.

9.9 Dimensioning

This section presents examples on how to convert the cell throughput values to the maximum number of broadband subscribers. Figure 9.30 shows two methods: a traffic volume based approach and a data rate based approach. The traffic volume based approach estimates the maximum traffic volume in gigabytes that can be carried by LTE 20 MHz $1+1+1$ configuration. The spectral efficiency is assumed to be 1.74 bps/Hz/cell using 2×2 MIMO. The busy

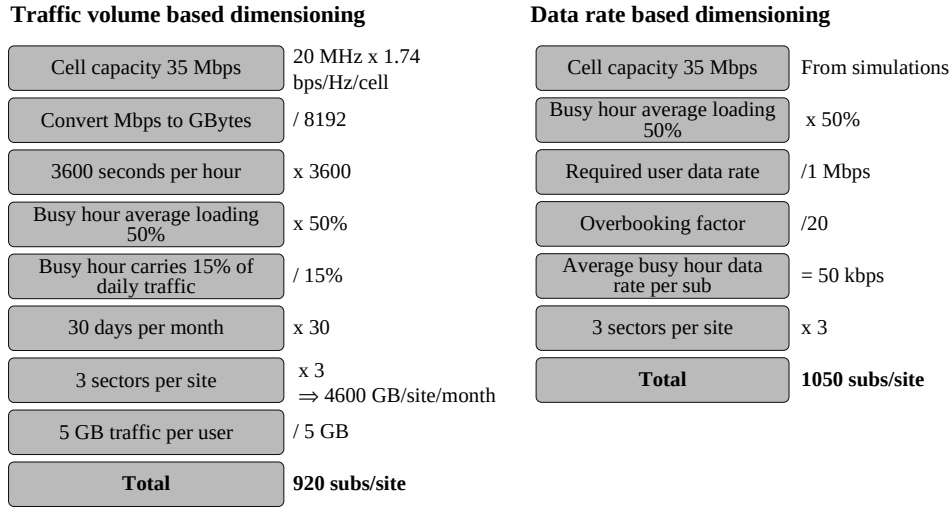


Figure 9.30 LTE dimensioning example for 1 + 1 + 1 at 20 MHz

hour is assumed to carry 15% of the daily traffic according to Figure 9.30 and the busy hour average loading is 50%. The loading depends on the targeted data rates during the busy hour: the higher the loading, the lower are the data rates. The maximum loading also depends on the applied QoS differentiation strategy: QoS differentiation pushes the loading closer to 100% while still maintaining the data rates for more important connections.

The calculation shows that the total site throughput per month is 4600 GB. To offer 5 GB data for every subscriber per month, the number of subscribers per site will be 920.

Another approach assumes a target of 1 Mbps per subscriber. Since only some of the subscribers are downloading data simultaneously, we can apply an overbooking factor, for example 20. This essentially means that the average busy hour data rate is 50 kbps per subscriber. The number of subscribers per site using this approach is 1050.

The calculation illustrates that LTE has the capability to support a large number of broadband data subscribers.

Figure 9.31 illustrates the technology and spectrum limits for the traffic growth assuming that HSPA and LTE use the existing GSM sites. The starting point is voice only traffic in a GSM network with 12 + 12 + 12 configuration, which corresponds to a high capacity GSM site found in busy urban areas. This corresponds to $12 \times 8 \times 0.016 = 1.5$ Mbps sector throughput assuming that each time slot carries 16 kbps voice rate. The voice traffic was the dominant part of the network traffic before flat rate HSDPA was launched. The data traffic has already exceeded the voice traffic in many networks in data volume. The second scenario assumes that the total traffic has increased 10 times compared to the voice only case. The sector throughput would then be 15 Mbps, which can be carried with three HSPA carriers using a 15 MHz spectrum.

The last scenario assumes 50 times more traffic compared to voice only, which leads to 75 Mbps and can be carried with two LTE carriers each of 20 MHz with a total 40 MHz of spectrum. The site throughput will be beyond 200 Mbps, setting corresponding requirements for the transport network capacity also.

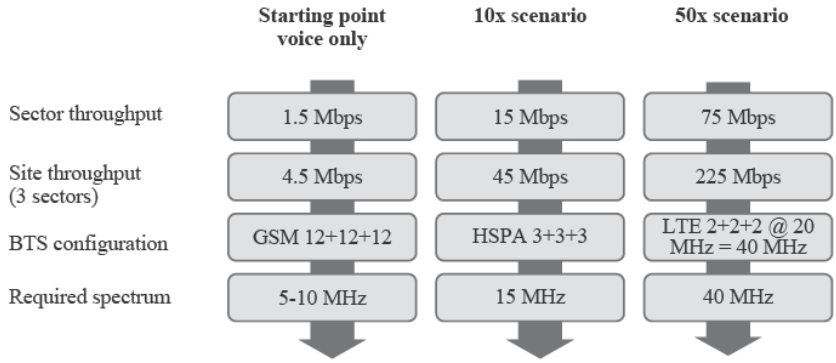


Figure 9.31 Traffic growth scenarios with 10 times and 50 times more traffic

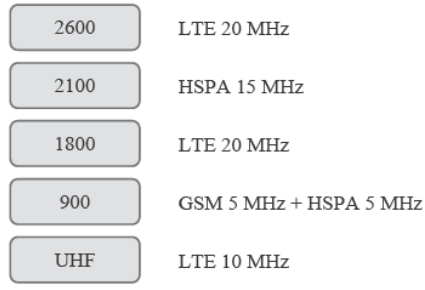


Figure 9.32 Example of a European operator with good spectrum resources

An example of the availability of the spectrum resources by a European operator within a few years is illustrated in Figure 9.32: GSM with 5 MHz, HSPA with 20 MHz and LTE with 50 MHz. Such an amount of spectrum would allow the traffic to increase more than 50 times compared to the voice only scenario. There are many operators in Asia and Latin America with less spectrum resources, which makes it more difficult to provide the high broadband wireless capacities.

9.10 Capacity Management Examples from HSPA Networks

In this section some of the HSDPA traffic analysis in RNC and at the cell level is shown and the implications discussed. It is expected that the analysis from broadband HSDPA networks will also be useful for the dimensioning of broadband LTE networks. All the analysis is based on the statistics of a single RNC and up to 200 NodeBs. The NodeBs were equipped with a 5-code round robin type shared HSDPA scheduler where five codes of HSDPA are shared among three cells and the transport solution was 2*E1 per NodeB. The maximum power for HSDPA was limited to 6 W. This area was selected to be analyzed as in this RNC the RF capability and transport capability are in line with each other, i.e. the transport solution can deliver the same throughput as the shared 5-code HSDPA scheduler. First it is necessary to evaluate the cell level data volume

fluctuations and contributions to RNC level total data volume so that the RNC throughput capacity can be checked. Then the cell level data volume and throughput limits are evaluated for when the new throughput improving features (proportional fair scheduler, cell dedicated scheduler, more codes, code multiplexing and so on) for the radio interface are introduced.

9.10.1 Data Volume Analysis

Figure 9.33 shows the data volume distribution over 24 h for the RNC on a typical working day. It can be seen that the single hour data volume share is a maximum of 6% from the total daily traffic volume and the fluctuation is just 3% to 6%. Also the hourly data volume share from the busy hour data volume share is 50% during the early morning hours and steadily increases towards the busiest hours from 8 pm to 1 am. The 3 hours from 9 pm to midnight are the busiest hours during the day. The usage increases heavily after about 6 pm, which indicates that as the working day ends then is the time for internet usage.

Looking into the individual cell contribution to the total RNC level data volume in Figure 9.34, it can be seen that during the night when the data volume is low and mobility is low, the traffic is also heavily concentrated on certain areas (cells) and during the day the share of cells contributing to the total data volume also increases.

As can be seen from Figure 9.34 during the early morning hours, 10% of the cells contribute 70–85% of the total RNC level data volume whereas during the busiest hours the same 70–85% data volume is contributed by 19–25% of the cells. During the early morning hours the data volume is very concentrated on just a couple of cells, which means that cells covering the residential areas should have very high data volume during the night time and early morning and due to low mobility the channel reservation times should be extremely long. This is shown in Figure 9.35, which indicates that during the early morning hours the

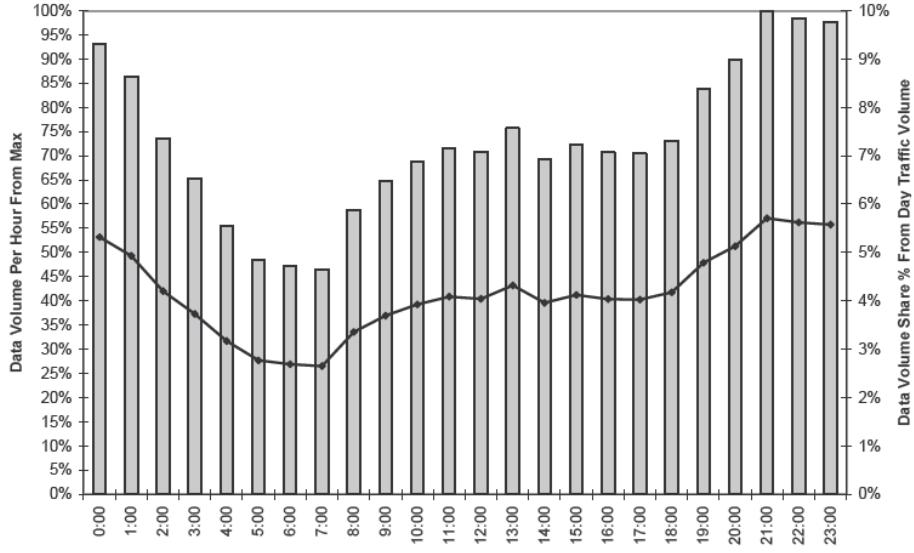


Figure 9.33 Daily RNC level hourly data volume deviation

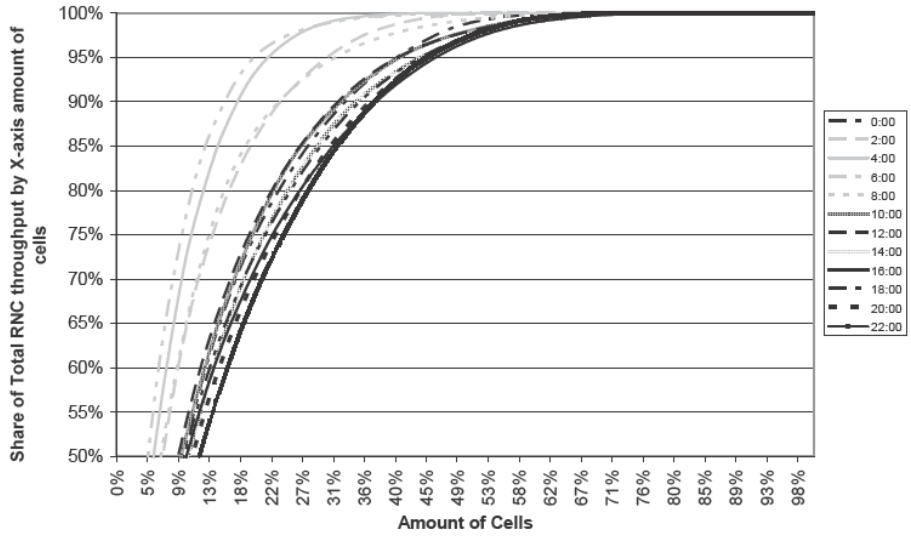


Figure 9.34 Cells' data volume contribution to total RNC data volume

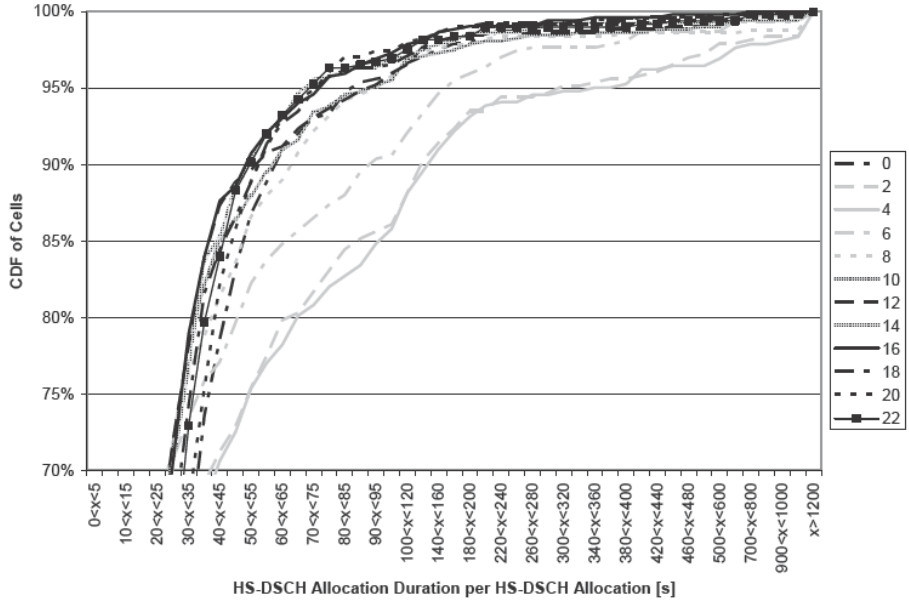


Figure 9.35 Average HS-DSCH allocation duration CDF of cells

average HS-DSCH allocation duration under the cell is a lot longer than during the highest usage. The lack of mobility and potentially some file sharing dominating usage during night time means that certain cells have extremely high data volumes at certain hours only and some cells have a fairly high usage and low data volume variation throughout the whole day. This makes efficient cell dimensioning a challenging task as if capacity is given to a cell according to the busy hour needs, then some cells are totally empty during most times of the day.

In Figure 9.36, the data volume fluctuation per hour is shown as the cumulative distribution of cells having a certain percentage of data volume from the highest data volume during the specific hour. From the graphs it can be seen that during the early morning and low data volume times there is also the highest number of cells having the lowest data volume. This indicates a very high fluctuation of data volume between all the cells.

Figure 9.37 shows the busy hour share of the total daily traffic on the cell level. The cells are sorted according to the data volume: the cells on the left have the highest traffic volume. The busy hour carries 10–20% of the daily traffic in busy cells. Figure 9.38 shows the distribution of the cells depending on the percentage of the data carried by the busy hour. The most typical values are 10–15% and 15–20%. We need to note that the busy hour share on the RNC level was only 6%. The difference between the cell and RNC level traffic distribution is explained by the trunking gain provided by RNC since the traffic is averaged over hundreds of cell.

9.10.2 Cell Performance Analysis

The cell performance analysis is carried out using the key indicators below:

- Active HS-DSCH Throughput – typically given in this analysis as kbps, it is the throughput under a cell when data have been sent in TTIs. Put simply it is the amount of data (kbit)/ number of active TTIs (s) averaged over a 1 h measurement period.
- HSDPA Data Volume – typically given in this analysis as Mbyte, kbit or Mbit, it is the amount of data sent per cell during the 1 h measurement period.
- Average number of simultaneous HSDPA users, during HSDPA usage – the amount of simultaneous users during the active TTIs, i.e. how many users are being scheduled during active TTIs per cell (the maximum amount depends on operator purchased feature set). When taking the used application into account, the average number of simultaneous users during HSDPA usage needs to be replaced by the number of active users who have data in the NodeB buffers. Averaged over the 1 h measurement period.
- Allocated TTI share – the average number of active TTIs during the measurement period (i.e. when there are data to send, the TTI is active) over all the possible TTIs during the 1 h measurement period.
- Average Throughput per User – typically given as kbps, it is the Active HS-DSCH Throughput/ Average number of simultaneous HSDPA users adjusted by the allocated TTI share; it is the average throughput that one single user experiences under a certain cell. When taking into account the used application, the Active HS-DSCH throughput needs to be divided by the number of active users who have data in the NodeB buffer. Averaged over the 1 h measurement period.
- Average reported Channel Quality Indicator (CQI) – every UE with HS-DSCH allocated measures and reports the CQI value back to the BTS. The average reported CQI is the

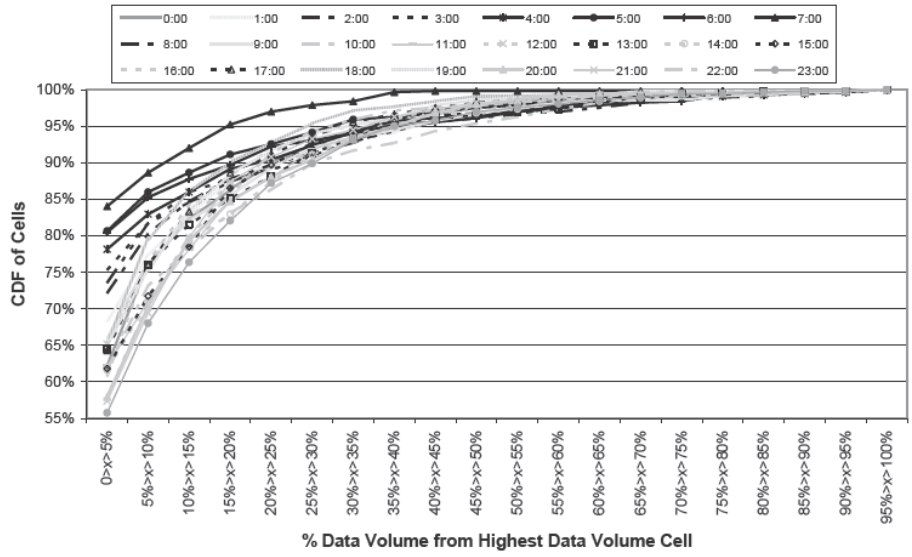


Figure 9.36 Percentage data volume per cell from highest data volume cell per hour

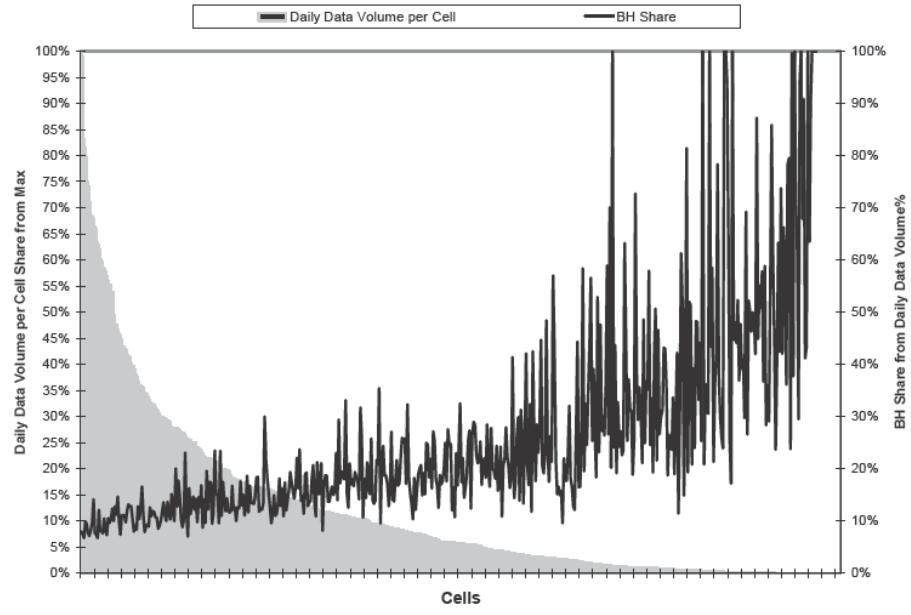


Figure 9.37 Busy hour share of cell level data volume

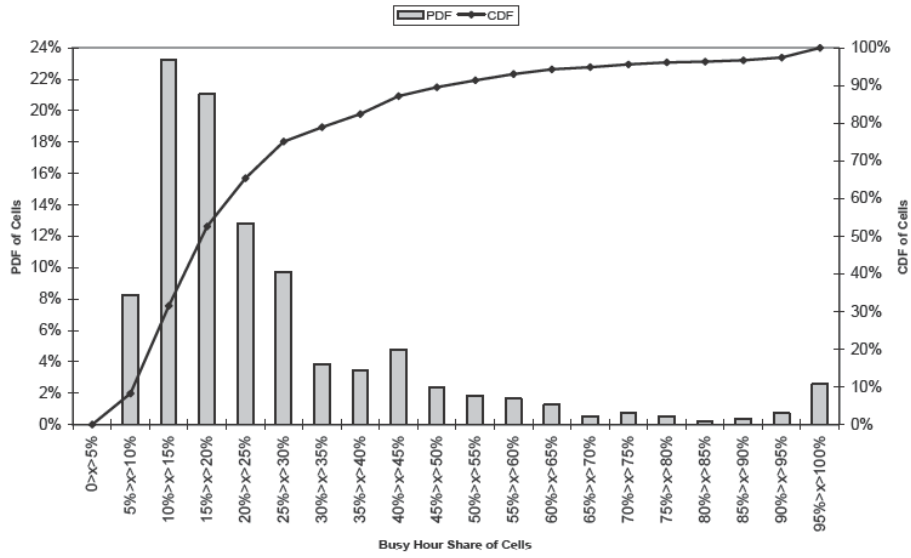


Figure 9.38 CDF and PDF of data volume busy hour share

CQI value reported by all the UEs under the cell averaged over the 1 h measurement period.

The end user throughput depends then on the average active HS-DSCH throughput and the average number of simultaneous HSDPA users. Figure 9.39 shows the average throughput as a function of the average number of simultaneous users. The average active HS-DSCH throughput is approximately 1.2 Mbps. When the allocated TTI share is low, the end user throughput approaches the HS-DSCH throughput. When the allocated TTI share increases, the end user throughput is reduced. This calculation is, however, pessimistic since it assumes that all users are downloading all the time.

When the used application activity is taken into account, i.e. the active HS-DSCH throughput is divided by the number of users who have data in the NodeB buffer, the average throughput per user is quite different from the formula used above. Figure 9.40 shows the comparison of the two different average throughput per user formulas and the Active HS-DSCH throughput. It should be noted that Figure 9.39 and Figure 9.40 are not from exactly the same time. The end user throughput is 400–800 kbps when only those users are considered that have data in the buffer. This can be explained when analyzing the difference between the number of simultaneous users and the number of simultaneous users who have data in the NodeB buffer as shown in Figure 9.41.

Figure 9.41 shows that the used application plays a significant role in the end user throughput evaluation and it should not be ignored. Therefore, the average user throughput may be low because the application simply does not need a high average data rate. The network performance analysis needs to separate the data rate limitations caused by the radio network and the actual used application data rate.

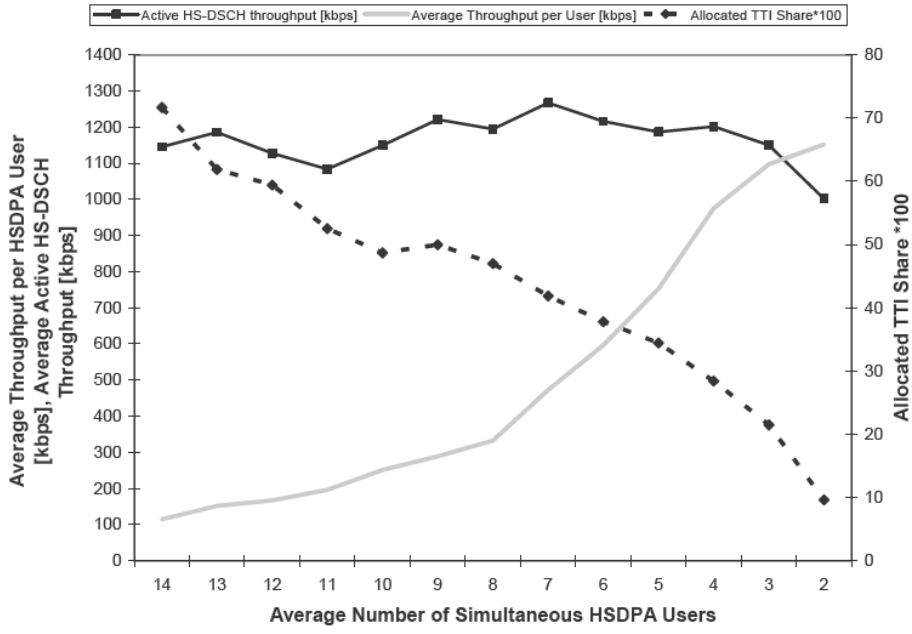


Figure 9.39 Average throughput per user

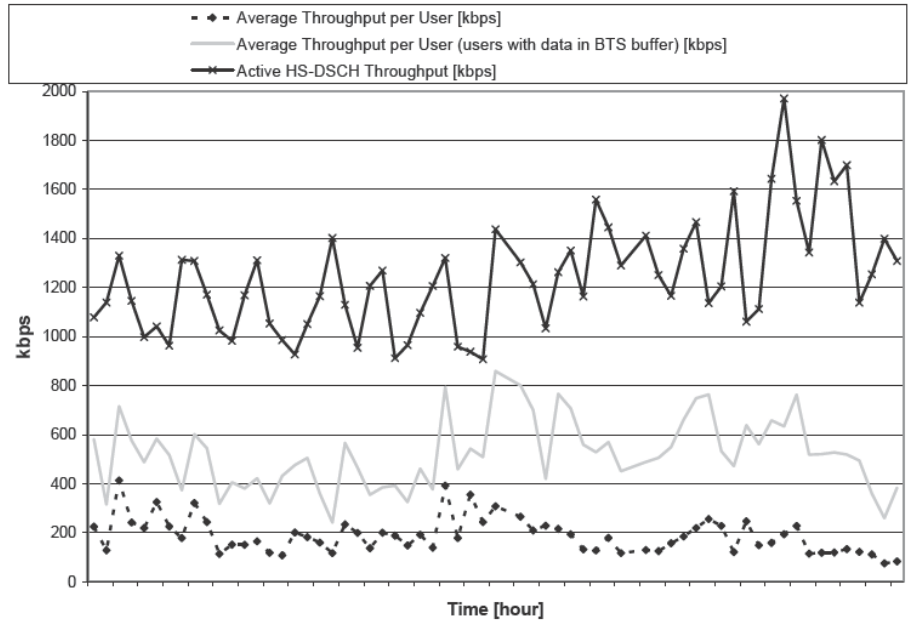


Figure 9.40 Average throughput per user, comparisons of different formulas

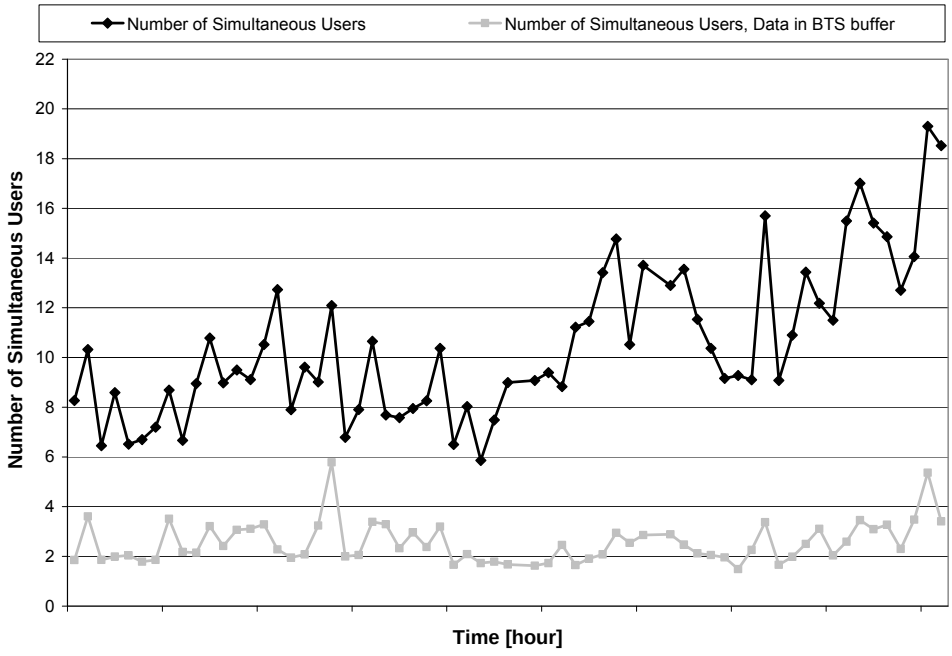


Figure 9.41 Number of simultaneous users with and without taking the used application into account

9.11 Summary

3GPP LTE Release 8 enables a peak bit rate of 150 Mbps in downlink and 50 Mbps in uplink with 2 × 2 MIMO antenna configuration in downlink and 16QAM modulation in uplink. The bit rates can be pushed to 300 Mbps in downlink with 4 × 4 MIMO and 75 Mbps in uplink with 64QAM. It is expected that the first LTE deployments will provide bit rates up to 150 Mbps.

The LTE link level performance can be modeled with the theoretical Shannon limit when suitable correction factors are included. LTE link level performance was shown to be robust with high mobile speeds, and the uplink LTE performance can be optimized by using adaptive transmission bandwidth. The link level simulations are used in the link budget calculations, which indicate that the LTE link budget is similar to the HSPA link budget with the same data rate and same spectrum.

The LTE system performance is optimized by orthogonal transmission schemes, by MIMO transmission and by frequency domain scheduling. The spectral efficiency can be further enhanced with multi-antenna transmission and higher order sectorization. The high efficiency can be maintained for different LTE bandwidths between 5 and 20 MHz, while the spectral efficiency is slightly lower with the narrowband LTE carriers 1.4 and 3.0 MHz. It was shown that LTE provides higher spectral efficiency compared to HSPA and HSPA evolution especially due to the frequency domain packet scheduling.

The user plane latency in LTE can be as low as 10–20 ms. The low latency is relevant for improving the end user performance since many applications and protocols benefit from low latency. The low latency is enabled by the short sub-frame size of 1 ms.

The flexible refarming of LTE to the GSM spectrum is enabled by the narrowband options: the refarming can be started with 1.4 or 3.0 MHz and later expanded when the GSM traffic has decreased. All UEs need to support all bandwidths between 1.4 and 20 MHz.

The dimensioning of the broadband wireless networks is different from voice networks. The examples from HSPA networks illustrate that the traffic distribution over the day, over the geographical area and the user mobility need to be considered.

References

[1] 3GPP Technical Specification 36.213 'Physical layer procedures', v. 8.3.0.
[2] 3GPP Technical Specification 36.306 v8.2.0: 'User Equipment (UE) radio access capabilities', August 2008.
[3] P. Mogensen *et al.* 'LTE Capacity compared to the Shannon Bound', *IEEE Proc. Vehicular Technology Conference*, pp. 699–703, April 2007.
[4] 3GPP Technical Specification 25.942 'Radio Frequency (RF) system scenarios', v. 7.0.0.
[5] 3GPP Technical Report 25.996 'Spatial channel model for Multiple Input Multiple Output (MIMO) simulations', V.7.0.0.
[6] I.Z. Kovács *et al* , 'Effects of Non-Ideal Channel Feedback on Dual-Stream MIMO OFDMA System Performance', *IEEE Proc. Veh. Technol. Conf.*, Oct. 2007
[7] I.Z. Kovács *et al.*, 'Performance of MIMO Aware RRM in Downlink OFDMA', *IEEE Proc. Veh. Technol. Conf.*, pp. 1171–1175, May 2008.
[8] S. Kumar, *et al*., 'Performance Evaluation of 6-Sector-Site Deployment for Downlink UTRAN Long Term Evolution', *IEEE Proc. Vehicular Technology Conference*, September 2008.
[9] B. Hagerman, D. Imbeni, J. Barta, A. Pollard, R. Wohlmuth, P. Cosimini, 'WCDMA 6 Sector Deployment-Case study of a real installed UMTS-FDD Network', *Vehicular Technology Conference*, Vol. 2, pp. 703–707, Spring 2006.
[10] K.I. Pedersen, P.E. Mogensen, B. Fleury, 'Spatial Channel Characteristics in Outdoor Environments and Their Impact on BS Antenna System Performance', *IEEE Proc. Vehicular Technology Conference*, pp. 719–723, May 1998.
[11] 3GPP TSG RAN R1-072444 'Summary of Downlink Performance Evaluation', May 2007.
[12] 3GPP TSG RAN R1-072261 'LTE Performance Evaluation – Uplink Summary', May 2007.

10

Voice over IP (VoIP)

Harri Holma, Juha Kallio, Markku Kuusela, Petteri Lundén, Esa Malkamäki, Jussi Ojala and Haiming Wang

10.1 Introduction

While the data traffic and the data revenues are increasing, the voice service still makes the majority of operators' revenue. Therefore, LTE is designed to support not only data services efficiently, but also good quality voice service with high efficiency. As LTE radio only supports packet services, the voice service will also be Voice over IP (VoIP), not Circuit Switched (CS) voice. This chapter presents the LTE voice solution including voice delay, system performance, coverage and inter-working with the CS networks. First, the general requirements for voice and the typical voice codecs are introduced. The LTE voice delay budget calculation is presented. The packet scheduling options are presented and the resulting system capacities are discussed. The voice uplink coverage challenges and solutions are also presented. Finally, the LTE VoIP inter-working with the existing CS networks is presented.

10.2 VoIP Codecs

GSM networks started with Full rate (FR) speech codec and evolved to Enhanced Full Rate (EFR). The Adaptive Multi-Rate (AMR) codec was added to 3GPP Release 98 for GSM to enable codec rate adaptation to the radio conditions. AMR data rates range from 4.75 kbps to 12.2 kbps. The highest AMR rate is equal to the EFR. AMR uses a sampling rate of 8 kHz, which provides 300–3400 Hz audio bandwidth. The same AMR codec was included also for WCDMA in Release 99 and is also used for running the voice service on top of HSPA. The AMR codec can also be used in LTE.

The AMR-Wideband (AMR-WB) codec was added to 3GPP Release 5. AMR-WB uses a sampling rate of 16 kHz, which provides 50–7000 Hz audio bandwidth and substantially better voice quality and mean opinion score (MOS). As the sampling rate of AMR-WB is double the sampling rate of AMR, AMR is often referred to as AMR-NB (narrowband). AMR-WB data rates range from 6.6 kbps to 23.85 kbps. The typical rate is 12.65 kbps, which is similar to the

LTE for UMTS: OFDMA and SC-FDMA Based Radio Access Edited by Harri Holma and Antti Toskala
 ISBN: 978-0-470-99401-6

normal AMR of 12.2 kbps. AMR-WB offers clearly better voice quality than AMR-NB with the same data rate and can be called wideband audio with narrowband radio transmission. The radio bandwidth is illustrated in Figure 10.1 and the audio bandwidth in Figure 10.2. The smallest bit rates, 1.8 and 1.75 kbps, are used for the transmission of Silence Indicator Frames (SID).

This chapter considers AMR codec rates of 12.2, 7.95 and 5.9 kbps. The resulting capacity of 12.2 kbps would also be approximately valid for AMR-WB 12.65 kbps.

When calling outside mobile networks, voice transcoding is typically required to 64 kbps Pulse Code Modulation (PCM) in links using ITU G.711 coding. For UE-to-UE calls, the transcoding can be omitted with transcoder free or tandem free operation [1]. Transcoding generally degrades the voice quality and is not desirable within the network.

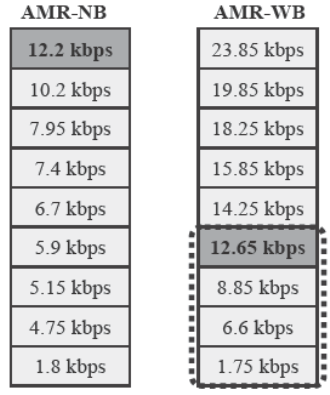


Figure 10.1 Adaptive Multirate (AMR) Voice Codec radio bandwidth

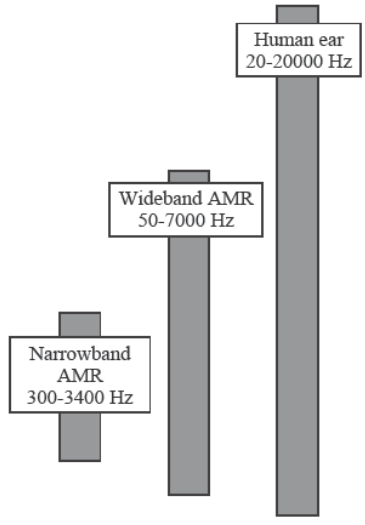


Figure 10.2 Adaptive Multirate (AMR) Voice Codec audio bandwidth

There are also a number of other voice codecs that are generally used for VoIP. A few examples are G.729, using an 8 kbps coding rate, and Internet Low Bit Rate Codec (iLBC), using 13 kbps which is used, for example, in Skype and in Googletalk.

10.3 VoIP Requirements

There is a high requirement set for the radio network to provide a reliable and good quality voice service. Some of the main requirements are considered below.

The impact of the mouth-to-ear latency on user satisfaction is illustrated in Figure 10.3. The delay preferably should be below 200 ms, which is similar to the delay in GSM or WCDMA voice calls. The maximum allowed delay for a satisfactory voice service is 280 ms.

IP Multimedia Subsystem (IMS) can be deployed to control VoIP. IMS is described in Chapter 3. IMS provides the information about the required Quality of Service (QoS) to the radio network by using 3GPP standardized Policy and Charging Control (PCC) [3]. The radio network must be able to have the algorithms to offer the required QoS better than Best Effort. QoS includes mainly delay, error rate and bandwidth requirements. QoS in LTE is described in Chapter 8.

The voice call drop rates are very low in the optimized GSM/WCDMA networks today – in the best case below 0.3%. VoIP in LTE must offer similar retainability including smooth interworking between GSM/WCDMA circuit switched (CS) voice calls. The handover functionality from VoIP in LTE to GSM/WCDMA CS voice is called Single radio Voice Call Continuity (SR-VCC) and described in detail in section 10.10.

The AMR 12.2 kbps packet size is 31 bytes while the IP header is 40-60 bytes. IP header compression is a mandatory requirement for an efficient VoIP solution. IP header compres-

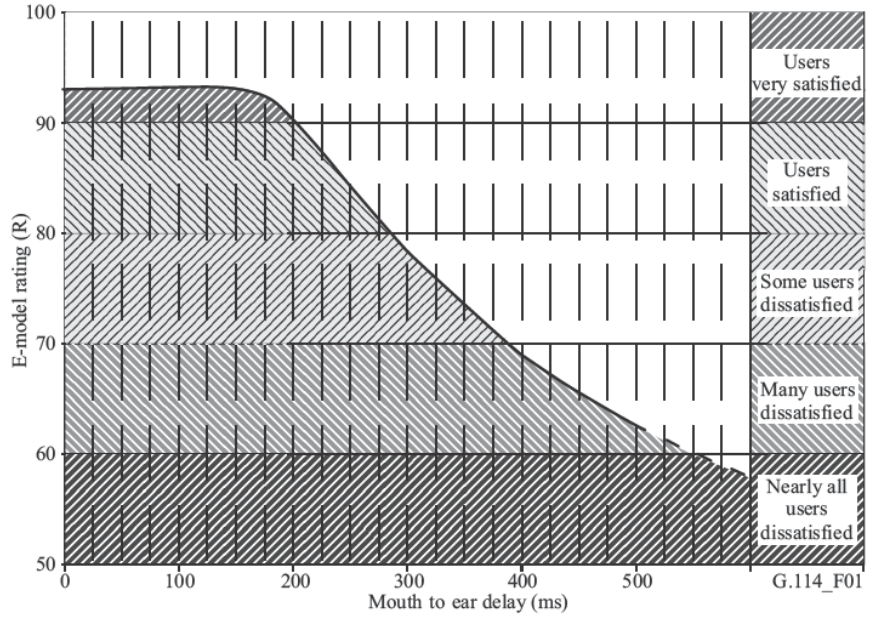


Figure 10.3 Voice mouth-to-ear delay requirements [2]

sion is required both in UE and in eNodeB. All the VoIP simulations in this chapter assume IP header compression.

The IP connectivity requires keep alive messages when the UE does not have a phone call running. The frequency of the keep alive messages depends on the VoIP solution: operator IMS VoIP can use fairly infrequent keep alive messages since IMS is within the operator's own network and no firewalls or Network Address Tables (NAT) are required in between. The internet VoIP requires very frequent keep alive message to keep the connectivity open through firewalls and NATs. The frequent keep alive message can affect UE power consumption and network efficiency.

VoIP roaming cases need further attention especially if there are some LTE networks designed for VoIP and data, while some networks are designed for data only transmission without the required voice features. VoIP roaming also requires IMS and GPRS roaming agreements and the use of visited GGSN/MME model. One option is to use circuit switched (CS) calls whenever roaming with CS Fallback for LTE procedures. Similarly CS calls can also be used for emergency calls since 3GPP Release 8 LTE specifications do not provide the priority information from the radio to the core network nor a specific emergency bearer.

10.4 Delay Budget

The end-to-end delay budget for LTE VoIP is considered here. The delay should preferably be below 200 ms, which is the value typically achieved in the CS network today. We use the following assumptions in the delay budget calculations. The voice encoding delay is assumed to be 30 ms including a 20 ms frame size, 5 ms look-ahead and 5 ms processing time. The receiving end assumes a 5 ms processing time for the decoding. The capacity simulations assume a maximum 50 ms air–interface delay in both uplink and downlink including scheduling delay and the time required for the initial and the HARQ retransmissions of a packet. We also assume a 5 ms processing delay in the UE, 5 ms in eNodeB and 1 ms in the SAE gateway. The transport delay is assumed to be 10 ms and it will depend heavily on the network configuration. The delay budget is presented in Figure 10.4. The mouth-to-ear delay is approximately 160 ms with these

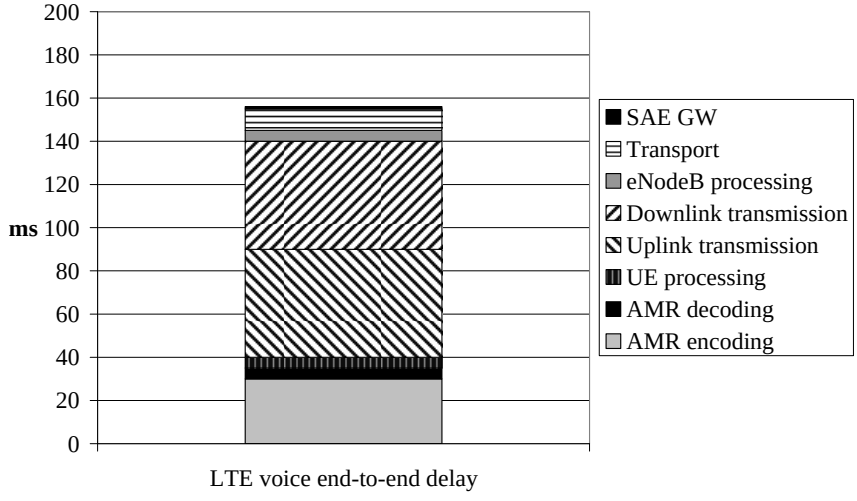


Figure 10.4 LTE voice end-to-end delay budget

assumptions, illustrating that LTE VoIP can provide lower end-to-end latency than CS voice calls today while still providing high efficiency.

10.5 Scheduling and Control Channels

By default the LTE voice service uses dynamic scheduling, where control channels are used to allocate the resources for each voice packet transmission and for the possible retransmissions. Dynamic scheduling gives full flexibility for optimizing resource allocation, but it requires control channel capacity. Multi-user channel sharing plays an important role when optimizing the air interface for VoIP traffic. Because each user is scheduled with control signaling with the dynamic scheduler, the control overhead might become a limiting factor for VoIP capacity. One solution in downlink for reducing control channel capacity consumption with a dynamic scheduler is to bundle multiple VoIP packets together at Layer 1 into one transmission of a user. The packet bundling is CQI based and is applied only for users whose channel conditions favor the use of bundling. The main benefit from the bundling is that more users could be fitted to the network with the same control channel overhead as users in good channel conditions are scheduled less often. The drawback from the bundling is that bundled packets experience a tighter delay budget, but the negative impact of this to VoIP capacity can be kept to a minimum by applying bundling for good enough users that are not relying on HARQ retransmissions. From the voice quality perspective it is also important to minimize the probability of losing consecutive packets. This can be achieved by making the link adaptation for the TTIs carrying bundled packets in a more conservative way leading to a reduced packet error rate for the first transmission. Due to the low UE transmission power and non-CQI based scheduling, packet bundling is not seen as an attractive technique in LTE uplink.

The 3GPP solution to avoid control channel limitation for VoIP capacity is the Semi-Persistent Scheduling (SPS) method [4, 5], where the initial transmissions of VoIP packets are sent without associated scheduling control information by using persistently allocated transmission resources instead. The semi-persistent scheduling is configured by higher layers (Radio Resource Control, RRC), and the periodicity of the semi-persistent scheduling is signaled by RRC. At the beginning of a talk spurt, the semi-persistent scheduling is activated by sending the allocated transmission resources by Physical Downlink Control Channel (PDCCH) and the UE stores the allocation and uses it periodically according to the periodicity. With semi-persistent scheduling only retransmissions and SID frames are scheduled dynamically, implying that the control channel capacity is not a problem for the semi-persistent scheduler. On the other hand, only limited time and frequency domain scheduling gains are available for semi-persistent scheduling. The semi-persistent allocation is released during the silence periods. Semi-persistent scheduling is adopted for 3GPP Release 8. The downlink operation of the semi-persistent scheduler is illustrated in Figure 10.5.

In the following the impact of control channel capacity to VoIP maximum capacity with dynamic scheduler is illustrated with theoretical calculations.

It is assumed that the number of Control Channel Elements (CCE) is approximately 20 per 5 MHz bandwidth. Two, four or eight CCEs can be aggregated per user depending on the channel conditions in low signal-to-noise ratios. We further apply the following assumptions for the calculations: voice activity 50%, voice packet arrival period 20 ms, downlink share of the traffic 50% and number of retransmissions 20%. The results are calculated without packet

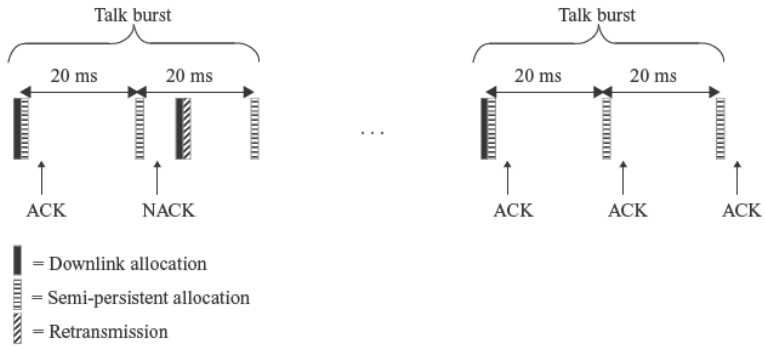


Figure 10.5 Semi-persistent scheduling in downlink

bundling and with two packet bundling. To simplify the calculations it is further assumed that SID frames are not taken into account. VoIP maximum capacity can be calculated by using Equation 10.1.

$$Max\_users = \frac{\#CCEs}{\#CCEs\_per\_user} \cdot \frac{Packet\_period[ms]}{Voice\_activity} \cdot Packet\_bundling \cdot Downlink\_share \cdot \frac{1}{1 + BLER} \tag{10.1}$$

The calculation results are shown in Figure 10.6 for a 5 MHz channel. As an example, the maximum capacity would be 330 users without CCE aggregation and without packet bundling. Assuming an average CCE aggregation of three brings the maximum capacity to 110 without packet bundling and 220 with packet bundling. These theoretical calculations illustrate the importance that multi-user channel sharing has for VoIP system performance, and furthermore, the maximum gain in capacity that packet bundling may provide with the dynamic

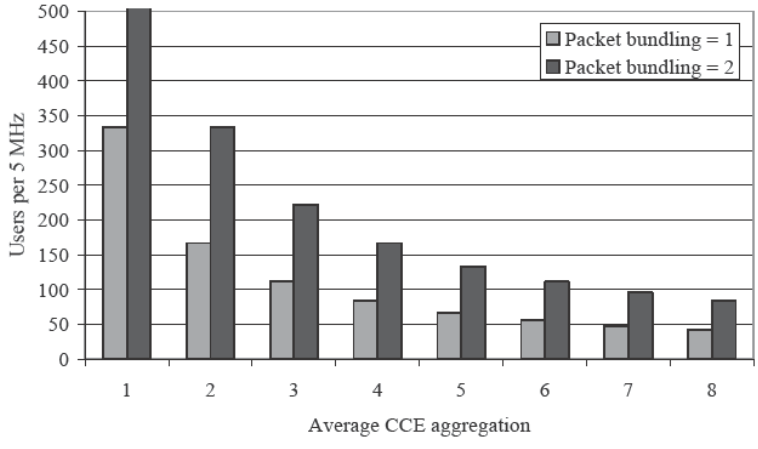


Figure 10.6 Maximum control channel limited capacity with fully dynamic scheduling

scheduler. According to the system level simulations, the average size of CCE aggregation is approximately 1.5 for a scheduled user. Comparison of the simulated downlink capacities[1] for a dynamic scheduler (presented in section 10.6) with the theoretical upper limits shows that the simulated capacities without packet bundling match rather nicely the theoretical calculations, being approximately 5% lower than the theoretical maximum. With packet bundling the simulated capacities are notably below the theoretical upper limits, as in practice the probability for using bundling is clearly limited below 100%.

10.6 LTE Voice Capacity

This section presents the system level performance of VoIP traffic in LTE at a 5 MHz system bandwidth. VoIP capacity is given as the maximum number of users that could be supported within a cell without exceeding 5% outage. A user is defined to be in an outage if at least 2% of its VoIP packets are lost (i.e. either erroneous or discarded) during the call. VoIP capacity numbers are obtained from system level simulations in a macro cell scenario 1 [6], the main system simulation parameters being aligned with [7]. An illustration of a VoIP capacity curve is presented in Figure 10.7, which shows how the outage goes up steeply when the maximum capacity is approached. This enables running the system with a relatively high load (compared to the maximum capacity) while still maintaining low outage.

The results of VoIP capacity simulations are summarized in Table 10.1 for three different AMR codecs (AMR 5.9, AMR 7.95 and AMR 12.2) and for both dynamic and semi-persistent schedulers. The capacity is 210–470 users in downlink and 210–410 users in uplink, which

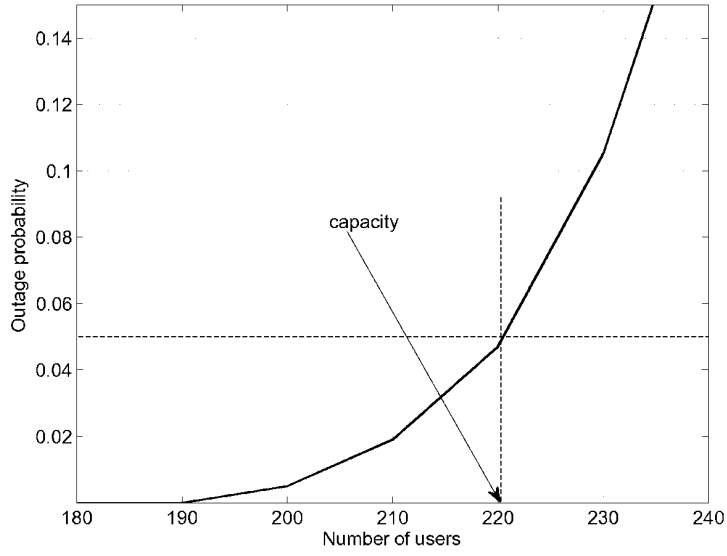


Figure 10.7 An example of VoIP capacity curve

[1] Simulated capacities should be up-scaled by 12.5% to remove the impact of SID transmissions on the capacity.

Table 10.1 VoIP capacity in LTE at 5 MHz

VoIP codec	AMR 5.9	AMR 7.95	AMR 12.2
Downlink capacity			
Dynamic scheduler, without bundling	210	210	210
Dynamic scheduler, with bundling	410	400	370
Semi-persistent scheduler	470	430	320
Uplink capacity			
Dynamic scheduler	230	230	210
Semi-persistent scheduler	410	320	240

corresponds to 42–94 users per MHz per cell in downlink and 42–82 users per MHz per cell in uplink. The lower AMR rates provide higher capacity than the higher AMR rates. The AMR codec data rate can be defined by the operator allowing a tradeoff between the capacity and the voice quality. The lower AMR rates do not increase the capacity with dynamic scheduling due to the control channel limitation.

The results also show that voice capacity is uplink limited, which can be beneficial when there is asymmetric data transmission on the same carrier taking more downlink capacity. The downlink offers higher capacity than uplink because the downlink scheduling uses a point-to-multipoint approach and can be optimized compared to the uplink.

In the following, the simulated capacities are analyzed in more detail. All the supporting statistics presented in the sequel are assuming load as close as possible to the 5% outage.

As described in section 10.5, VoIP system performance with the dynamic scheduler is limited by the available PDCCH resources, and therefore the dynamic scheduler cannot fully exploit the Physical Downlink Shared Channel (PDSCH) air interface capacity as there are not enough CCEs to schedule the unused Physical Resource Blocks (PRBs). This is illustrated in Figure 10.8 for downlink, which contains a cumulative distribution function for the scheduled PRBs per Transmission Time Interval (TTI) with AMR 12.2. The total number of PRBs on 5 MHz bandwidth is 25. As can be seen from Figure 10.8, the average utilization rate for the dynamic scheduler is only 40% if packet bundling is not allowed.

Due to the control channel limitations, the savings in VoIP packet payload size with lower AMR rates are not mapped to capacity gains with the dynamic scheduler, if packet bundling is not used. The different codecs provide almost identical performance in downlink whereas there are small gains in capacity in uplink with lower AMR rates. This uplink gain originates from the improved coverage as more robust MCS could be used when transmitting a VoIP packet in uplink – this gain does not exist in the downlink direction as it is not coverage limited due to higher eNodeB transmission power.

When packet bundling is enabled, the average size of an allocation per scheduled user is increased and hence the air interface capacity of PDSCH can be more efficiently exploited, which leads to an increased average PRB utilization rate of 70%. Gains of 75–90% are achieved for capacity when packet bundling is enabled. The probability of using packet bundling in a macro cell scenario is approximately 70% for AMR 12.2 codec, and it slightly increases when the VoIP packet payload decreases. This explains the small (< 10%) gains in capacity achieved by reducing VoIP packet payload.

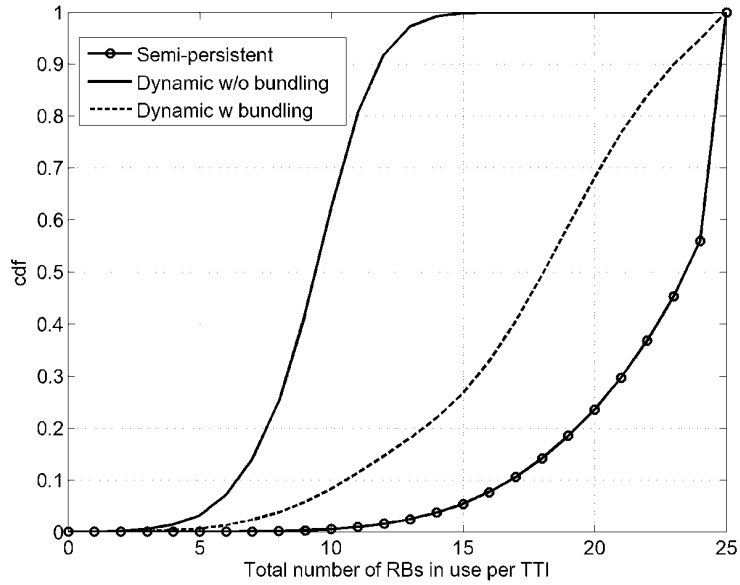


Figure 10.8 Cumulative distribution function for scheduled Physical Resource Blocks (PRB) per TTI in downlink (out of a total of 25 PRBs on 5 MHz bandwidth)

The performance of the semi-persistent scheduler does not suffer from control channel limitations, as initial transmissions are scheduled without associated control information by using persistently allocated time and frequency resources instead. The difference in control channel consumption for simulated packet scheduling methods is presented for the downlink direction in Figure 10.9, which contains the cumulative distribution function for the total number of consumed CCEs per TTI. With the dynamic scheduler all control channel elements are in use 70% of the time, if bundling is not allowed. With packet bundling, load is higher at 5% outage and hence control channel capacity could be more efficiently exploited implying that all CCEs are used almost 100% of the time. With the semi-persistent scheduler, the control channel overhead is clearly lower, the probability of using all CCEs being only slightly higher than 10%. Similar observations can be made from the distribution of consumed CCEs per TTI in the uplink direction, which is presented in Figure 10.10. As expected, the CCE consumption for semi-persistent scheduling is much lower than that for dynamic scheduling. Here we assume that the total number of CCEs reserved for downlink/uplink scheduling grants is 10.

The performance of the semi-persistent scheduler is not limited by the control channel resources, but it is limited by the available PDSCH bandwidth, which is illustrated in Figure 10.8, showing that the average PDSCH utilization rate for a semi-persistent scheduler is more than 90%. As the performance of the semi-persistent scheduler is user plane limited, the size of a VoIP packet has a significant impact on the capacity: an approximately 35% higher capacity is obtained if AMR 7.95 is used instead of AMR 12.2, whereas the use of AMR 5.9 instead of AMR 7.95 provides capacity gains of only 10%. The reason for the reduced gain for AMR 5.9 is presented in Figure 10.11, which shows the distribution of the size of the persistent resource allocation in terms of PRBs for all simulated codecs in the downlink direction. According to

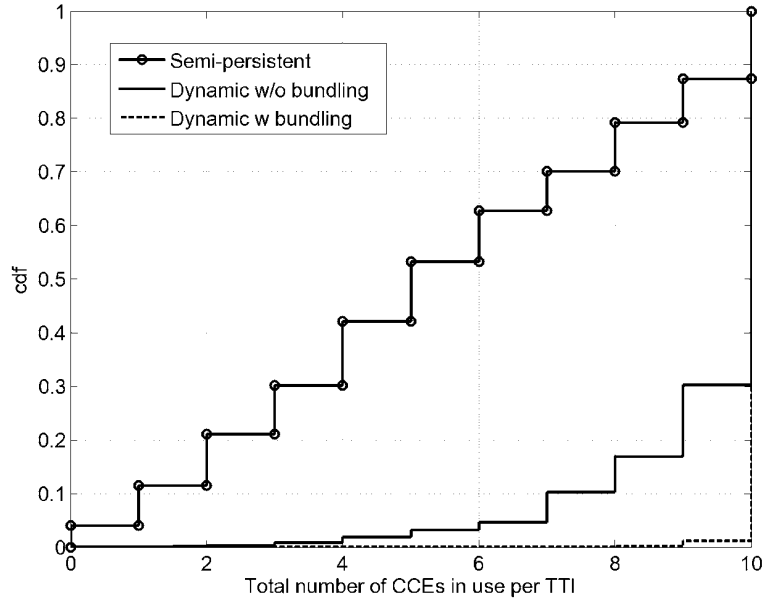


Figure 10.9 Cumulative distribution function for the total number of CCEs per TTI in downlink

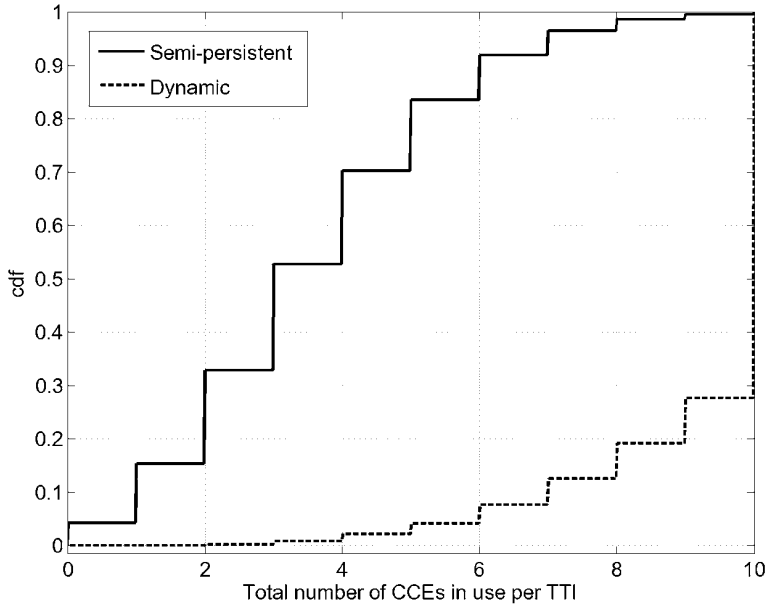


Figure 10.10 Cumulative distribution function for the total number of CCEs per TTI in uplink

the distribution, the size of the persistent allocation is one (1) PRB for almost 50% of the users with AMR 7.95, and hence no savings in allocated bandwidth can be achieved for those users when using AMR 5.9 instead of AMR 7.95. In the downlink the size of the persistent allocation

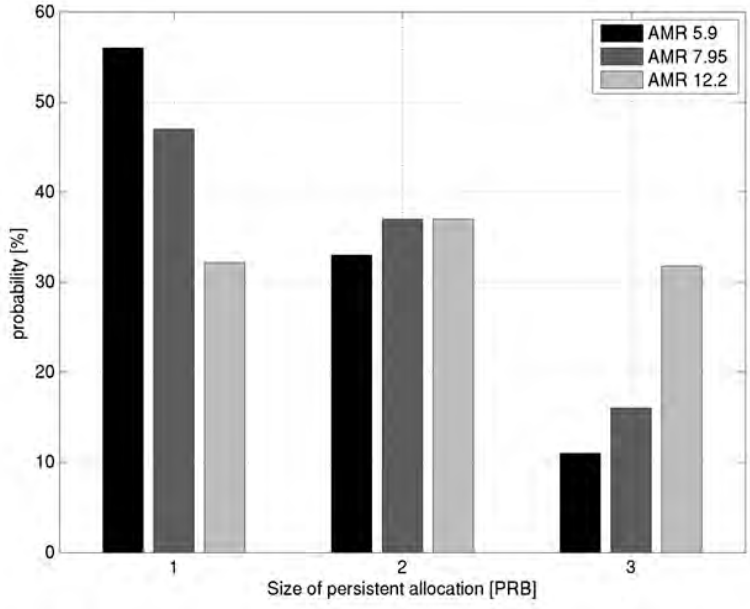


Figure 10.11 Probability distribution function for the size of the persistent resource allocation for different codecs in downlink

is calculated dynamically from wideband CQI for each talk spurt separately, whereas in the uplink, due to practical constraints, the size of the persistent resource allocation is decided from the path loss measurements of a user. The reason for the uplink solution is that the sounding is not accurate enough with a large number of users. The size of the persistent resource allocation has the following distribution during the uplink simulation: with AMR12.2 all the users have a persistent allocation of size two (2) PRBs, whereas with AMR 7.95 half of the users have a persistent allocation of size one (1) PRB, and half of the users have a persistent allocation of size two (2) PRBs. With AMR 5.9, 90% of the users have a persistent allocation of size one (1) PRB, and others have a persistent allocation of size two (2) PRB. An approximately 33% higher capacity is obtained if AMR 7.95 is used instead of AMR 12.2. Furthermore, AMR 5.9 provides approximately 28% gains in capacity over AMR 7.95. The reduced gain is mainly due to a slightly increased number of retransmissions for AMR 5.9.

When comparing the schedulers in the downlink direction for experienced air interface delay, it is observed that due to the control channel limitations the relative amount of packets experiencing delay close to the used delay bound (50 ms) is clearly higher for the dynamic scheduler. With packet bundling, control channel limitations can be relaxed, and therefore delay critical packets can be scheduled earlier. Bundling itself adds delays in steps of 20 ms. For the semi-persistent scheduler, the time the packet has to wait in the packet scheduler buffer before initial scheduling takes place depends on the location of the persistent resource allocation in the time domain, which has a rather even distribution with a full load. Therefore the packet delay distribution for the semi-persistent scheduler is smooth. In all these schemes, retransmissions are scheduled at the earliest convenience, after an 8 ms HARQ cycle. Due to the first HARQ retransmission the distributions plateau at 10 ms. A cumulative distribution function for the packet delay is presented in Figure 10.12.

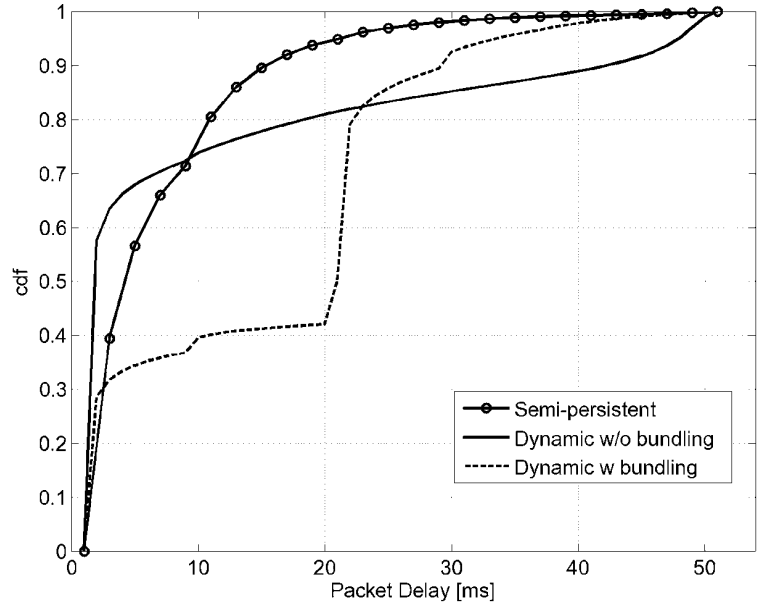


Figure 10.12 Cumulative distribution function for the experienced air interface delay in downlink with AMR 12.2

When summarizing VoIP system performance in downlink, it is concluded that the performance of the dynamic scheduler is control channel limited without packet bundling and hence the semi-persistent scheduler is able to have 50–125% higher capacities than the dynamic scheduler. The performance of the semi-persistent scheduler is user plane limited and the gains in capacity over the dynamic scheduler are smallest for AMR 12.2, which has the highest data rate amongst the simulated codecs and hence the biggest VoIP packet size. When the packet bundling is used with AMR 12.2, the control channel limitation for the performance of the dynamic scheduler is less significant compared with lower rate codecs as the number of supported users at 5% outage is lower. Therefore, the dynamic scheduler can have a 15% gain in capacity over the semi-persistent scheduler if AMR 12.2 is used with the packet bundling. When bundling is used with lower rate codecs, the control channel capacity starts to limit the performance of the dynamic scheduler due to the high number of supported users at 5% outage, and hence the semi-persistent scheduler achieves 8–15% higher capacities than the dynamic scheduler.

When comparing the performances of different scheduling methods in the uplink, it is observed that (similar to downlink) the performance of the dynamic scheduler is purely control channel limited whereas the semi-persistent scheduler suffers much less from control channel limitation due to a looser control channel requirement. Therefore, the semi-persistent scheduler is able to have 40–80% capacity gains over the dynamic scheduler with AMR 7.95 and AMR 5.9. Moreover, even with AMR 12.2 the semi-persistent scheduler achieves a 14% higher capacity than the dynamic scheduler.

The presented downlink results are obtained by using frequency dependent CQI reporting. In practice the large number of supported VoIP users in LTE may necessitate the use of wideband CQI to keep the overhead from CQI feedback at a reasonable level. This would mean a lower

Table 10.2 Relative loss in capacity (%) due to increased CQI granularity

	CQI reporting sub-band size	
	4 PRBs	Wideband CQI
Semi-persistent scheduling	2%	7%
Dynamic without bundling	0%	3%
Dynamic with bundling	7%	15%

uplink signaling overhead from CQI feedback at the cost of reduced capacity, as the frequency domain scheduling gains will be lost. To keep the capacity reduction as small as possible, the impact of lost frequency domain scheduling gains to performance should be compensated with more efficient use of frequency diversity. Therefore the use of the semi-persistent scheduler becomes even more attractive for VoIP traffic in LTE, as the performance of the semi-persistent scheduler is less dependent on the frequency domain scheduling gains than the dynamic scheduler. This is illustrated in Table 10.2, which presents the relative losses in capacity with AMR 12.2 when increasing the size of CQI reporting sub-band in the frequency domain. Relative losses are calculated against the presented capacity numbers, which were obtained assuming narrowband CQI reporting [7].

According to the simulations, the capacity of the dynamic scheduler is reduced by 15% due to the use of wideband CQI, whereas for the semi-persistent scheduler the corresponding loss is only 7%. Hence with wideband CQI the semi-persistent scheduler provides a similar performance to the dynamic scheduler for AMR 12.2, and for lower rate codecs the gains in capacity over the dynamic scheduler are increased further compared to the gains achieved with narrowband CQI.

Finally, in the light of the above results analysis, it seems that semi-persistent scheduling is the most attractive scheduling method for VoIP traffic in LTE. Nevertheless, as the used persistent allocation is indicated to the user via PDCCH signaling, some sort of combination of dynamic and semi-persistent scheduling methods seems to be the best option for VoIP in LTE. Furthermore, when comparing the performances of downlink and uplink, it is concluded that the VoIP performance in LTE is uplink limited. Hence downlink capacity can accommodate additional asymmetric data traffic, e.g. web browsing.

10.7 Voice Capacity Evolution

This section presents the evolution of the voice spectral efficiency from GSM to WCDMA/HSPA and to LTE. Most of the global mobile voice traffic is carried with GSM EFR or AMR coding. The GSM spectral efficiency can typically be measured with Effective Frequency Load (EFL), which represents the percentage of the time slots of full rate users that can be occupied in case of frequency reuse one. For example, EFL = 8% corresponds to 8% × 8 slots/200 kHz = 3.2 users/MHz. The simulations and the network measurements show that GSM EFR can achieve EFL 8% and GSM AMR can achieve 20% assuming all terminals are AMR capable and the network is optimized. The GSM spectral efficiency can be further pushed by the network feature Dynamic Frequency and Channel Allocation and by using Single Antenna Interference Cancellation (SAIC), known also as Downlink Advanced Receiver Performance (DARP). We assume up to EFL 35% with those enhancements. For further information see [8] and [9]. The

overhead from the Broadcast Control Channel (BCCH) is not included in the calculations. BCCH requires higher reuse than the hopping carriers.

WCDMA voice capacity is assumed to be 64 users with AMR 12.2 kbps, and 100 users with AMR 5.9 kbps on a 5 MHz carrier. HSPA voice capacity is assumed to be 123 users with AMR 12.2 kbps, and 184 users with AMR 5.9 kbps. For HSPA capacity evolution, see Chapter 13.

Capacity evolution is illustrated in Figure 10.13. LTE VoIP 12.2 kbps can provide efficiency which is 15× more than GSM EFR. The high efficiency can squeeze the voice traffic into a smaller spectrum. An example calculation is shown in Figure 10.14 assuming 1500 subscrib-

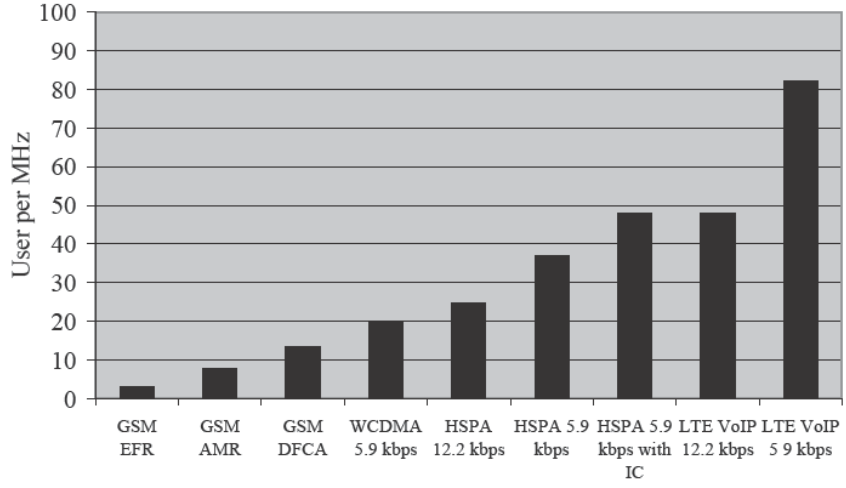


Figure 10.13 Voice spectral efficiency evolution (IC = Interference Cancellation)

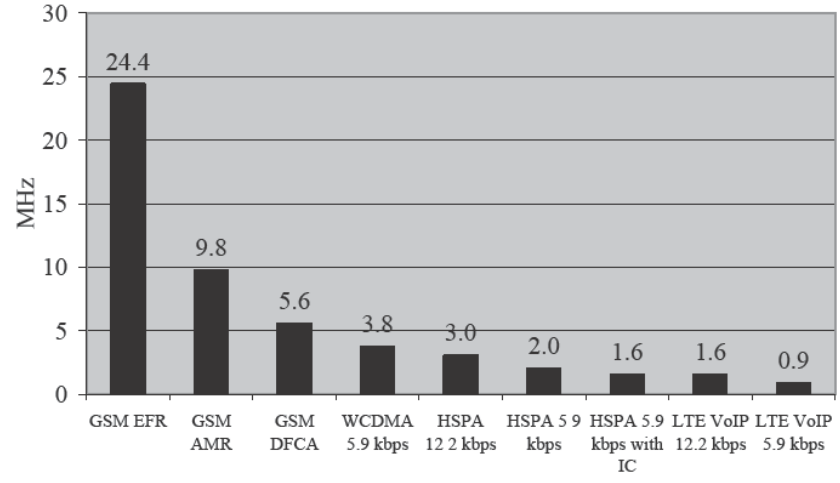


Figure 10.14 Spectrum required for supporting 1500 subscribers per sector at 40 mErl

ers per sector and 40 mErl traffic per subscriber. GSM EFR would take 25 MHz of spectrum while LTE will be able to carry that voice traffic in less than 2 MHz. LTE can then free up more spectrum for data usage.

10.8 Uplink Coverage

Uplink coverage can be maximized when UE transmits continuously with maximum power. Since VoIP packets are small (< 40 bytes), they can easily fit into 1 ms TTI. The AMR packets arrive every 20 ms leading to only 5% activity in the uplink. The uplink coverage could be improved by increasing the UE transmission time by using retransmissions and TTI bundling. These solutions are illustrated in Figure 10.15. The number of retransmissions must be limited for VoIP to remain within the maximum delay budget. The maximum number of retransmissions is six assuming a maximum 50 ms total delay, since the retransmission delay is 8 ms in LTE.

TTI bundling can repeat the same data in multiple (up to four) TTIs [4,10]. TTI bundling effectively increases the TTI length allowing the UE to transmit for a longer time. A single transport block is coded and transmitted in a set of consecutive TTIs. The same hybrid ARQ process number is used in each of the bundled TTIs. The bundled TTIs are treated as a single resource where only a single grant and a single acknowledgement are required. TTI bundling can be activated with a higher layer signaling per UE. The trigger could be, for example, UE reporting its transmit power is getting close to the maximum value.

The resulting eNodeB sensitivity values with retransmission and TTI bundling are shown in Table 10.3. The eNodeB sensitivity can be calculated as follows:

$$eNodeB\_sensitivity\,[dBm] = -174 + 10 \cdot \log_{10}(Bandwidth) + Noise\_figure + SNR \quad (10.2)$$

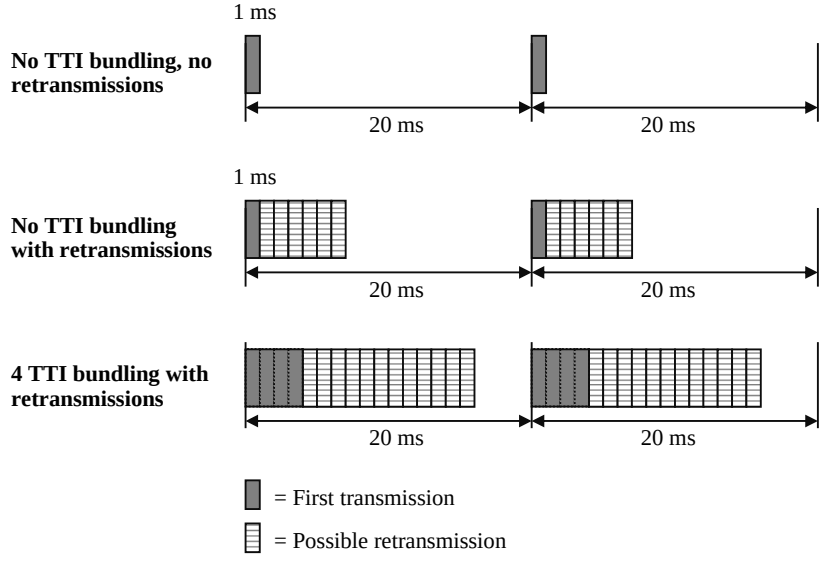


Figure 10.15 TTI bundling for enhancing VoIP coverage in uplink

Table 10.3 Uplink VoIP sensitivity with TTI bundling [11]

Number of TTIs bundled	1	4
Transmission bandwidth	360 kHz (2 resource blocks)	360 kHz (2 resource blocks)
Number of retransmissions	6	3
Required SNR	−4.2 dB	−8.0 dB
Receiver sensitivity	−120.6 dBm	−124.4 dBm

The eNodeB receiver noise figure is assumed to be 2 dB, two resource blocks are used for voice and no interference margin is included. The bundling can improve the uplink voice coverage by 4 dB.

The WCDMA NodeB sensitivity can be estimated by assuming Eb/N0 = 4.5 dB [12], which gives −126.6 dBm. To get similar voice coverage in LTE as in WCDMA we need to apply TTI bundling with a sufficient number of retransmissions. In terms of Hybrid-ARQ (HARQ) acknowledgement and retransmission timing, the TTI bundling method illustrated in Figure 10.16 is adopted for the LTE specifications [4,10]. According to this method four (4) subframes are in one bundle for the Frequency Division Duplex (FDD) system. Within a bundle, the operation is like autonomous retransmission by the UE in consecutive subframes without waiting for ACK/NACK feedback. The Redundancy Version (RV) on each autonomous retransmission in consecutive subframes changes in a pre-determined manner.

HARQ acknowledgement is generated after receiving the last subframe in the bundle. The timing relation between the last subframe in the bundle and the transmission instant of the HARQ acknowledgement is identical to the case of no bundling. If the last subframe in a bundle is subframe N then the acknowledgement is transmitted in subframe $N+4$. If the first subframe in a bundle is subframe k then any HARQ retransmissions begin in subframe k + 2 × HARQ RTT.

In the following, the impact of four (4) TTI bundling on capacity is studied via system level simulations in a bad coverage limited macro cell scenario 3 [2]. Main system simulation parameters are aligned with [7]. Table 10.4 presents both the energy accumulation and obtained

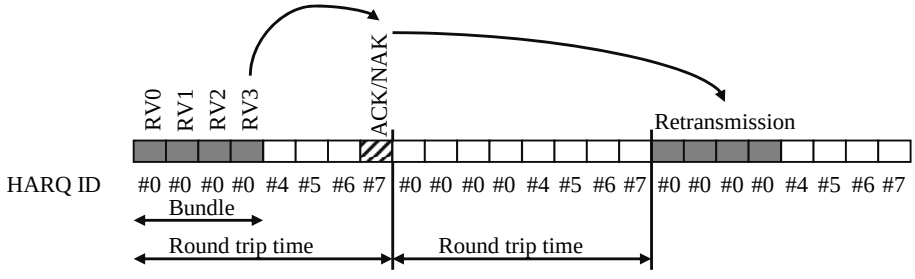


Figure 10.16 TTI bundling combining with HARQ

Table 10.4 Impact of TTI bundling on performance

	No bundling	TTI bundling
Maximal collected TTI per packet	7	12
VoIP capacity with TTI bundling at 5 MHz	<120	144

Table 10.5 Performance comparison for different delay budgets

Delay budget (ms)	50	60	70
Maximal collected TTI per packet	12	16	20
VoIP capacity with TTI bundling at 5 MHz	144	155	163

capacity for AMR 12.2 when dynamic scheduling with 50 ms packet delay budget is used. By using TTI bundling, 2.34 dB energy accumulation gain can be achieved, so VoIP capacity can be improved by 20% at least by using TTI bundling compared to no bundling.

Moreover, as the results in Table 10.5 show, the air interface delay has a clear impact on the achieved VoIP capacity and also on the gains that can be obtained from TTI bundling, i.e., with a looser delay budget capacity gains from TTI bundling are increased. This is evident as the longer air interface delay allows more energy to be accumulated from bundled TTIs, e.g. for 50 ms air interface delay, the packet will cumulate the energy from at most 12 TTIs assuming four (4) TTI bundling. Furthermore, for 70 ms air interface delay the corresponding number is increased to 20 TTIs, which would mean that 66% more energy could be accumulated. Hence the longer air interface delay implies bigger combining gains for the received signal from the bundled TTIs, and therefore the coverage is increased as the air interface delay is increased. This is again mapped to capacity gains. Besides, a longer delay budget means that the realized time domain scheduling gains are increased as each packet has more opportunities to be scheduled, and this further improves the coverage and hence improves the capacity. All in all, due to this reasoning the capacity with a 70 ms delay bound is approximately 13% higher than the capacity with a 50 ms delay bound.

10.9 Circuit Switched Fallback for LTE

Voice service with LTE terminals can also be offered before high quality VoIP support is included into LTE radio and before IP Multimedia Subsystem (IMS) is deployed. The first phase can use so-called Circuit Switched (CS) fallback for LTE where the LTE terminal will be moved to GSM, WCDMA or CDMA network to provide the same services that exists in CS networks today, for example voice calls, video call or Location Services [13]. The CS fallback procedures require that the Mobility Management Entity (MME) as well as Mobile Switching Center (MSC)/Visiting Location Register (VLR) network elements are upgraded to support procedures described in this section. The CS fallback is described in [13].

When LTE UE executes the attach procedure towards the LTE core network, the LTE core network will also execute a location update towards the serving CS core network to announce the presence of the terminal to the CS core network via the LTE network in addition to executing the normal attach procedures. The UE sends the attach request together with specific 'CS Fallback Indicator' to the MME, which starts the location update procedure towards MSC/VLR via an IP based SGs interface. The new Location Area Identity (LAI) is determined in the MME based on mapping from the Tracking area. A mapped LAI could be either GERAN or UTRAN based on the operator configuration. Additionally GERAN and UTRAN can be served by different MSC/VLR in the network architecture, which causes execution of a roaming retry for the CS fallback procedure as described later. The VLR creates an association with

the MME. This procedure is similar to the combined LA/RA update supported in GERAN by using Network Mode of Operation 1 (NMO1) with Gs interface.

If MME supports the connection to the multiple core network nodes similarly as SGSN does in GERAN/UTRAN, then a single MME can be connected to multiple MSC/VLR network elements in the CS core network. The benefit is the increased overall network resiliency compared to the situation when the MME is connected to only a single MSC/VLR.

For a mobile terminated call, MSC/VLR sends a paging message via SGs interface to the correct MME based on the location update information. eNodeB triggers the packet handover or network assisted cell change from LTE to the target system. The UE sends the paging response in the target system and proceeds with the CS call setup.

An overview of the mobile terminated fallback handover is shown in Figure 10.17.

An overview of the mobile originated call is shown in Figure 10.18. The UE sends a CS call request to the eNodeB, which may ask the UE to measure the target cell. eNodeB triggers the handover to the target system. the UE starts the normal CS call establishment procedure in the target cell. Once the CS call ends, the UE again reselects LTE to get access to high data rate capabilities.

If the service areas covered by LTE Tracking Area (TA) and GERAN/UTRAN Location Area (LA) are not similar, the serving MSC/VLR of LTE UE is different from MSC/VLR that serves the target LA. The target MSC/VLR is not aware of the UE and is not expecting a response to the paging message. This case is handled with a procedure called as Roaming Retry for CS fallback. When LTE UE has received the paging request, it executes an additional

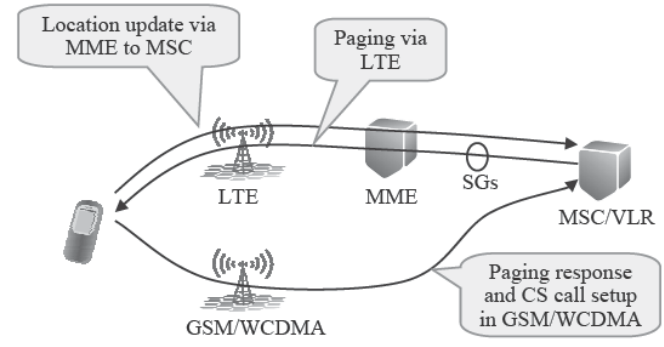


Figure 10.17 Circuit switched fallback handover – mobile terminated call

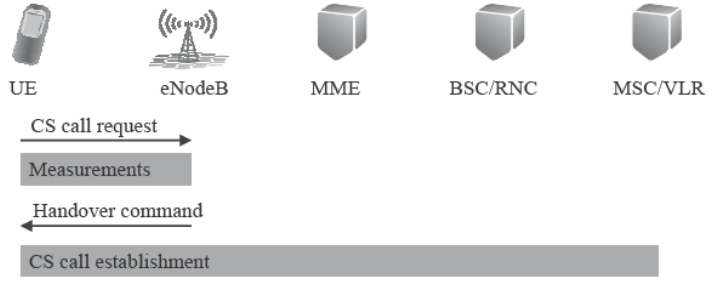


Figure 10.18 Circuit switched fallback for LTE – mobile originated call

location update procedure to the target MSC/VLR to make the target MSC/VLR aware of the existence of the UE. Based on this location update the HLR of the LTE UE will send a cancel message to the MSC/VLR that was serving the LTE UE prior to the CS fallback attempt. This cancel message will stop the paging process in that MSC/VLR and will trigger the resume call handling procedure. This procedure is similar to the Mobile Terminating Roaming Retry Call procedure used in GERAN/UTRAN to request a call to be re-routed to a new MSC/VLR from a gateway MSC. The gateway MSC will re-execute the mobile terminating call procedure by executing a HLR enquiry resulting in the call being routed to the correct MSC/VLR.

If there is an active packet session in LTE and if the target system supports simultaneous voice and data connection, the packet session is relocated from LTE to the target system. If the target system does not support simultaneous voice and data, the data connection is suspended during the CS voice call. This happens when LTE UE is moved to a GERAN that does not support simultaneous voice and data connection, called Dual Transfer Mode (DTM).

Voice is one of the most demanding services in terms of delay requirements and in terms of call reliability. Many optimized GERAN (GSM) and UTRAN (WCDMA) networks today can provide call drop rates below 0.5% or even below 0.3%. The advantage of the fallback handover is that there is no urgent need to implement QoS support in LTE or SAE. The network optimization may also be less stringent if the voice is not supported initially. Also, there is no need to deploy IMS just because of the voice service. From the end user point of view, the voice service in GSM or in WCDMA is as good as voice service in LTE, except that the highest data rates are not available during the voice call. On the other hand, the CS fallback for LTE adds some delay to the call setup process as the UE must search for the target cell. With high mobility, the CS fallback for LTE may also affect the call setup success rate, which naturally needs to be addressed in practical implementation.

Similar fallback procedures can also be used for other services familiar from today's network such as Unstructured Supplementary Service Data (USSD) and Location Services (LCS). Due to the popularity of SMS, the number of SMSs can be very high and the CS fallback handover may create a large number of reselections between LTE and GERAN/UTRAN. Therefore, it was seen as the preferred way to transmit SMSs over LTE even if the voice calls would use the CS fallback solution. If the core network has support for IMS, then SMS body can be transferred over SIP messages in LTE as defined by 3GPP [14].

10.10 Single Radio Voice Call Continuity (SR-VCC)

Once the VoIP call is supported in LTE, there may still be a need to make a handover from LTE VoIP to a GSM, WCDMA or CDMA CS voice network when running out of LTE coverage. The handover functionality from VoIP to the CS domain is referred to as Single Radio Voice Call Continuity (SR-VCC). The solution does not require UE capability to simultaneously signal on two different radio access technologies – therefore it is called a Single Radio Solution. SR-VCC is illustrated in Figure 10.19. SR-VCC is defined in [15]. The Dual radio Voice call continuity was already defined in 3GPP Release 7 for voice call continuation between WLAN and GSM/WCDMA networks [16].

The selection of the domain or radio access is under the network control in SR-VCC and SR-VCC enhanced MSC Server (called 'S-IWF' in this chapter) deployed into the CS core network. This architecture has been defined to enable re-use of already deployed CS core network assets to provide the necessary functionality to assist in SR-VCC. This architecture option is illustrated in Figure 10.20.

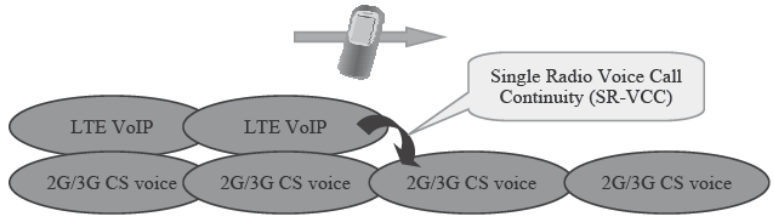


Figure 10.19 Single radio voice call continuity (SR-VCC)

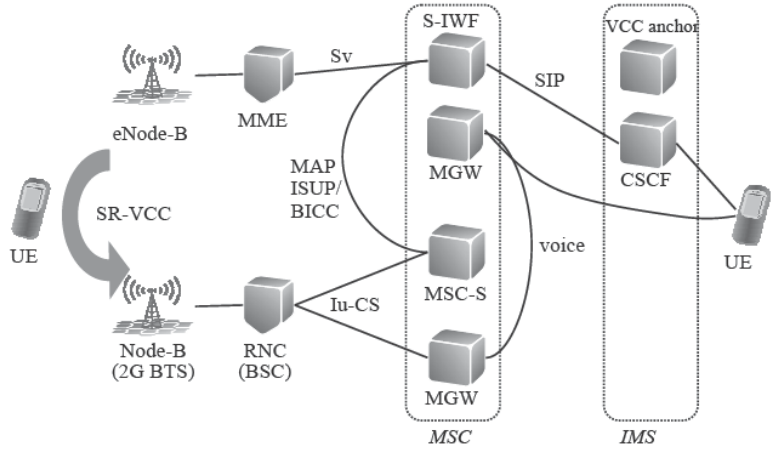


Figure 10.20 Architecture for SR-VCC. CSCF, call session control function; MGW, media gateway; MAP, mobile application port; ISUP, ISDN user part; BICC, bearer independent call control protocol

S-IWF, co-located in the MSC Server, uses a new GTP based Sv interface towards MME function in the LTE domain to trigger the SR-VCC procedure. Architecturally the S-IWF acts similarly to an anchor MSC Server towards target 2G/3G CS domain and is also responsible for preparing the needed target radio access network resources jointly with target GERAN or UTRAN. S-IWF also connects the speech path from the target CS domain towards the other subscriber in the voice call and together with the VCC anchor also hides mobility between LTE VoIP and 2G/3G CS domain from the other side of the call. The VCC anchor is located at the IMS application server and based on the same concept that was defined by 3GPP for Release 7 WLAN Voice Call Continuity.

S-IWF implementation has been specified so that it may be deployed with IMS Centralized Service (ICS) architecture so that the S-IWF then uses Mw interface (acting as the MSC Server enhanced for ICS) towards IMS instead of, for instance, the ISUP interface. In the former situation, handover causes registration prior to a new call towards IMS on behalf of the served subscriber whereas in the latter situation no registration is needed. This kind of support for SR-VCC with both ICS and non-ICS architectures is essential to achieve the required deployment flexibility.

During the LTE attach procedure, the MME will receive the required SR-VCC Domain Transfer Number from HSS that is then given to S-IWF via the Sv interface. S-IWF uses this

number to establish a connection to the VCC anchor during the SR-VCC procedure. When the VoIP session is established, it will be anchored within IMS in the VCC anchor to use SR-VCC in case it is needed later in the session. This anchoring occurs in both the originating and terminating sessions based on IMS subscription configuration.

The SR-VCC procedure is presented in Figure 10.21. LTE eNodeB first starts inter-system measurements of the target CS system. eNodeB sends the handover request to the MME, which then triggers the SR-VCC procedure via the Sv interface to the MSC-Server S-IWF functionality with Forward Relocation Request. S-IWF in the MSC-Server initiates the session transfer procedure towards IMS by establishing a new session towards the VCC anchor that originally anchored the session. The session is established by S-IWF using the SR-VCC number provided by MME. S-IWF also coordinates the resource reservation in the target cell together with the target radio access network. The MSC Server then sends a Forward Relocation Response to MME, which includes the necessary information for the UE to access the target GERAN/UTRAN cell.

After the SR-VCC procedure has been completed successfully the VoIP connection is present from Media Gateway that is controlled by S-IWF towards the other side of an ongoing session. The CS connection exists towards the radio access network to which the UE was moved during the procedure.

S-IWF functionality is not involved in a normal LTE VoIP session if the handover to CS domain is not required.

For a simultaneous voice and non-voice data connection, the handling of the non-voice bearer is carried out by the bearer splitting function in the MME. The MME may suppress

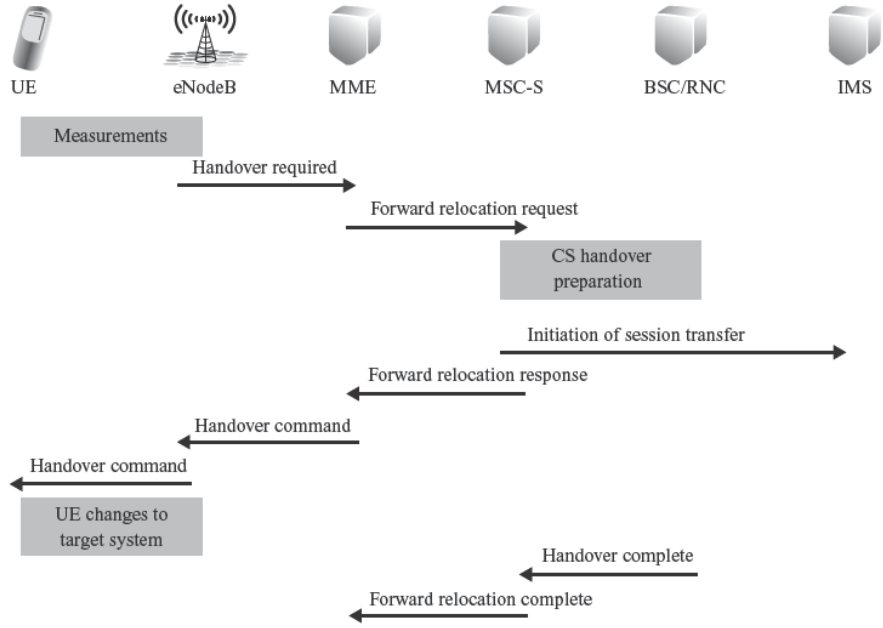


Figure 10.21 SR-VCC procedure

the handover of a non-voice PS bearer during the SR-VCC procedure. This may happen if the target GERAN system does not support simultaneous voice and data functionality (DTM). If the non-voice bearer is also handed over, the process is done in the same way as the normal inter-system handover for packet services. The MME is responsible for coordinating the Forward Relocation Response from the SR-VCC and the packet data handover procedure.

For roaming, the Visited Public Land Mobile Network (PLMN) controls the radio access and domain change while taking into account any related Home PLMN policies.

10.11 Summary

LTE radio is not only optimized for high data rates and high data capacities, but also for high quality voice with high capacity. The system simulations show that LTE can support 50–80 simultaneous voice users per MHz per cell in the macro cellular environment with AMR data rates between 12.2 and 5.9 kbps. The lower AMR rates offer the highest capacity. The highest capacity is achieved using semi-persistent packet scheduling where the first transmission has pre-allocated resources and only retransmissions are scheduled. The capacity with fully dynamic scheduling is limited by control channel capacity since each voice packet needs to be scheduled separately.

Voice quality depends on the voice codec and on the end-to-end delay. It is assumed that in the first phase, VoIP in LTE uses already standardized 3GPP speech codec AMR narrowband and AMR wideband due to backward compatibility with the legacy network. For end-to-end voice delay, LTE can provide values below 200 ms offering a similar or lower delay than the existing CS networks.

LTE uplink voice coverage can be optimized by using TTI bundling, which allows the combination of multiple 1-ms TTIs to increase the uplink transmission time and the average transmitted energy. The LTE voice coverage with TTI bundling is similar to the existing 2G/3G voice service.

As LTE supports only packet based services, then the voice service is also based on Voice over IP (VoIP). LTE specifications include definitions for inter-working with the existing Circuit Switched (CS) networks. One option is to use CS fallback for LTE where the UE moves to the 2G/3G network to make the voice call using a CS connection and returns to LTE when the voice call is over. Another option is to use VoIP in LTE and then allow inter-system handover to the 2G/3G CS network when LTE coverage is running out. This handover is called Single Radio Voice Call Continuity (SR-VCC). The CS fallback handover has been assumed to enable re-use of existing deployed CS core network investments as well as to provide a similar look and feel for the end users who are accustomed to current mobile network services. In some cases this kind of solution can be assumed to provide the first phase solution, but the IMS based core network architecture with support for VoIP and Multimedia Telephony is considered to be the long term solution.

Both CS fallback for LTE as well as the LTE based VoIP with SR-VCC solution can be deployed either in a phased manner or simultaneously depending on terminal capabilities. Additionally it is also possible to deploy IMS for non-VoIP services such as video sharing, presence or instant messaging, even primary voice service, by using CS fallback for LTE procedures. This kind of architecture ensures maximum flexibility for the operator to plan and deploy LTE as well as IMS solutions.

References

[1] 3GPP Technical Specifications 23.053 'Tandem Free Operation (TFO); Service description; Stage 2', V.8.3.0
[2] ITU-T Recommendation G.114 'One way transmission time', May 2003.
[3] 3GPP Technical Specification 23.203 'Policy and charging control architecture', V.8.3.1.
[4] 3GPP Technical Specifications 36.321 'Medium Access Control (MAC) protocol specification', V.8.3.0.
[5] 3GPP TS36.300, 'Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 8)'.
[6] 3GPP TR 25.814 v7.0.0, 'Physical layer aspect for evolved Universal Terrestrial Radio Access (UTRA)'.
[7] 'Next Generation Mobile Networks (NGMN) Radio Access Performance Evaluation Methodology', A White paper by the NGMN Alliance, January 2008.
[8] T. Halonen, J. Romero, J. Melero, 'GSM, GPRS and EDGE Performance', 2nd Edition, John Wiley & Sons, 2003.
[9] A. Barreto, L. Garcia, E. Souza, 'GERAN Evolution for Increased Speech Capacity', Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th, April 2007.
[10] 3GPP R2-082871, 'LS on TTI Bundling', RAN1#62, Kansas City, USA, May 2008.
[11] R1-081856, 'Coverage Comparison between PUSCH, PUCCH and RACH', Nokia/Nokia Siemens Networks, Catt, RAN1#54, Kansas City, USA, May 2008.
[12] H. Holma, A. Toskala, 'WCDMA for UMTS', 4th Edition, John Wiley & Sons, 2007.
[13] 3GPP Technical Specifications 23.272 'Circuit Switched (CS) fallback in Evolved Packet System (EPS)', V. 8.1.0.
[14] 3GPP Technical Specifications 23.204 'Support of Short Message Service (SMS) over generic 3GPP Internet Protocol (IP) access', V.8.3.0.
[15] 3GPP Technical Specifications 23.216 'Single Radio Voice Call Continuity (SRVCC)', V. 8.1.0.
[16] 3GPP Technical Specifications 23.206 'Voice Call Continuity (VCC) between Circuit Switched (CS) and IP Multimedia Subsystem (IMS)', V.7.5.0.

11

Performance Requirements

Andrea Ancora, Iwajlo Angelow, Dominique Brunel, Chris Callender, Harri Holma, Peter Muszynski, Earl McCune and Laurent Noël

11.1 Introduction

3GPP defines minimum Radio Frequency (RF) performance requirements for terminals (UE) and for base stations (eNodeB). These performance requirements are an essential part of the LTE standard as they facilitate a consistent and predictable system performance in a multi-vendor environment. The relevant specifications are given in [1], [2], [3] covering both duplex modes of LTE: Frequency Division Duplex (FDD) and Time Division Duplex (TDD).

Some of the RF performance requirements, e.g. limits for unwanted emissions, facilitate the mutual coexistence of LTE with adjacent LTE systems or adjacent 2G/3G systems run by different operators without coordination. The corresponding requirements may have been derived from either regulatory requirements or from coexistence studies within 3GPP; see [4].

Other performance requirements, e.g. modulation accuracy and baseband demodulation performance requirements, provide operators with assurance of sufficient performance within their deployed LTE carrier.

[2] is the basis for the eNodeB test specification [5], which defines the necessary tests for type approval. Correspondingly, [1] is the basis on the UE side for the test specifications [6] and [7].

This chapter presents the most important LTE minimum performance requirements, the rationale underlying these requirements and the implications for system performance and equipment design. Both RF and baseband requirements are considered. First the eNodeB requirements are considered and then the UE requirements.

11.2 Frequency Bands and Channel Arrangements

11.2.1 Frequency Bands

Table 11.1 lists the currently defined LTE frequency bands, together with the corresponding duplex mode (FDD or TDD). There are currently 17 bands defined for FDD and 8 bands for

LTE for UMTS: OFDMA and SC-FDMA Based Radio Access Edited by Harri Holma and Antti Toskala
 ISBN: 978-0-470-99401-6

Table 11.1 LTE frequency bands

LTE Band	Uplink eNode B receive UE transmit		Downlink eNode B transmit UE receive		Duplex mode
1	1920 MHz	– 1980 MHz	2110 MHz	– 2170 MHz	FDD
2	1850 MHz	– 1910 MHz	1930 MHz	– 1990 MHz	FDD
3	1710 MHz	– 1785 MHz	1805 MHz	– 1880 MHz	FDD
4	1710 MHz	– 1755 MHz	2110 MHz	– 2155 MHz	FDD
5	824 MHz	– 849 MHz	869 MHz	– 894 MHz	FDD
6	830 MHz	– 840 MHz	875 MHz	– 885 MHz	FDD
7	2500 MHz	– 2570 MHz	2620 MHz	– 2690 MHz	FDD
8	880 MHz	– 915 MHz	925 MHz	– 960 MHz	FDD
9	1749.9 MHz	– 1784.9 MHz	1844.9 MHz	– 1879.9 MHz	FDD
10	1710 MHz	– 1770 MHz	2110 MHz	– 2170 MHz	FDD
11	1427.9 MHz	– 1452.9 MHz	1475.9 MHz	– 1500.9 MHz	FDD
12	698 MHz	– 716 MHz	728 MHz	– 746 MHz	FDD
13	777 MHz	– 787 MHz	746 MHz	– 756 MHz	FDD
14	788 MHz	– 798 MHz	758 MHz	– 768 MHz	FDD
17	704 MHz	– 716 MHz	734 MHz	– 746 MHz	FDD
18	815 MHz	– 830 MHz	860 MHz	– 875 MHz	FDD
19	830 MHz	– 845 MHz	875 MHz	– 890 MHz	FDD
…					
33	1900 MHz	– 1920 MHz	1900 MHz	– 1920 MHz	TDD
34	2010 MHz	– 2025 MHz	2010 MHz	– 2025 MHz	TDD
35	1850 MHz	– 1910 MHz	1850 MHz	– 1910 MHz	TDD
36	1930 MHz	– 1990 MHz	1930 MHz	– 1990 MHz	TDD
37	1910 MHz	– 1930 MHz	1910 MHz	– 1930 MHz	TDD
38	2570 MHz	– 2620 MHz	2570 MHz	– 2620 MHz	TDD
39	1880 MHz	– 1920 MHz	1880 MHz	– 1920 MHz	TDD
40	2300 MHz	– 2400 MHz	2300 MHz	– 2400 MHz	TDD

TDD. Whenever possible, the RF requirements for FDD and TDD have been kept identical to maximize the commonality between the duplex modes.

While the physical layer specification and many RF requirements are identical for these frequency bands, there are some exceptions to this rule for the UE RF specifications, as will be discussed further in UE related sections in this chapter. On the other hand, the eNodeB RF requirements are defined in a frequency band agnostic manner as there are less implementation constraints for base stations. If the need arises, further LTE frequency bands can be easily added affecting only isolated parts of the RF specifications. Furthermore, the specified LTE frequency variants are independent of the underlying LTE release feature content (Release 8, 9, etc.). Those frequency variants that will be added during the Release 9 time frame can still be implemented using just Release 8 features.

The band numbers 15 and 16 are skipped since those numbers were used in ETSI specifications.

Not all of these bands are available in each of the world's regions. Table 11.2 illustrates the typical deployment areas for the different FDD frequency variants.

The most relevant FDD bands in Europe are:

Table 11.2 Usage of the frequency variants within the world's regions

Band	Europe	Asia	Japan	Americas
1	X	X	X	
2				X
3	X	X		
4				X
5		X		X
6			X	
7	X	X		
8	X	X		
9			X	
10				X
11			X	
12				X
13			X	
14				X
15				X
16				X
17				X
18			X	

- Band 7 is the new 2.6 GHz band. The 2.6 GHz auctions have been running in a few countries during 2007 and 2008, and continue during 2009.
- Band 8 is currently used mostly by GSM. The band is attractive from a coverage point of view due to the lower propagation losses. The band can be reused for LTE or for HSPA. The first commercial HSPA900 network started in Finland in November 2007. LTE refarming is considered in Chapter 9.
- Band 3 is also used by GSM, but in many cases Band 3 is not as heavily used by GSM as Band 8. That makes refarming for LTE simpler. Band 3 is also important in Asia where Band 7 is not generally available.

Correspondingly, the new bands in the USA are Bands 4, 12, 13, 14 and 17. Bands 2 and 5 can be used for LTE refarming. The LTE deployment in Japan will use Bands 1, 9, 11 and 18. In summary, LTE deployments globally will use several different frequency bands from the start.

Further frequencies for International Mobile Telephony (IMT) have been identified at the ITU WRC 2007 conference (see Chapter 1).

11.2.2 Channel Bandwidth

The width of a LTE carrier is defined by the concepts of Channel bandwidth ($BW_{Channel}$) and Transmission bandwidth configuration (N_{RB}) (see Figure 11.1**)**. Their mutual relationship is specified in Table 11.3.

The transmission bandwidth configuration, N_{RB}, is defined as the maximum number of Resource Blocks (RB) that can be allocated within a LTE RF channel. A RB comprises 12

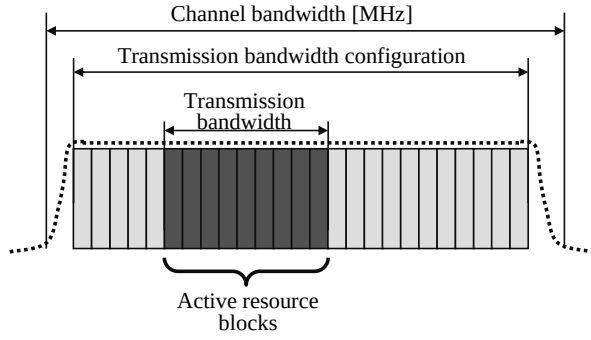


Figure 11.1 Definition of Channel Bandwidth and Transmission Bandwidth Configuration for one LTE carrier

Table 11.3 Transmission bandwidth configuration N_{RB} in LTE channel bandwidths

Channel bandwidth $BW_{Channel}$ [MHz]	1.4	3	5	10	15	20
Transmission bandwidth configuration N_{RB}	6	15	25	50	75	100

sub-carriers and can thus be understood to possess a nominal bandwidth of 180 kHz. While the physical layer specifications allow N_{RB} to assume any value within the range $6 \leq N_{RB} \leq 110$, all RF requirements (and thus a complete LTE specification) are only defined for the values in Table 11.3. This flexibility within the LTE specifications, however, supports the addition of further options for transmission bandwidth configurations (channel bandwidths), should the need arise.

The channel bandwidth is the relevant RF related parameter for defining Out-of-band (OOB) emission requirements (e.g. spectrum emission mask, Adjacent Channel Leakage Ratio [ACLR]). The RF channel edges are defined as the lowest and highest frequencies of the carrier separated by the channel bandwidth, i.e. at $F_C \pm BW_{Channel}/2$.

Note from Table 11.3 that the transmission bandwidth configuration measures only 90% of the channel bandwidth for 3, 5, 10, 15, and 20 MHz LTE and less so for 1.4 MHz LTE; e.g. for 5 MHz LTE a transmission bandwidth of $25 \times 180\,\text{kHz} = 4.5\,\text{MHz}$ is obtained. In fact, the basic OFDM spectrum comprises only slowly decaying sidelobes, with a rolloff as described by the *sinc()* function. Therefore, efficient usage of the spectrum requires the use of filtering to effectively confine the OOB emissions. Such filters, however, require a certain amount of transition bandwidth in order to be (a) practical and (b) to consume only a small amount of the cyclic prefix duration due to the inevitable time dispersion they will cause. The values in Table 11.3, in particular for the LTE channel bandwidth of 1.4 MHz, are therefore a compromise between in- and out-of-channel distortions and were extensively studied in 3GPP. The transmission bandwidth is 77% of the channel bandwidth for LTE 1.4 MHz.

The 1.4 and 3 MHz channel bandwidth options have been chosen to facilitate a multitude of cdma2000® and GSM migration scenarios within Bands 2, 3, 5 and 8.

Not all combinations of LTE frequency bands and channel bandwidth options are meaningful, and Table 11.4 provides the combinations supported by the standard. These combinations were based on inputs from the operators. Note from Table 11.4 that the options with a wider channel

Table 11.4 Supported transmission bandwidths with normal sensitivity ('X') and relaxed sensitivity ('O')

Channel bandwidth						
E-UTRA Band	1.4 MHz	3 MHz	5 MHz	10 MHz	15 MHz	20 MHz
1			X	X	X	X
2	X	X	X	X	O	O
3	X	X	X	X	O	O
4	X	X	X	X	X	X
5	X	X	X	O		
6			X	O		
7			X	X	X	O
8	X	X	X	O		
9			X	X	O	O
10			X	X	X	X
11			X	O	O	O
12						
13	X	X	O	O		
14	X	X	O	O		
17						
…						
33			X	X	X	X
34			X	X	X	
35	X	X	X	X	X	X
36	X	X	X	X	X	X
37			X	X	X	X
38			X	X		
39			X	X	X	X
40				X	X	X

bandwidth are typically supported for bands with a larger amount of available spectrum, e.g. 20 MHz LTE is supported in Bands 1, 2, 3 and 4 but not in Bands 5, 8, etc. Conversely, LTE channel bandwidths below 5 MHz are typically supported in frequency bands with either a smaller amount of available spectrum (e.g. Bands 5, 8, etc.) or bands exhibiting 2G migration scenarios (e.g. Bands 2, 5 and 8).

Furthermore [1] defines two levels of UE sensitivity requirements depending on the combination of bandwidths and operating band; these are also listed in Table 11.4. This is because there will be an increased UE self-interference in those frequency bands which possess a small duplex separation and/or gap. For large bandwidth and small duplex separation, certain relaxations of the UE performance are allowed or UE functionality is limited. For other cases, the UE needs to meet the baseline RF performance requirements.

11.2.3 Channel Arrangements

The channel raster is 100 kHz for all E-UTRA frequency bands, which means that the carrier centre frequency must be an integer multiple of 100 kHz. For comparison, UMTS uses a 200 kHz channel raster and for some frequency bands requires additional RF channels, offset by 100 kHz relative to the baseline raster, in order to effectively support the deployment within

a 5 MHz spectrum allocation. The channel raster of 100 kHz for LTE will support a wide range of migration cases.

The spacing between carriers will depend on the deployment scenario, the size of the frequency block available and the channel bandwidths. The nominal channel spacing between two adjacent LTE carriers is defined as follows:

$$\text{Nominal channel spacing} = (BW_{Channel(1)} + BW_{Channel(2)})/2 \tag{11.1}$$

where $BW_{Channel(1)}$ and $BW_{Channel(2)}$ are the channel bandwidths of the two respective LTE carriers. The LTE–LTE coexistence studies within 3GPP [4] assumed this channel spacing.

The nominal channel spacing can be adjusted, however, to optimize performance in a particular deployment scenario, e.g. for coordination between adjacent LTE carriers.

11.3 eNodeB RF Transmitter

LTE eNodeB transmitter RF requirements are defined in [2] and the corresponding test specification in [5]. In this section we discuss two of the most important transmitter requirements:

- Unwanted emissions, both inside and outside the operating band. The corresponding requirements will ensure the RF compatibility of the LTE downlink with systems operating in adjacent (or other) frequency bands.
- Transmitted signal (modulation) quality, or also known as Error Vector Magnitude (EVM) requirements. These requirements will determine the in-channel performance for the transmit portion of the downlink.

Coexistence of LTE with other systems, in particular between the LTE FDD and TDD modes, will also be discussed, both in terms of the 3GPP specifications as well as within the emerging regulative framework within Europe for Band 7.

11.3.1 Operating Band Unwanted Emissions

For WCDMA the spurious emissions requirements as recommended in ITU-R SM.329 are applicable for frequencies that are greater than 12.5 MHz away from the carrier centre frequency. The 12.5 MHz value is derived as 250% of the necessary bandwidth (5 MHz for WCDMA) as per ITU-R SM.329. The frequency range within 250% of the necessary bandwidth around the carrier centre may be referred to as the OOB domain. Transmitter intermodulation distortions manifest themselves predominantly within the OOB domain and therefore more relaxed emission requirements, such as ACLR, are typically applied within the OOB domain.

In LTE, the channel bandwidth can range from 1.4 to 20 MHz. A similar scaling by 250% of the channel bandwidth would result in a large OOB domain for LTE: for the 20 MHz LTE channel bandwidth option the OOB domain would extend to a large frequency range of up to ± 50 MHz around the carrier centre frequency. To protect services in adjacent bands in a more predictable manner and to align with WCDMA, the LTE spurious domain is defined to start from 10 MHz below the lowest frequency of the eNodeB transmitter operating band and from 10 MHz above the highest frequency of the eNodeB transmitter operating band, as shown in

Figure 11.2. The operating band plus 10 MHz on each side are covered by the LTE operating band unwanted emissions.

LTE spurious domain emission limits in [2] are defined as per ITU-R SM.329 and are divided into several categories, where Categories A and B are applied as regional requirements. Within Europe the Category B limit of −30 dBm/ MHz is required in the frequency range 1 GHz ↔ 12.75 GHz while the corresponding Category A limit applied in the Americas and in Japan is −13 dBm/ MHz.

In addition to the ITU-R SM.329 spurious emission limits, [2] defines more stringent limits across the operating bands of various wireless systems, including WCDMA, GSM and Personal Handyphone System (PHS).

The Operating band unwanted emission limits are defined as absolute limits by a mask that stretches from 10 MHz below the lowest frequency of the eNodeB transmitter operating band up to 10 MHz above the highest frequency of the eNodeB transmitter operating band, as shown in Figure 11.2.

This mask depends on the LTE channel bandwidth and is shown in Figure 11.3 for LTE bands >1 GHz and the limit of −25 dBm in 100 kHz as the lower bound for the unwanted emission limits. The limit of -25 dBm in 100 kHz is consistent with the level used for UTRA as a spurious emission limit inside the operating band (−15 dBm/ MHz). The measurement bandwidth of 100 kHz was chosen to be of similar granularity as the bandwidth of any victim system's smallest radio resource allocation (LTE 1 RB at 180 kHz; GSM 200 kHz carrier).

The operating band unwanted emission limits must also be consistent with the limits according to ITU-R SM.329 for all LTE channel bandwidth options. This means that outside 250% of the necessary bandwidth from the carrier centre frequency, the corresponding Category B limit of −25 dBm/100 kHz must be reached (see Figure 11.3). Even though the frequency range in which transmitter intermodulation distortions occur scales with the channel bandwidth, it was found feasible to define a common mask for the 5, 10, 15 and 20 MHz LTE options which meets the SM.329 limits at 10 MHz offset from the channel edge. However, this −25 dBm/100 kHz limit must already be reached with offsets of 2.8 MHz and 6 Mhz respectively from the channel edge for the LTE 1.4 MHz and LTE 3 MHz, and this necessitated definition of separate masks. For

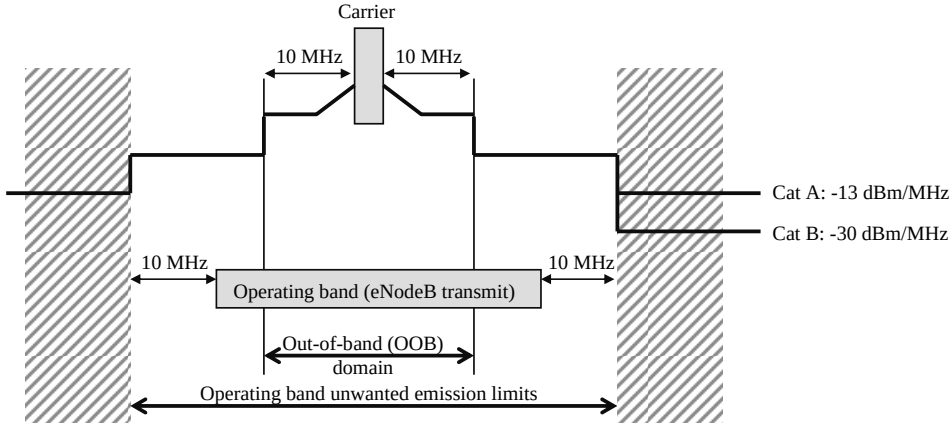


Figure 11.2 Defined frequency ranges for spurious emissions and operating band unwanted emissions

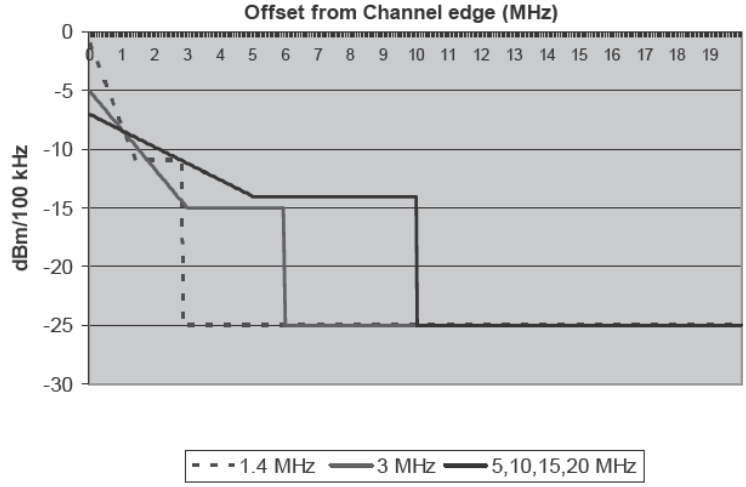


Figure 11.3 Operating band unwanted emission requirements levels relative to channel edge (E-UTRA bands >1 GHz)

the 1.4, 3 and 5 MHz LTE the same total eNodeB transmission power of 46 dBm was assumed resulting in a higher power spectral density for the smaller LTE bandwidth options, hence the mask permits higher emission levels at the channel edge.

For the North American LTE Bands (Bands 2, 4, 5, 10, 12, 13, 14, 17) an additional unwanted emission limit is derived in [2] from FCC Title 47 Parts 22, 24 and 27.These requirements are interpreted as −13 dBm in a measurement bandwidth defined as 1% of the '−26 dB modulation bandwidth' within the first MHz from the channel edge and −13 dBm/MHz elsewhere.

11.3.2 Coexistence with Other Systems on Adjacent Carriers Within the Same Operating Band

The RAN4 specifications also include ACLR requirements of 45 dBc for the 1st and 2nd adjacent channels of (a) the same LTE bandwidth and (b) 5 MHz UTRA (see Figure 11.4).

[4] contains simulation results for the downlink coexistence of LTE with adjacent LTE, UTRA or GSM carriers. The required Adjacent Channel Interference Ratio (ACIR) to ensure ≤5% cell edge throughput loss for the victim system was found to be about 30 dB for all of these cases. With an assumed UE Adjacent Channel Selectivity (ACS) of 33 dB for LTE and UTRA, the choice of 45 dB ACLR ensures minimal impact from the eNodeB transmit path and also aligns with the UTRA ACLR1 minimum performance requirements.

For LTE–LTE coexistence the ACLR1, ACLR2 are only specified for the same LTE bandwidth, i.e. mixed cases such as 5 MHz LTE ↔ 20 MHz LTE are not covered by explicit ACLR requirements according to all possible values for the victim carrier bandwidth. Analysis in [4], however, has shown that these cases are sufficiently covered by the 45 dBc ACLR requirements measured within the same bandwidth as the aggressing carrier. A large matrix of ACLR require-

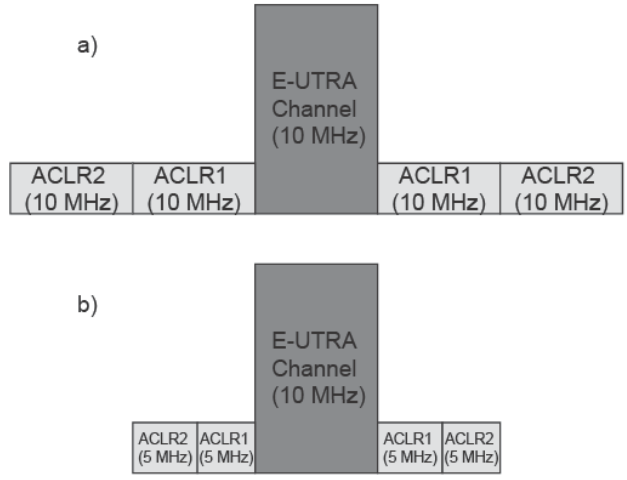


Figure 11.4 The two defined ACLR measures, one for 1st and 2nd adjacent E-UTRA carriers and one for 1st and 2nd adjacent UTRA carrier

ments for all LTE bandwidth options and, additionally, for multiple other radio technologies would have led to a large eNodeB conformance testing effort without materially adding more robust protection.

Furthermore, the unwanted emission mask specified for LTE provides a baseline protection for any victim system. For this reason no ACLR requirements were specified, for example, for GSM victim carriers – the corresponding ACLR values can be obtained by integrating the unwanted emission mask accordingly and are well above the required ~30 dB suggested by system simulations [4]. However, ACLR requirements for the 1st and 2nd WCDMA channels have been added to explicitly protect adjacent 'legacy' WCDMA carriers by the same level from aggressing LTE systems as specified for WCDMA as LTE migration within WCDMA bands was considered an important deployment scenario.

The ACLR2/3.84 MHz for a WCDMA victim is specified to be the same value as the ACLR1/3.84 MHz (45 dBc), not 50 dBc as for WCDMA. This is reasonable, as the second adjacent channel interference contributes only little to overall ACIR,

$$ACIR = \frac{1}{\frac{1}{ACLR1} + \frac{1}{ACS1} + \frac{1}{ACLR2} + \frac{1}{ACS2}} \tag{11.2}$$

because the WCDMA UE ACS2 in the second adjacent channel is significantly higher (~42 dBc) than the ACS1 (~33 dBc). Therefore, decreasing the ACLR2/3.84 MHz from 50 to 45 dBc has only negligible impact. On the other hand, an ACLR2/3.84 MHz requirement of 50 dBc for 10, 15 and 20 MHz LTE would be overly restrictive from an eNodeB implementation perspective

as the WCDMA ACLR2 frequency region (within 5 … 10 MHz offset from the LTE carrier edge) would still be within the ACLR1 region of the transmitting LTE carrier for which 45 dBc is a more appropriate requirement.

The ACLR values specified for LTE can also be compared to the corresponding values obtained by integrating the unwanted emission mask, together with an assumption on the eNodeB output power, e.g. 46 dBm. In fact, the integrated ACLR values obtained from the mask are ~2 … 5 dB more relaxed when compared with the specified ACLR values. This was specified on purpose, as downlink power control on RBs may lead to some 'ripples' of the leakage power spectrum within the OOB domain and the LTE unwanted emission mask was designed to be merely a 'roofing' requirement. These ripples would be averaged out within the ACLR measurement bandwidth and are therefore not detrimental from a coexistence perspective. Hence, in LTE, unlike WCDMA, the transmitter (PA) linearity is set by the ACLR and not the unwanted emission mask requirements, facilitating increased output power – for more information see [8].

Finally, unwanted emission mask and ACLR requirements apply whatever the type of transmitter considered (single carrier or multi-carrier). For a multi-carrier eNodeB, the ACLR requirement applies for the adjacent channel frequencies below the lowest carrier frequency used by the eNodeB and above the highest carrier frequency used by the eNodeB, i.e. not within the transmit bandwidth supported by the eNodeB, which is assumed to belong to the same operator.

11.3.3 Coexistence with Other Systems in Adjacent Operating Bands

[2] contains additional spurious emission requirements for the protection of UE and/or eNodeB of other wireless systems operating in other frequency bands within the same geographical area. The system operating in the other frequency band may be GSM 850/900/1800/1900, Personal Handyphone System (PHS), Public Safety systems within the US 700 MHz bands, WCDMA FDD/TDD and/or LTE FDD/TDD. Most of these requirements have some regulatory background and are therefore mandatory within their respective region (Europe, Japan, North America). The spurious emission limits for the protection of 3GPP wireless systems assume typically ~67 dB isolation (Minimum Coupling Loss, MCL) between the aggressor and victim antenna systems, including antenna gain and cable losses.

Moreover, to facilitate co-location (i.e. assuming only 30 dB MCL) with the base stations of other 3GPP systems, an additional set of optional spurious emission requirements has been defined in [2].

However, as shown by Figure 11.2, spurious emissions do not apply for the 10 MHz frequency range immediately outside the eNodeB transmit frequency range of an operating band. This is also the case when the transmit frequency range is adjacent to the victim band, hence the above coexistence requirements do not apply for FDD/TDD transition frequencies at, e.g., 1920, 2570, 2620 MHz. However, the deployment of FDD and TDD technologies in adjacent frequency blocks can cause interference between the systems as illustrated in Figure 11.5. There may be interference between FDD and TDD base stations and between the terminals as well.

From an eNodeB implementation perspective it is challenging to define stringent transmitter and receiver filter requirements to minimize the size of the needed guard band around the FDD/TDD transition frequencies. Furthermore, different regions of the world may impose local regulatory requirements to facilitate FDD/TDD deployment in adjacent bands. It was therefore

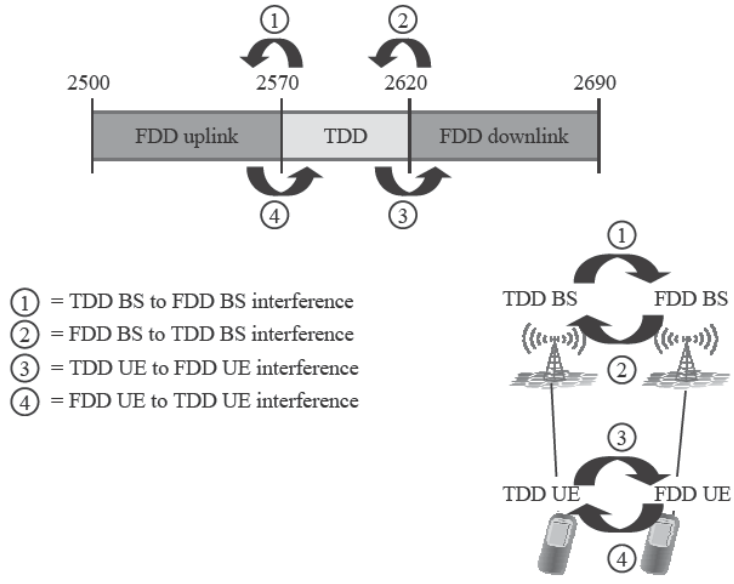


Figure 11.5 Interference scenarios between FDD and TDD (example for Band 7)

seen as impractical in 3GPP to make generic assumptions about the available guard bands and subsequently to define general FDD/TDD protection requirements, which would need to be in line with the various regulatory requirements throughout the world. Hence, additional FDD/TDD protection emission limits for the excluded frequency range of a 10 MHz frequency range immediately outside the eNodeB transmit frequency range of an operating band are left to local or regional requirements.

As an example of such local requirements, the remainder of this section will discuss the rules defined in [9] for using FDD and TDD within the European 2.6 GHz band allocation (3GPP Band 7). The cornerstones of the emission requirements in [9] are as follows:

- In-block and out-of-block Equivalent Isotropic Radiated Power (EIRP) emission masks are defined without any reference to a specific wireless system's BS standard, i.e. the limits can be considered as technology neutral. The in-block EIRP limits provide a safeguard against blocking of an adjacent system's BS receive path. The out-of-block masks, also called Block Edge Masks (BEM), impose radiated unwanted emissions limits across adjacent license block uplink/downlink receive band frequency ranges.
- The BEMs should facilitate FDD/TDD coexistence without any detailed coordination and cooperation arrangements between operators in neighboring spectrum blocks, for base station separations greater than 100 m. In this, the assumed MCL has been > 53 dB and the interference impact relative to BS thermal noise, *I/N*, has been assumed as −6 dB, with a noise figure of 5 dB. Antenna isolation measurements show that 50 dB MCL can be easily achieved on the same rooftop by using separate, vertically and/or horizontally displaced, antennas for FDD and TDD, respectively.

- Some of the BEMs have been derived from the WCDMA BS Spectrum Emission Mask (SEM), by converting the WCDMA SEM (i.e. a conducted emission requirement to be met at the BS antenna connector) into an EIRP mask by making assumptions on antenna gain and cable losses (17 dBi).
- EIRP limits are requirements for radiated (not conducted) unwanted emissions and they apply within or outside a license block. Hence, these are not testable BS equipment requirements as such, but they involve the whole RF site installation, including antennas and cable losses. Furthermore, a BEM requirement does not necessarily start at the RF channel (carrier) edge, but at the edge of the licensee's block, which may differ from the channel edge depending on the actual deployment scenario. Finally, the BEM must be met for the sum of all radiated out-of-block emissions, for example from multiple RF carriers or even multiple BSs present at a site, not only for those pertaining to a single RF channel as is the case for the UTRA SEM.

These requirements will be discussed in the following in more detail assuming the Band 7 frequency arrangements with 70 MHz FDD uplink, 50 MHz TDD and 70 MHz FDD downlink.

Figure 11.6 shows the maximum allowed in-block EIRP limits. The nominal EIRP limit of 61 dBm/5 MHz can be relaxed by local regulation up to 68 dBm/5 MHz for specific deployments, for example in areas of low population density where higher antenna gain and/or BS output power are desirable.

The lower 5 MHz part of the TDD allocation is a restricted block with a maximum allowed base station EIRP of 25 dBm/5 MHz where antennas are placed indoors or where the antenna height is below a certain height specified by the regulator. The lower transmission power in the restricted 5 MHz TDD block serves to relax the FDD BS selectivity and blocking requirements just below 2570 MHz.

Figure 11.7 shows the resulting out-of-block EIRP emission mask (BEM) for an unrestricted (high power) 20 MHz TDD block allocation starting 5 MHz above the FDD uplink band allocation. The following can be noted:

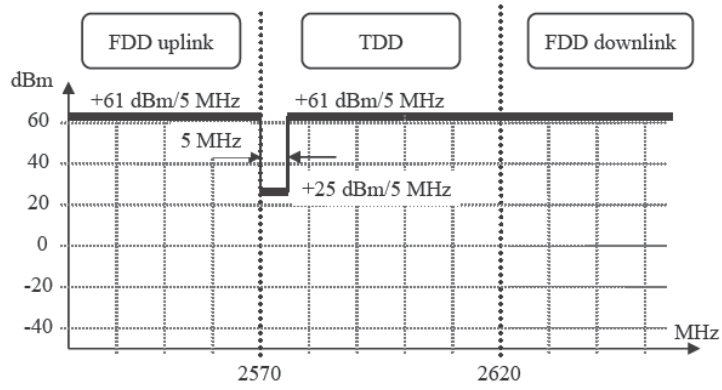


Figure 11.6 Maximum allowed in-block EIRP including restricted block

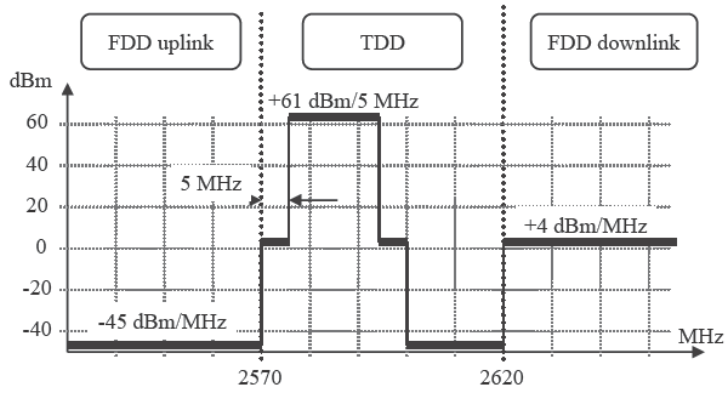


Figure 11.7 Out-of-block EIRP emission mask (BEM) for a 20 MHz TDD block above the FDD uplink band allocation

- The whole FDD uplink band allocation is protected by a −45 dBm/MHz EIRP limit, which is chosen as a baseline requirement for protection of the uplink in the BS transmitter → BS receiver interference scenario. Given a MCL > 53 dB, there will be minimal impact on the FDD BS uplink.
- Possible unsynchronized TDD operation is also protected by the −45 dBm/MHz EIRP limit from 5 MHz above the TDD block edge onwards.
- As the relevant interference mechanism across the FDD downlink band will be TDD BS → FDD UE, a more relaxed limit of +4 dBm is sufficient. This limit is also stipulated as out-of-block EIRP limit for a FDD BS across this part of the band (see Figure 11.7).
- Not shown in Figure 11.7 is the possible use of a TDD restricted block within 2570–2575 MHz, which has a tighter BEM requirement across the first 5 MHz just below the 2570 MHz transition frequency in order to protect the FDD uplink.
- Also not shown is the detailed slope of the BEM within the first MHz from either side of the TDD block edge, which follows essentially the shape of the UTRA SEM.

Figure 11.8 shows the resulting out-of-block EIRP emission mask (BEM) for 10 MHz FDD block allocation starting just above the TDD band allocation. The following can be noted:

- The first 5 MHz TDD block just below the FDD downlink block may receive a higher level of interference due to the more relaxed BEM EIRP limit of +4 dBm. This is to facilitate FDD BS transmit filters; no restricted power blocks are foreseen for the FDD blocks.
- The rest of the TDD band and FDD uplink band allocation is protected by the −45 dBm/MHz EIRP baseline requirement.

11.3.4 Transmitted Signal Quality

[2] contains the modulation-specific Error Vector Magnitude (EVM) requirements to ensure sufficiently high quality of the transmitted base station signal (see Table 11.5).

Typical impairments of the transmitted signal modulation accuracy are analog RF distortions (frequency offset, local oscillator phase noise, amplitude/phase ripple from analog fil-

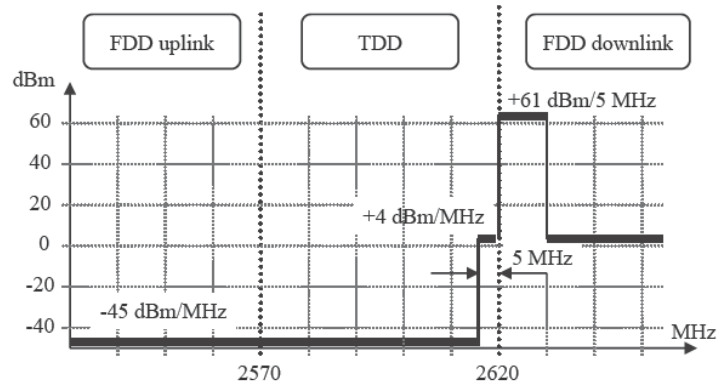


Figure 11.8 Out-of-block EIRP emission mask (BEM) for a 10 MHz FDD block above the TDD band allocation

Table 11.5 EVM requirements

Modulation scheme for PDSCH	Required EVM (%)
QPSK	17.5%
16QAM	12.5%
64QAM	8%

ters) as well as distortions created within the digital domain such as inter-symbol interference from digital filters used for spectrum shaping, finite wordlength effects and, most important for conditions near maximum transmit power, the clipping noise from peak-to-average radio reduction schemes.

The EVM requirement ensures that the downlink throughput due to the base station non-ideal waveform is only marginally reduced, typically by 5% assuming ideal reception in the UE. The required EVM must be fulfilled for all transmit configurations and across the whole dynamic range of power levels used by the base station.

11.3.4.1 Definition of the EVM

More precisely, the EVM is a measure of the difference between the ideal constellation points and the measured constellation points obtained after equalization by a defined 'reference receiver'. Unlike WCDMA, the EVM is not measured on the transmitted composite time-domain signal waveform, but within the frequency domain, after the FFT, by analyzing the modulation accuracy of the constellation points on each sub-carrier. Figure 11.9 shows the reference point for the EVM measurement.

The OFDM signal processing blocks prior to the EVM reference point, in particular the constrained Zero-Forcing (ZF) equalizer, are fundamentally not different from those of an actual UE. The rationale for this rather complex EVM definition has been that only those transmitter impairments should 'pollute' the EVM that would not be automatically removed by the UE recep-

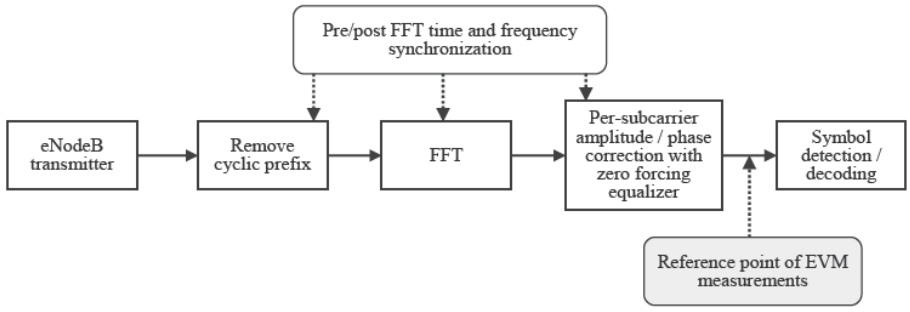


Figure 11.9 Reference point for EVM measurement

tion anyway. For example, clipping noise from peak-to-average ratio reduction has AWGN-like characteristics and cannot be removed by UE equalizers, whereas a non-linear phase response from analog filter distortion can be estimated from the reference signals and subsequently be removed by the ZF equalizer. The qualifier 'constrained' within the ZF-equalization process refers to a defined frequency averaging process of the raw channel estimates used to compute the ZF equalizer weights. Some form of frequency averaging of the raw channel estimates at the reference signal locations will be present within an actual UE receiver implementation and this will therefore limit the removal of some of the impairments such as filter amplitude (or phase) ripple with a spatial frequency in the order of the sub-carrier spacing. Therefore an equivalent of this frequency averaging process has been defined for the EVM measurement process.

Yet another feature of the EVM definition is the chosen time synchronization point for the FFT processing. The EVM is deliberately measured not at the ideal time synchronization instant, which would be approximately the centre of the cyclic prefix, but rather at two points shortly after the beginning and before the end of the cyclic prefix, respectively. This also ensures that in the presence of pre- and post-cursors from the transmit spectrum shaping filters, the UE will see sufficiently low inter-symbol interference even under non-ideal time tracking conditions.

The details of this EVM measurement procedure, including all the aspects above are defined within [5], Annex F.

11.3.4.2 Derivation of the EVM requirement

Next we provide a rationale for the EVM requirements based on an analytic approach presented in [10]. Let us calculate the required EVM for 5% throughput loss for each instantaneous C/I value, corresponding to a certain selected Modulation and Coding Scheme (MCS).

To start with, we consider the MCS throughput curves and approximating MCS envelope shown in Figure 11.10. The MCS curves were generated from link level simulations under AWGN channel conditions for a 1×1 antenna configuration (1 transmit and 1 receive antenna) using ideal channel estimation within the receiver.

We approximate the set of these throughput curves by the expression for Shannon's channel capacity with the fitting parameter α=0.65.

$$C = \alpha \log 2(1 + \frac{S}{N}) \tag{11.3}$$

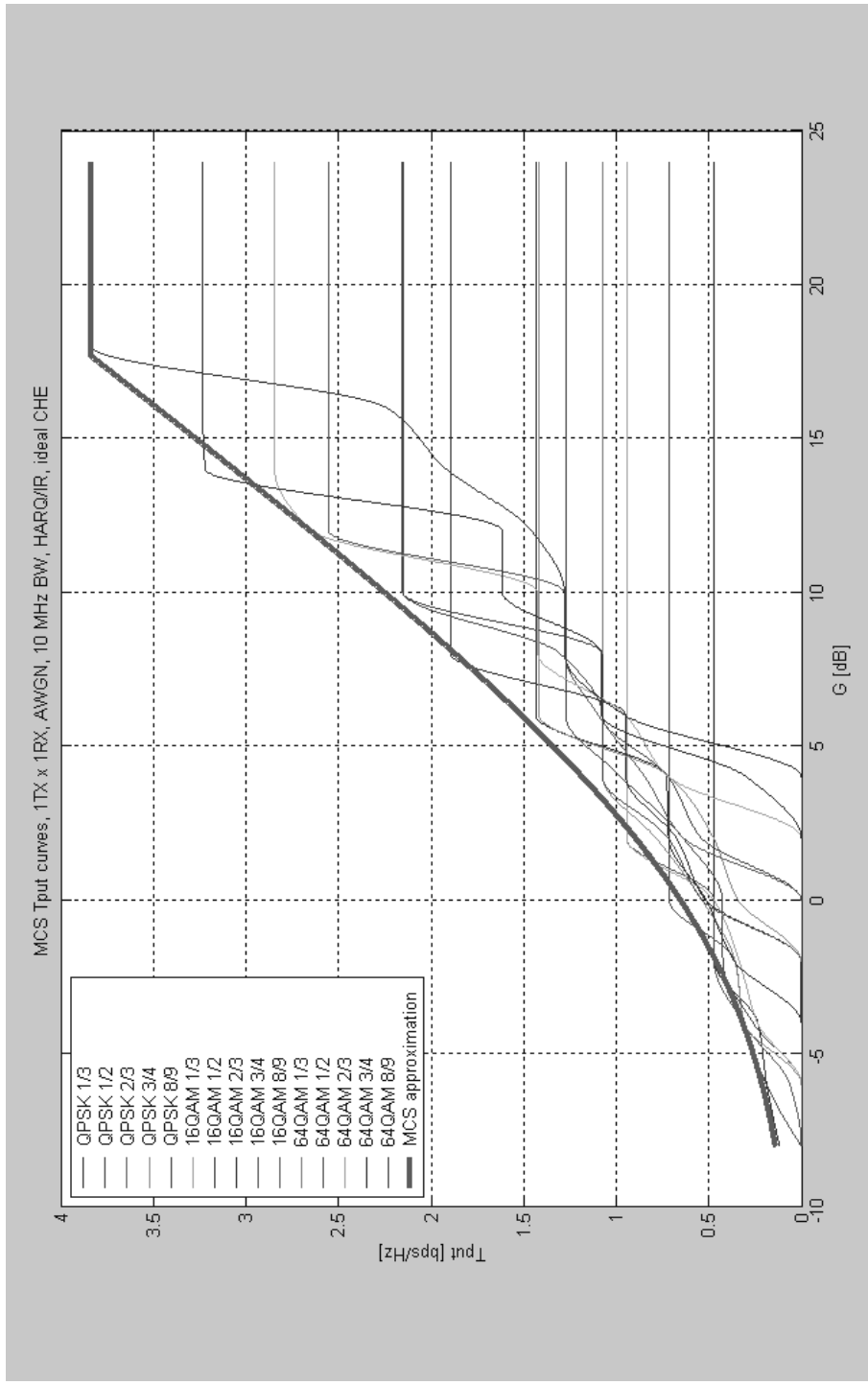


Figure 11.10 MCS throughput curves and approximating MCS envelope

where S is the signal power and N is the noise power. Assume next that the transmit impairments can be modeled as AWGN with power M and the receiver noise as AWGN with power N. Then the condition that the transmitter AWGN should lead to 5% throughput loss can be expressed as:

$$\alpha \log 2(1+\frac{S}{M+N}) = 0.95 \cdot \alpha \log 2(1+\frac{S}{N}) \tag{11.4}$$

Solving this condition for $EVM_{req} \equiv \frac{M}{S}$ we obtain

$$EVM_{req} \equiv \frac{M}{S} = \left[\left(1+\frac{S}{N}\right)^{0.95} - 1\right]^{-1} - \frac{N}{S} \tag{11.5}$$

The required EVM is plotted in Figure 11.11 for the C/I range of the approximating envelope MCS throughput curve of Figure 11.10. As can been seen ~6.3% EVM would be required for the highest throughput MCS (64QAM 8/9, operating S/N ~17.7 dB). However, when assuming a single EVM requirement for all 64QAM modulated MCSs, this would be too stringent as 64QAM may be chosen from a C/I range from 12 to 17.7 dB according to the chosen MCS set. This would indicate a required EVM in the range of 10–6.3%. Looking at the midpoint S/N of ~15 dB for 64QAM MCS selection one obtains a 7.9% EVM requirement. Similarly, one obtains for:

- 16QAM: C/I range from 6 to 12 dB, midpoint C/I of ~9 dB with an EVM requirement of 12.9%;
- QPSK: C/I range from −8 to 6 dB, midpoint C/I of ~ −1 dB with EVM requirements of 29.6% and 16.3% respectively (for 6 dB C/I).

The required EVM values for 64QAM, 16QAM and QPSK are 8%, 12.5% and 17.5%.

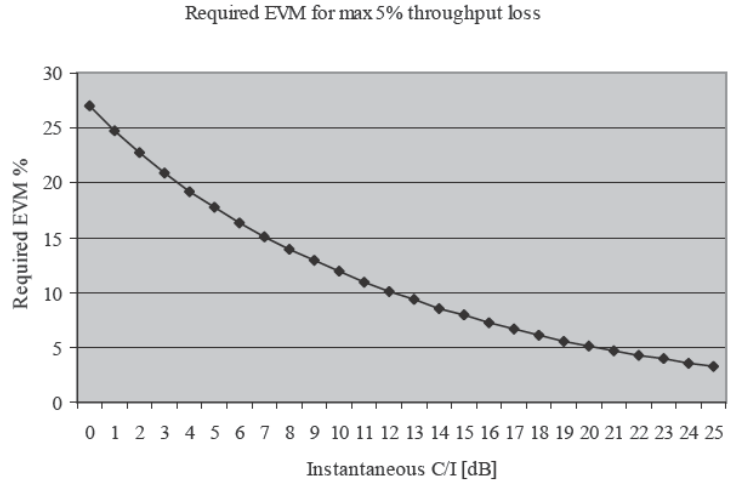


Figure 11.11 Required EVM for 5% throughput loss

However, the above derivation assumes a smooth approximating MCS envelope. In reality, the MCS envelope has a 'waterfall' shape in AWGN, as shown in Figure 11.10 and may exhibit either a larger impact from EVM in regions with steep slopes or a smaller impact in regions with flat response. In an actual system scenario the C/I distribution will average these unequal throughput losses across the 'waterfall' MCS envelope. The quasi-static system simulations in [10] do indeed verify that the resulting average throughput loss for 64QAM is in line with the above simplified derivation.

11.4 eNodeB RF Receiver

The purpose of the eNodeB RF receiver requirements is to verify different RF impairments that have an impact on the network performance. These impairments include noise figure, receiver EVM, selectivity on different frequencies, including adjacent channel, etc. The following base station RF receiver requirements are described: reference sensitivity level, dynamic range, in-channel selectivity, Adjacent Channel Selectivity (ACS), blocking, receiver spurious emissions, and receiver intermodulation.

11.4.1 Reference Sensitivity Level

The reference sensitivity level is the minimum mean power received at the antenna connector, at which a throughput requirement is fulfilled. The throughput will be equal to or higher than 95% of the maximum throughput for a specified reference measurement channel.

The purpose of this requirement is to verify the receiver noise figure. Other receiver impairments such as receiver EVM are included within the receiver demodulation performance requirements at high SNR points. Therefore, the maximum throughput is defined at low SNR points for the sensitivity case. The reference measurement channel is based on a QPSK modulation and 1/3 coding rate.

For channel bandwidths lower than or equal to 5 MHz, the reference measurement channels are defined on the basis of all resource blocks allocated to this channel bandwidth. For channel bandwidths higher than 5 MHz, the sensitivity is measured using consecutive blocks consisting of 25 RBs.

The reference sensitivity level calculation is given by Equation 11.6. Noise Figure (NF) is equal to 5 dB and implementation margin (IM) is equal to 2 dB.

$$P_{REFSENS}[\text{dBm}] = -174[\text{dBm/Hz}] + 10\log_{10}(N_{RB} \cdot 180k) + NF + SNR + IM \qquad (11.6)$$

For example, for a 10 MHz channel bandwidth ($N_{RB} = 25$), the reference sensitivity level is equal to −101.5 dBm. The simulated SNR for 95% of the maximum throughput is equal to −1.0 dB.

The eNodeB noise figure is relevant for the coverage area. The uplink link budget calculation in Chapter 9 assumes a base station noise figure of 2 dB, so clearly better than the minimum performance requirement defined in 3GPP specifications. The typical eNodeB has better performance than the minimum requirement since the optimized performance is one of the selling arguments for the network vendors.

11.4.2 Dynamic Range

The dynamic range requirement is a measure of the capability of the receiver to receive a wanted signal in the presence of an interfering signal inside the received frequency channel, at which a throughput requirement is fulfilled. The throughput will be equal to or higher than 95% of the maximum throughput for a specified reference measurement channel.

The intention of this requirement is to ensure that the base station can receive a high throughput in the presence of increased interference and high wanted signal levels. Such a high interference may come from neighboring cells in the case of small cells and high system loading. This requirement measures the effects of receiver impairments such as receiver EVM and is performed at high SNR points. The mean power of the interfering signal (AWGN) is equal to the receiver noise floor increased by 20 dB, in order to mask the receiver's own noise floor. The maximum throughput is defined at high SNR points, thus the reference measurement channel is based on 16QAM modulation and 2/3 coding rate.

The mean power calculation of the wanted signal is given by Equation 11.7. NF is equal to 5 dB and IM is equal to 2.5 dB.

$$P_{wanted}[\text{dBm}] = -174[\text{dBm/Hz}] + 10\log(N_{RB} \cdot 180\text{k}) + 20 + \text{NF} + \text{SNR} + \text{IM} \qquad (11.7)$$

For example, for 10 MHz channel bandwidth ($N_{RB} = 25$), the wanted signal mean power is equal to −70.2 dBm. The simulated SNR for 95% of the maximum throughput is equal to 9.8 dB. The interfering signal mean power is set equal to −79.5 dBm. The dynamic range measurement for the 5 MHz case is illustrated in Figure 11.12.

11.4.3 In-channel Selectivity

The in-channel selectivity requirement is a measure of the receiver's ability to receive a wanted signal at its assigned resource block locations in the presence of an interfering signal received at a larger power spectral density, at which a throughput requirement is fulfilled. The throughput will equal or exceed 95% of the maximum throughput for a specified reference measurement channel.

The intention of this requirement is to address in-band adjacent resource block selectivity, i.e. the reception of simultaneous user signals at greatly different power spectral density levels due to used modulation format, power control inaccuracies, other-cell interference levels, etc. The uplink signal is created by two signals, where one is the wanted QPSK modulated signal

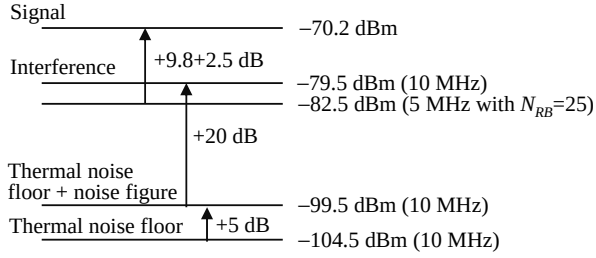


Figure 11.12 Dynamic range measurement for 10 MHz LTE

Table 11.6 Number of resource blocks allocated for wanted and interfering signal

Channel bandwidth (MHz)	Wanted signal	Interfering signal
1.4	3	3
3	9	6
5	15	10
10	25	25
15	25	25
20	25	25

and the other is the interfering 16QAM modulated signal at elevated power. Table 11.6 presents the allocation of resource blocks for wanted and interfering signals, for different channel bandwidths. The high power level difference may happen if the interfering user is close to the base station and can use a high signal-to-noise ratio while the wanted user is far from the base station and can only achieve a low signal-to-noise ratio.

For channel bandwidths equal to 10, 15 and 20 MHz, the 25 resource block allocations of the wanted and interfering signals are adjacently around Direct Current (DC), in order to be sensitive to the RF impairments of the receiver image leakage, EVM, 3rd order intermodulation (IMD3) and the local oscillator phase noise.

The mean power of the interfering signal is equal to the receiver noise floor increased by 9.5 dB (required SNR) and additionally by 16 dB (assumed interference over thermal noise). The desensitization of the wanted resource block allocation, in the presence of the interfering resource block allocation, is equal to 3 dB.

The wanted signal mean power calculation is given by Equation 11.8. NF is equal to 5 dB and IM is equal to 2 dB.

$$P_{wanted}[\text{dBm}] = -174[\text{dBm/Hz}] + 10\log(N_{RB} \cdot 180\text{k}) + 3 + \text{NF} + \text{SNR} + \text{IM} \qquad (11.8)$$

For example, for 10 MHz channel bandwidth ($N_{RB} = 25$), the wanted signal mean power is equal to –98.5 dBm. The simulated SNR for 95% of the maximum throughput is equal to −1.0 dB. The interfering signal mean power is equal to −77 dBm. The in-channel measurement case is presented in Figure 11.13.

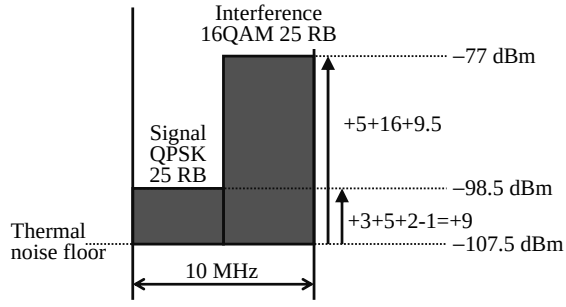


Figure 11.13 In-channel selectivity measurement for 5 MHz LTE

11.4.4 Adjacent Channel Selectivity (ACS) and Narrow-band Blocking

The ACS (narrow-band blocking) requirement is a measure of the receiver's ability to receive a wanted signal at its assigned channel frequency in the presence of an interfering adjacent channel signal, at which a throughput requirement is fulfilled. The throughput will equal or exceed 95% of the maximum throughput, for a specified reference measurement channel.

The intention of this requirement is to verify the selectivity on the adjacent channel. The selectivity and the narrowband blocking are important to avoid interference between operators. The adjacent operator's mobile may use a high transmission power level if its own base station is far away. If such a mobile happens to be close to our base station, it can cause high interference levels on the adjacent channel.

Table 11.7 presents the ACS requirement relationship between E-UTRA interfering signal channel bandwidth, wanted signal channel bandwidth, wanted signal mean power and interfering signal centre frequency offset to the channel edge of the wanted signal. The interfering signal mean power is equal to −52 dBm. Table 11.8 shows how the adjacent channel selectivity can be calculated from the performance requirements for 10 MHz channel bandwidth with 25 resource blocks. The adjacent channel selectivity test case is illustrated in Figure 11.14. Base station noise floor is given by Equation 11.9. NF is equal to 5 dB and $N_{RB} = 25$.

$$D[\text{dBm}] = -174[\text{dBm/Hz}] + 10\log(N_{RB} \cdot 180\text{k}) + \text{NF} \tag{11.9}$$

The narrow-band blocking measurement is illustrated in Figure 11.15. The wanted signal mean power is equal to $P_{REFSENS} + 6$ dB. The interfering 1RB E-UTRA signal mean power is equal to −49 dBm. The interfering signal is located in the first five worst case resource block allocations. Additionally, for 3 MHz channel bandwidth, the interfering signal is located in every third resource block allocation. For 5, 10, 15 and 20 MHz channel bandwidths, the interfering signal is located additionally in every fifth resource block allocation. Such a location of the interfering signal verifies different possible impairments of the receiver performance. For GSM

Table 11.7 Derivation of ACS requirement

Wanted signal channel bandwidth [MHz]	Wanted signal mean power [dBm]	Offset [MHz]	Interfering signal channel bandwidth [MHz]
1.4	$P_{REFSENS} + 11$	0.7	1.4
3	$P_{REFSENS} + 8$	1.5	3
5, 10, 15, 20	$P_{REFSENS} + 6$	2.5	5

Table 11.8 Calculation of the ACS requirement for 10 MHz channel bandwidth

Interfering signal mean power (A)	–52 dBm
Base station noise floor with 5 dB noise figure (B)	–102.5 dBm
Allowed desensitization (C)	6 dB
Total base station noise (D = B + C in dB)	–96.5 dBm
Allowed base station interference (E = D – B in absolute value)	–98 dBm
Base station adjacent channel selectivity (F = A – E)	46 dB

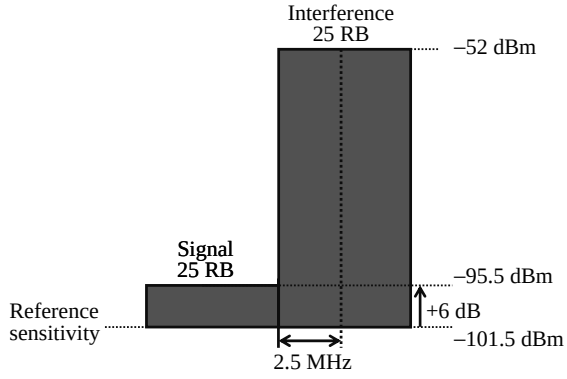


Figure 11.14 Adjacent channel selectivity measurement for 10 MHz LTE

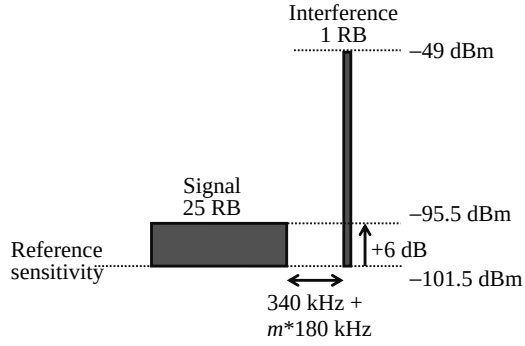


Figure 11.15 Narrowband blocking measurement for 10 MHz LTE

band refarming, the blocker can be a narrowband GSM signal. There are no specific requirements with a GSM signal used as the blocker. The GSM signal, however, is sufficiently similar to 1 RB LTE blocker for practical performance purposes.

11.4.5 Blocking

The blocking requirement is a measure of the receiver's ability to receive a wanted signal at its assigned channel frequency in the presence of an unwanted interferer, at which a throughput requirement is fulfilled. The throughput will equal or exceed 95% of the maximum throughput for a specified reference measurement channel.

The intention of this requirement is to verify the selectivity on different frequencies, excluding the adjacent channel. The in-band blocking can also be called adjacent channel selectivity for the second adjacent channel.

For in-band blocking the unwanted interferer is an LTE signal. For example, for Operating Band 1 (1920–1980 MHz), the in-band blocking refers to the centre frequencies of the interfering signal from 1900 MHz to 2000 MHz. For out-of-band blocking, the unwanted interferer

is a Continuous Wave (CW) signal. For Operating Band 1, the out-of-band blocking refers to the centre frequencies of the interfering signal from 1 MHz to 1900 MHz and from 2000 MHz to 12 750 MHz. The measurements are shown in Figure 11.16.

The in-band blocking requirement for 10 MHz LTE is illustrated in Figure 11.17**,** and out-of-band in Figure 11.18. The wanted signal mean power is equal to $P_{REFSENS}$ + 6 dB. The E-UTRA interfering signal mean power is equal to −43 dBm. The CW interfering signal mean power is equal to −15 dBm.

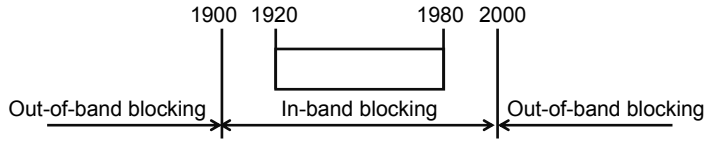


Figure 11.16 In-band and out-of-band blocking measurement for Band 1

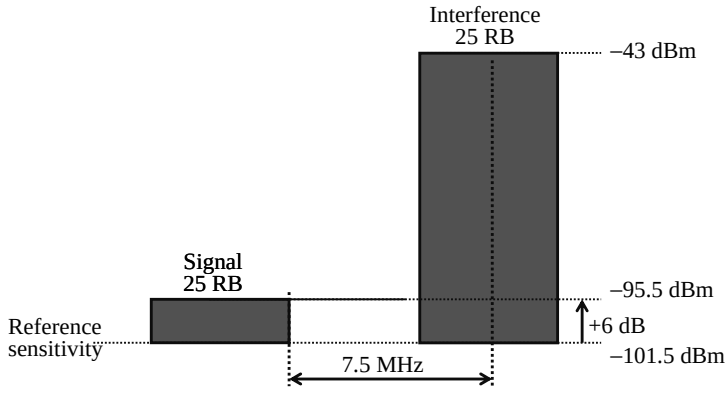


Figure 11.17 In-band blocking measurement for 10 MHz LTE

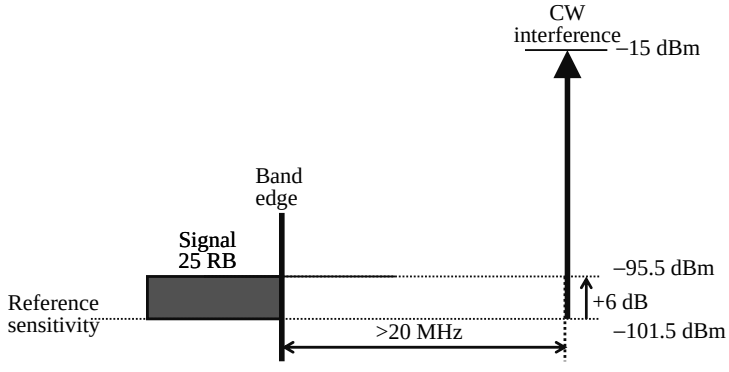


Figure 11.18 Out-of-band blocking measurement for 10 MHz LTE

11.4.6 Receiver Spurious Emissions

The spurious emissions power is the power of emissions generated or amplified in a receiver that appears at the base station receiver antenna connector.

The frequency range between 2.5 × channel bandwidth below the first carrier frequency and 2.5 × channel bandwidth above the last carrier frequency, transmitted by the base station, may be excluded from the requirement. Frequencies that are more than 10 MHz below the lowest frequency or more than 10 MHz above the highest frequency of the base station transmitter operating band will not be excluded from the requirement.

Additionally, the power of any spurious emission will not exceed the levels specified for coexistence with other systems in the same geographical area and for protection of the LTE FDD base station receiver of own or different base station.

11.4.7 Receiver Intermodulation

Intermodulation response rejection is a measure of the capability of the receiver to receive a wanted signal on its assigned channel frequency in the presence of two interfering signals which have a specific frequency relationship to the wanted signal, at which a throughput requirement is fulfilled. The throughput will equal or exceed 95% of the maximum throughput for a specified reference measurement channel.

The intermodulation requirement is relevant when there are two terminals from other operators transmitting at high power level close to our base station.

The mean power of the wanted signal is equal to $P_{REFSENS} + 6$ dB (Figure 11.19). The interfering signal mean power (both LTE and CW) is equal to −52 dBm. The offset between the CW interfering signal centre frequency and the channel edge of the wanted signal is equal to 1.5 LTE interfering signal channel bandwidth. The offset between the LTE interfering signal centre frequency and the channel edge of the wanted signal is specified on the basis of the worst case scenario – the intermodulation products fall on the edge resource blocks of an operating channel bandwidth.

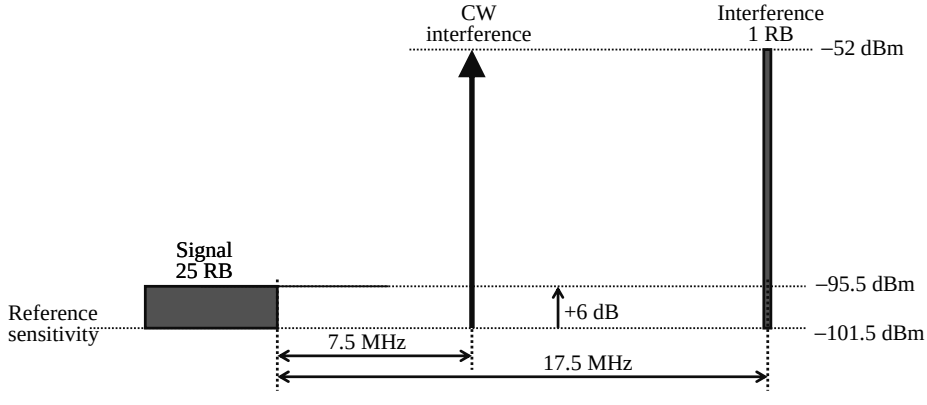


Figure 11.19 Intermodulation measurement for 5 MHz LTE

11.5 eNodeB Demodulation Performance

The purpose of the base station demodulation performance requirements is to estimate how the network is performing in practice and to verify different base station impairments which have an impact on the network performance. These impairments include RF and baseband impairments, receiver EVM, time and frequency tracking, frequency estimation, etc. The base station demodulation performance requirements are described for the following uplink channels: Physical Uplink Shared Channel (PUSCH), Physical Uplink Control Channel (PUCCH) and Physical Random Access Channel (PRACH).

11.5.1 PUSCH

The PUSCH demodulation performance requirements are determined by a minimum required throughput for a given SNR. The throughput will equal or exceed 30% or 70% of the maximum throughput for a specified reference measurement channel that contains data and reference signals only.

The PUSCH demodulation performance requirements are specified for all LTE channel bandwidths. Additionally, for each channel bandwidth, the following various network related parameters are selected to match different radio system configurations:

- number of receive antennas – 2 or 4;
- modulation and coding scheme – QPSK 1/3, 16QAM 3/4 or 64QAM 5/6;
- channel model – EPA5, EVA5, EVA70, ETU70 or ETU300, where the number indicates the Doppler shift. PA is Pedestrian A, VA is Vehicular A and TU is Typical Urban channel model;
- cyclic prefix type – normal or extended;
- number of resource blocks allocated for channel bandwidth – single or all possible.

Each channel model is described by the Doppler frequency which corresponds to various velocities depending on the frequency band. For example, EPA5 corresponds to velocities equal to 7.7 kmph, 2.7 kmph and 2.1 kmph for 0.7 GHz (Band 12), 2 GHz (Band 1) and 2.6 GHz (Band 7), respectively.

The incremental redundancy HARQ allowing up to three retransmissions is used. Table 11.9 presents a set of PUSCH base station tests for each channel bandwidth and for each configuration of receive antennas.

For a base station supporting multiple channel bandwidths, only tests for the largest and the smallest channel bandwidths are applicable.

The SNR requirements were specified on the basis of average link level simulation results with implementation margins presented by various companies during 3GPP meetings.

An example of the relationship between the net data rate and required SNR follows: 10 MHz channel bandwidth, normal cyclic prefix and QSPK 1/3 modulation and coding scheme were taken into account.

For a fixed reference channel A3–5 [2], the SNR requirements were specified for full resource block allocation (50 resource blocks), for 2 and 4 receive antennas, for EPA5 and EVA70 channel models and for 30% and 70% fractions of the maximum throughput. Table 11.10 presents the SNR requirements for a 30% fraction of the maximum throughput.

Table 11.9 Set of PUSCH base station tests

Cyclic prefix	Channel model, RB allocation	Modulation and coding scheme	Fraction of maximum throughput [%]
Normal	EPA5, all	QPSK 1/3 16QAM 3/4 64QAM 5/6	30 and 70 70 70
	EVA5, single	QPSK 1/3 16QAM 3/4 64QAM 5/6	30 and 70 30 and 70 70
	EVA70, all	QPSK 1/3 16QAM 3/4	30 and 70 30 and 70
	ETU70, single	QPSK 1/3	30 and 70
	ETU300, single	QPSK 1/3	30 and 70
Extended	ETU70, single	16QAM 3/4	30 and 70

Table 11.10 SNR requirements for 30% fraction of the maximum throughput (fixed reference channel A3–5)

Number of receive antennas	Channel model	SNR requirement [dB]
2	EPA5 EVA70	–4.2 –4.1
4	EPA5 EVA70	–6.8 –6.7

For the fixed reference channel A3–5 the payload size is equal to 5160 bits and corresponds to a 5.16 Mbps instantaneous net data rate. For one resource block it is 103.2 kbps, accordingly.

A 30% fraction of the maximum throughput corresponds to 30.9 kbps and 61.9 kbps for one and two resource blocks, respectively.

The link budget in Chapter 9 assumed an SNR equal to −7 dB with 64 kbps and two resource blocks corresponding to 32 kbps and one resource block. The link budget assumes better eNodeB receiver performance because of more assumed HARQ retransmissions and also because typical eNodeB performance is closer to the theoretical limits than the 3GPP minimum requirement.

Additionally, the uplink timing adjustment requirement for PUSCH was specified in [2]. The rationale for this requirement is to check if the base station sends timing advance commands with the correct frequency and if the base station estimates appropriate uplink transmission timing. The uplink timing adjustment requirements are determined by a minimum required throughput for a given SNR and are specified for moving propagation conditions as shown in Figure 11.20. The time difference between the reference timing and the first tap is described by Equation 11.10, where $A = 10$ μs.

$$\Delta\tau = \frac{A}{2} \cdot \sin(\Delta\omega \cdot t) \tag{11.10}$$

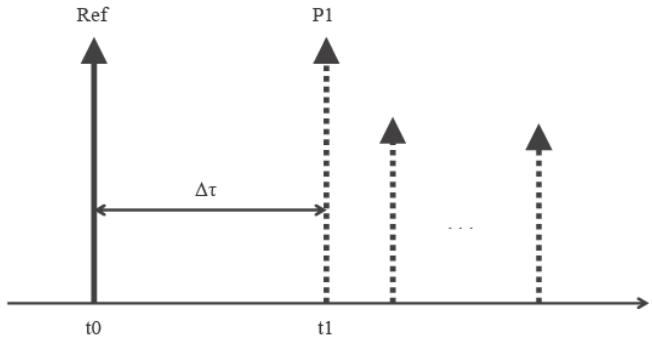


Figure 11.20 Moving propagation conditions

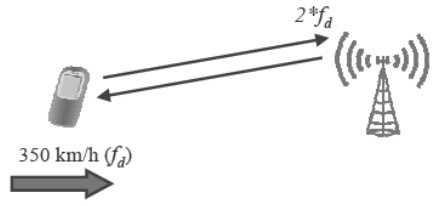


Figure 11.21 High speed train demodulation requirement

The uplink timing adjustment requirement is specified for normal and extreme conditions. For normal conditions the ETU channel model and UE speed of 120 km/h is considered ($\Delta\omega = 0.04$ s^{-1}). Uplink timing adjustment in extreme conditions is an optional requirement and corresponds to AWGN channel model and UE speed of 350 km/h ($\Delta\omega = 0.13$ s^{-1}).

[2] also includes eNodeB decoding requirements for the high speed train conditions up to 350 km/h. eNodeB can experience a two times higher Doppler shift in the worst case if the UE synchronizes to the downlink frequency including the Doppler shift (f_d) (see Figure 11.21). The maximum Doppler shift requirement is 1340 Hz which corresponds to 350 km/h at 2.1 GHz assuming that eNodeB experiences double Doppler shift.

11.5.2 PUCCH

The PUCCH performance requirements are specified for PUCCH format 1a and for PUCCH format 2. The PUCCH format 1a performance requirements are determined by a minimum required DTX to ACK probability and ACK missed detection probability for a given SNR. The DTX to ACK probability, i.e. the probability that ACK is detected when nothing is sent, will not exceed 1%. The ACK missed detection probability, i.e. the probability that ACK is not detected properly, will not exceed 1%.

The PUCCH format 1a performance requirements are specified for all E-UTRA channel bandwidths. Additionally, for each channel bandwidth, the following various network related parameters were selected to match different radio system configurations:

Table 11.11 Set of PUCCH base station tests

Cyclic prefix	Channel model
Normal	EPA5 EVA5 EVA70 ETU300
Extended	ETU70

- number of receive antennas – 2 or 4;
- channel model – EPA5, EVA5, EVA70, ETU70 or ETU300;
- cyclic prefix type – normal or extended.

Table 11.11 presents a set of PUCCH base station tests for each channel bandwidth and for each configuration of receive antennas.

The PUCCH format 2 performance requirements are determined by a CQI missed detection BLER probability for a given SNR. The CQI missed detection probability, i.e. the probability that CQI is not detected properly, will not exceed 1%. The PUCCH format 2 performance requirements are specified for all E-UTRA channel bandwidths, normal cyclic prefix, 2 receive antennas and ETU70 channel model only.

For base station supporting multiple channel bandwidths, only tests for the greatest and the smallest channel bandwidths are applicable.

11.5.3 PRACH

The PRACH performance requirements are specified for burst format 0, 1, 2, 3 and are determined by a minimum required total false alarm probability and missed detection probability for a given SNR. The total false alarm probability, i.e. the probability that preamble is detected (the sum of all errors from all detectors) when nothing is sent, will not exceed 0.1%. The missed detection probability will not exceed 1% and depends on the following errors:

- preamble is not detected properly;
- different preamble is detected than the one that is sent;
- correct preamble is detected but with wrong timing estimation.

The PRACH performance requirements are specified for 2 and 4 receive antennas and for AWGN and ETU70 channel models. For AWGN and ETU70, the timing estimation error occurs if the estimation error of the timing of the strongest path is larger than 1.04 μs and 2.08 μs, respectively. For ETU70, the strongest path for the timing estimation error refers to the strongest path in the power delay profile, i.e. the average of the delay of all paths having the same highest gain equal to 310 ns.

The PRACH performance requirements are specified for normal mode and for high speed mode. Different frequency offsets are tested for these modes. For high speed mode, when the receiver is in demanding conditions, additional frequency offsets are specified – 625 Hz

Table 11.12 Set of PRACH base station tests for normal mode

Channel model	f offset [Hz]
AWGN	0
ETU70	270

Table 11.13 Set of PRACH base station tests for high speed mode

Channel model	f offset [Hz]
AWGN	0
ETU70	270
ETU70	625
ETU70	1340

and 1340 Hz. For a frequency offset of 625 Hz, which corresponds to half of the preamble length (0.5·1/0.8 ms), the receiver is in difficult conditions because several correlation peaks are observed; a 1340 Hz frequency offset corresponds to a velocity of 350 km/h at 2.1 GHz frequency band.

Table 11.12 and Table 11.13 present a set of PRACH base station tests for each burst format and each configuration of receive antennas, for normal mode and for high speed mode, respectively.

11.6 UE Design Principles and Challenges

11.6.1 Introduction

The main requirements related to the LTE UE design are described in this section. As all LTE UEs will have to support legacy air interfaces, the LTE functionality will be constructed on top of the 2G GSM/EDGE and 3.5G WCDMA/HSPA architecture. This section presents the main differences and new challenges of an LTE terminal compared to a WCDMA terminal. Both data card and phone design are discussed.

11.6.2 RF Subsystem Design Challenges

11.6.2.1 Multi-mode and Multi-band Support

LTE equipment has to provide connection to legacy air interfaces to offer customers roaming capability in areas where LTE base stations are not yet deployed. It is essential for the acceptance of the new technology that there is continuity in the service to the user. The equipment must also support different operator, regional and roaming requirements, which results in the need to support many RF bands. For the successful introduction of a new technology the performance

of the UE must be competitive with existing technology in terms of key criteria such as cost, size and power consumption [11].

The definition of 3GPP bands can be found in Table 11.1. Although the Phase Locked Loop (PLL) and RF blocks of a Transceiver (TRX) can be designed to cover almost all the bands, the designer still has to decide how many bands will be supported simultaneously in a given phone to optimize the radio. This drives the number and frequency range of Low Noise Amplifiers (LNAs) and transmit buffers. The same considerations exist for the Front-End (FE) components in terms of the number and combination of Power Amplifiers (PAs), filters and thus the number of antenna switch ports. Similarly, the number of supported bands required in diversity path needs to be decided.

The support of legacy standards in combination with band support results in a complex number of use cases. These need to be studied to provide an optimum solution in terms of size and cost. Here are some of the anticipated multi-mode combinations:

- EGPRS + WCDMA + LTE FDD
- EGPRS + TD-SCDMA + LTE TDD
- EVDO + LTE FDD.

The first two combinations are natural migration paths through 3GPP cellular technologies and standardization has taken care of measurement mechanisms to allow handover back and forth between each technology without requiring operation (receive or transmit) in two modes simultaneously. This allows all three modes to be covered with a single receive or transmit path, or two receive paths when diversity is required. In the latter, handover support is more complex but again a transmit and receive path combination is feasible since two receivers are available from the LTE diversity requirement.

These multi-mode requirements can be supported by one highly reconfigurable TRX Integrated Circuit (IC), which is often better known under the well-used term of 'Software Defined Radios'. In reality, this should be understood as software reconfigurable hardware. This re-configurability for multi-mode operation takes place mainly in the analogue baseband section of the receiver and the transmitter. The multi-band operation takes place in the RF and Local Oscillator (LO) section of the TRX and in the RF-FE.

The combination of the high number of bands to be supported together with the need for multi-mode support is driving RF sub-system architectures that optimize hardware reuse, especially in the FE where the size and number of components becomes an issue. Improvement in this area builds on the optimizations already realized in EGPRS/WCDMA terminals, driving them further to meet LTE functionality. This means that not only does the LTE functionality need to comply with the architecture framework used for 2G and 3G but it also explores new opportunities of hardware reuse:

- LTE performance should be achieved without the use of external filters: neither between LNA and mixer nor between the transmitter and the PA as this is already realized in some WCDMA designs. This not only removes two filters per band but also allows simplification of the design of a multi-band TRX IC. This is especially critical for the FDD mode and where large channel bandwidth is used in bands where the duplex spacing is small. These new design tradeoffs are discussed in section 11.8.2. Similarly the interface to the Baseband (BB*)* needs to multiplex every mode to a minimum number of wires; this is best achieved by a digital interface as described in section 11.6.3.

- Reuse of the same RF FE path for any given band irrespective of its operation mode. This implies using:
  - Co-banding: reuse of the same receive filter for any mode, especially EGPRS (half duplex) reuses the duplexer required for FDD modes.
  - Multi-mode PA: reuse of same PA whatever mode or band.

The details associated with these two techniques are developed in further sections. However it is instrumental to illustrate their benefits in terms of reduced complexity and thus size and cost. Table 11.14 shows the difference in complexity between a design with none or all of the above optimizations. The example is that of a US phone with international roaming supporting: quad-band EGPRS (bands 2/3/5/8), quad-band WCDMA (bands 1/2/4/5), and triple band LTE (bands 4/7/13). The fully optimized solution block diagram is illustrated in Figure 11.22.

The two scenarios show a difference of almost a factor of two in the number of components. In addition the lower component count also significantly simplifies the TRX IC and the FE layout and size. A number of partially optimized solutions also exist, however, that fall somewhere between these scenarios.

11.6.2.2 Antenna Diversity Requirement

One of the main features introduced with LTE is MIMO operation to enhance the available data rate. MIMO operation requires the UE to be equipped with two receive antennas and paths. The conformance testing is done through RF connectors and assumes totally uncorrelated antennas. This scenario is far from representative of real operation in the field especially for small terminals operating in the lower 700 MHz frequency band. Small terminals such as smart phones have dimensions that only allow a few centimeters of separation between the two antennas. At low frequencies this distance results in a high correlation between the two signals received at each antenna, which results in degraded diversity gain. Furthermore at these frequencies and with small terminal sizes, the hand effect (modification of antenna's radiation patterns due to the hand or head proximity) further degrades the diversity gain. For devices demanding higher data rates, such as laptops or PC tablets, the device sizes allow proper antenna design. Also in higher frequency bands even a small terminal can provide sufficient antenna separation to grant good MIMO operation.

Table 11.14 Component count for different front end design

Blocks	Implementation	
	Un-optimized 'side-by-side'	Fully optimized
Low Noise Amplifiers	13	10
Duplex filters	6	5.5*
Band-pass filters	17	5
Power Amplifiers	8	2
Switches	2	4
Transceivers	3	1
Total number of components	49	27

*Reuse of RX side for band 1 and 4.

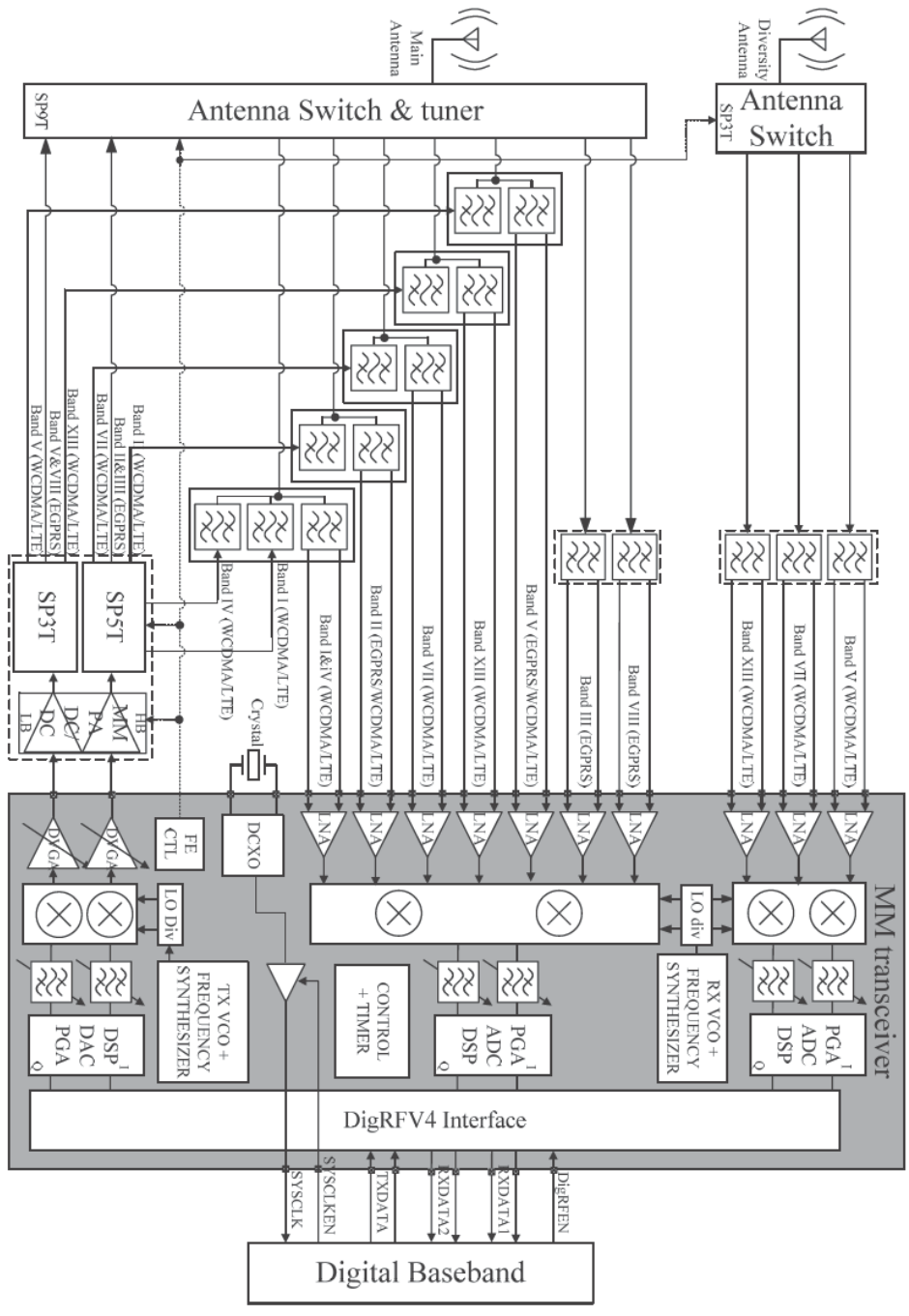


Figure 11.22 EGPRS/WCDMA/LTE optimized RF subsystem block diagram. HB, high band; LB, low band; DCXO, digital crystal oscillator; PGA, programmable gain amplifier; MM, multi-mode; FE, front-end

When the antenna design of an LTE UE is considered relative to current WCDMA phones the following design challenges are encountered:

1 The overall band support: the current WCDMA antenna frequency range is from 824 to 2170 MHz whereas future LTE devices will have to cover 698 to 2690 MHz. This stretches the current state of the art for antenna matching and also for maintaining antenna gain throughout the larger bandwidth. This will probably drive the introduction of new technology such as active antenna tuning modules.
2 Some LTE bands create new antenna coupling issues with other systems potentially present in the UE. The other antennas to be considered are the Global Positioning System (GPS) antenna, the Bluetooth (BT) and WLAN antenna, the analog Frequency Modulated (FM) radio antenna, and the Digital TV (Digital Video Broadcast – Handheld, DVB-H) antenna. The related critical coexistence use cases are discussed in section 11.6.2.3.
3 The support of antenna diversity: introducing an extra antenna in an already complex and relatively small UE presents a significant challenge if reasonable diversity gain is required.

The last two issues are easier to handle in data card designs where only the cellular modem is present. To some extent this is also true for the larger portable devices such as PC tablets and video viewers but smart phone mechanical design may prove particularly challenging for antenna placement.

11.6.2.3 New RF Coexistence Challenges

In the context of the multi-mode UE where multiple radio systems and multiple modems, such as BT, FM radio, GPS, WLAN and DVB-H must coexist, the larger bandwidth, the new modulation scheme and the new bands introduced in LTE create new coexistence challenges. In general coexistence issues are due to the transmit (TX) signal of one system (aggressor) negatively impacting another system's receiver ('RX'-victim) performance and notably its sensitivity. There are two aspects to consider: the direct rise of the victim's noise floor by the aggressor's transmitter out-of-band noise in the receiver band, and degradation of the receiver's performance due to blocking mechanisms.

Noise leakage from an aggressor TX in a victim's RX band is added to the RX noise floor further degrading its sensitivity, as shown in Figure 11.23. This noise leakage is only a function of the intrinsic TX noise, TX output filter attenuation in the victim's band and antenna isolation. This desensitization mechanism depends on the aggressor's TX design.

The desensitization of a victim due to an aggressor's transmitter out-of-band noise can be calculated as follows:

$$DES_{OOBN} = 10 * \mathrm{Log}(10^{(-174+NFvictim)} + 10^{(POUTaggressor-OOBN-ANTisol)}) + 174 - NFvictim \quad (11.11)$$

Where DES_{OOBN} is the resulting degradation of the victim's sensitivity in dB due to aggressor's TX noise, *NFvictim* is the victim's RX Noise Figure (NF) referred to the antenna in dB, *POUTaggressor* is the aggressor's TX maximum output power in dBm, *Out-of-band Noise (OOBN)* is the aggressor's TX noise in the victim's band in dBc/Hz, and *ANTisol* is the antenna isolation in dB.

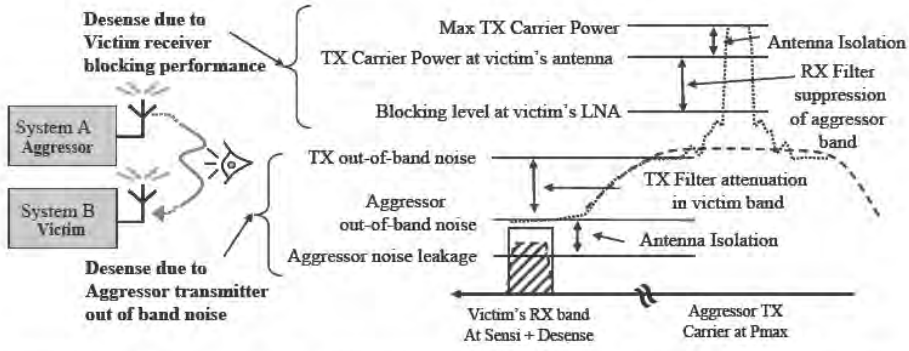


Figure 11.23 Victim/aggressor block diagram and aggressor transmitter leakage to victim receiver

The blocker power level present at the victim's LNA input depends on the aggressor maximum transmitted signal power, the antenna isolation and the victim's FE filter attenuation in the aggressor's TX band. Mitigation techniques can only be implemented in the victim's RX design. In both cases improved antenna isolation helps but the degrees of freedom are limited in small form factor UE especially when the aggressor and the victim operating frequencies are close to one another.

Victim desensitization due to the presence of a blocker may result from multiple mechanisms, as described in Figure 11.24.

A victim's RX LO phase noise is added to the wanted signal due to reciprocal mixing with the aggressor's transmit signal leakage:

1 Victim's RX reduced gain for wanted signal due to cross-compression on the TX signal leakage.
2 Second order intermodulation distortion (IMD_2) product of the TX signal leakage AM content falling at DC for Direct Conversion Receivers (DCR) which potentially overlap with the wanted signal.

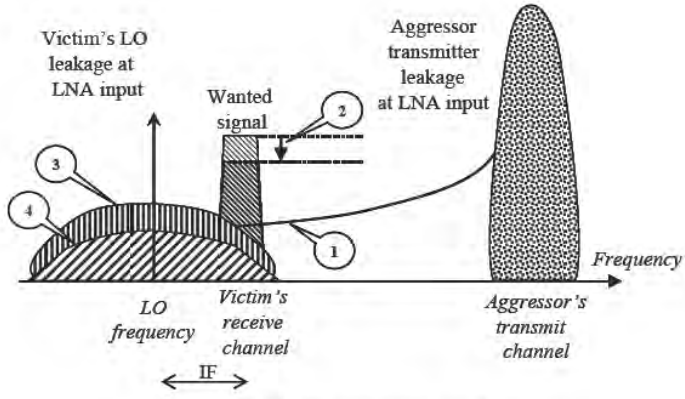


Figure 11.24 Desensitization mechanisms

3 Third order intermodulation product of victim's LO leakage with TX signal leakage AM content falling at DC which potentially overlap with wanted signal (known as cross-modulation).

For the first two cases, the actual aggressor signal characteristics and bandwidth do not play a role in the interference mechanism:

- Cross-compression is directly related to the signal peak power, which has been kept constant between WCDMA and LTE uplink thanks to the Maximum Power Reduction (MPR) introduced in 3G.
- Reciprocal mixing is directly related to interference leakage and the victim's LO Phase Noise at an offset equal to the difference between the aggressor's TX frequency and the victim's RX frequency. The larger distance improves both the selectivity on the aggressor leakage and the LO Phase Noise.

Similarly the aggressor's noise leakage is lower for higher frequency distance due to higher selectivity and lower Phase Noise at larger frequency offsets. From the above it can be seen that for mechanisms such as TX noise leakage, cross-compression and reciprocal mixing, the prime factor is the vicinity in the frequency domain of the aggressor and the victim. New band allocation for LTE has created new cases described in the band specific coexistence challenge section and in section 11.8.2.

The cross-modulation and the IMD_2 products, however, have a spectrum bandwidth (BW) directly related to the interferer's BW. Furthermore, the spectrum shape of these distortions depends on the signal statistics and modulation scheme. For these contributors it is difficult to propose a generic analysis of how the LTE interference compares to WCDMA. Analysis of which portion of the interfering spectrum is captured within the victim's channel bandwidth needs to be conducted for each LTE bandwidth and modulation combination. A comparison of IMD_2 products for WCDMA and LTE QPSK uplink modulated carriers can be found in section 11.8.2.1.

Band Specific Coexistence Challenges

As far as the LTE out-of-band TX noise emissions are concerned, the duplexer must provide enough attenuation in the following frequency bands:

- In band 11, the TX frequency is very close to GPS band. With the anticipated antenna isolation, more than 40 dB attenuation needs to be provided. GPS receivers may also have to improve their blocker handling capability to cope with LTE TX signal leakage.
- In bands 11/12/13/14, the TX frequencies are close to some of the TV bands, and consequently adequate attenuation of LTE TX noise is required.
- In bands 7/38/40, the TX frequencies are very close to the 2.4-GHz ISM band where BT and WLAN operate. With the anticipated antenna isolation, approximately 40 dB attenuation needs to be provided. Similarly the BT and WLAN TX noise needs to be kept low in the Band 7/38/40 RX frequencies. Given the blocker handling capability already required for the LTE FDD RX, the BT and WLAN blockers are not anticipated to be an issue. BT/WLAN receivers, however, may have to improve their blocker handling capability to cope with the LTE TX signal leakage.

As far as the LTE TX modulated carrier is concerned, with a 23 dBm output power, special attention must be paid to controlling the LTE TX second harmonic power level in the following aggressor/victim combinations:

- In Band 7/38, the TX second harmonic falls within the 5 GHz WLAN band, thus requiring the LTE transmitter to attenuate it to close to −90 dBm. In comparison to the 3GPP LTE which requires a harmonic level to be lower than −40 dBm, this represents a significant challenge. Also, the WLAN RX must be sufficiently linear to prevent regenerating this second harmonic in the presence of the LTE fundamental TX signal leakage.
- In Band 13, the TX second harmonic falls within the GPS band thus requiring the LTE transmitter to attenuate it to lower than −110 dBm. This level of attenuation might be hard to achieve practically. Similarly to the above coexistence test case, the LTE TX leakage presence at the GPS RX input also imposes linearity requirements to prevent regenerating this second harmonic.

There are also other cases where a harmonic of the LTE TX falls in the 5 GHz WLAN or UWB bands. These are usually higher harmonic orders, which should be less of an issue.

11.6.3 RF–Baseband Interface Design Challenges

In mobile handsets, the RF transceiver and the BB processor are often implemented on separate ICs. The TRX IC normally contains analog signal processing, while the BB IC is predominantly digital. Therefore analog-to-digital (A/D) and digital-to-analog (D/A) conversions are required in receive and transmit paths respectively. The location of these converters is a critical choice in wireless system design. If the converters are implemented in the RF transceiver, discrete time domain (digital) data are transferred across the interface between the BB and the TRX. On the other hand, if the converters are located in the BB IC, the interface comprises continuous time domain (analog) differential I/Q signals.

Figure 11.25 shows a typical example of an analog based interface with receive diversity. Here this mobile is designed to support all of these radio operating options: '3G' Wideband Code-Division Multiple Access (WCDMA) on one frequency band plus GSM on four frequency bands. It is readily seen that 24 separate interconnections are required.

Despite being in mass production from many vendors, analog interface solutions face the following challenges and criticisms. The large number of interconnecting pins increases package size and cost on both sides, and complicates printed circuit board routing. This is a particular problem because the sensitive analog interconnections also require special shielding to achieve the required performance. Furthermore, such interfaces are proprietary, forcing handset makers to use particular pairs of BB and RF devices. While some IC vendors prefer this forced restriction, it is a disservice to the industry by restricting competition and disabling creative approaches in either the BB or RF side alone from adoption in products. BB devices must include the necessary large number of analog blocks, increasing their die size.

This last point is particularly important, involving much more than the obvious economic issues. Clearly, larger die areas increase the cost of IC manufacture. Also these analog designs do not generally shrink as well as digital cells with progressively smaller CMOS processes. Thus, the fraction of the BB IC that these analog designs take up increases. Additionally, the analog circuitry yield in production may be lower than the digital circuitry, so a perfectly fine

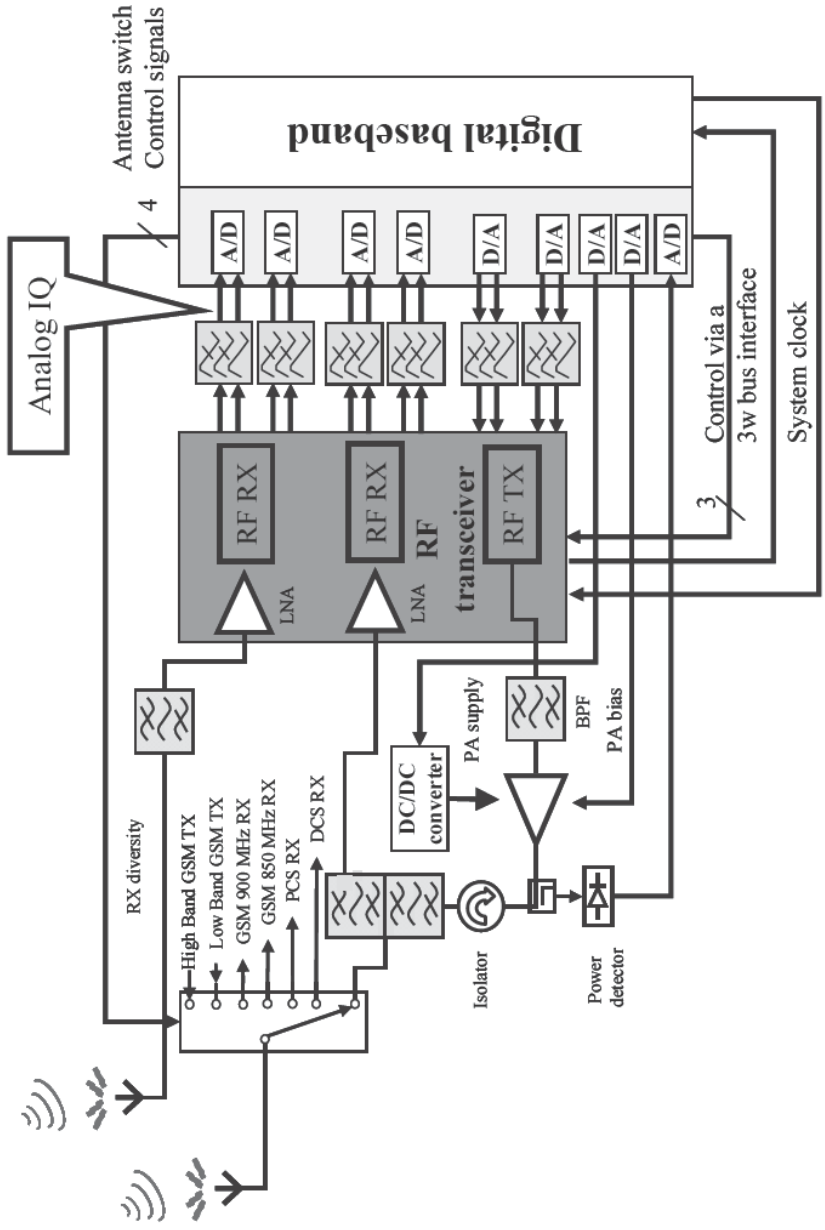


Figure 11.25 Typical HSPA monoband 3G, quad band 2G transceiver/BB block partitioning in analog I/Q interface

digital BB would be thrown away when an analog section fails. Even more of a problem is that analog design information on a new CMOS process is always provided later than the digital design information – sometimes much later. When new CMOS processes become available, this forces BB designs to wait until all the necessary design information is available and sufficiently qualified. In the newer CMOS processes, analog designs are actually getting more difficult, not easier.

Taking all of this together, it becomes clear that a purely digital interface is greatly desired, and would be a huge benefit to the mobile industry. The main organization leading the standardization effort is MIPI® – the Mobile Industry Processor Interface Alliance [12]. Combining the terms 'digital' and 'RF' together into the name 'DigRF^{SM}', this interface is already in its third evolutionary step, as listed in Table 11.15.

Adopting these digital interfaces changes the earlier block diagrams to that of Figure 11.26. The dual objective of eliminating analog designs from the BB device and reducing pin count down to only 7 pins is met. For DigRF^{SM} v4, the interface data rate necessary to support just a single antenna receive chain in a 20 MHz LTE application reaches 1248 Mbps.

One of the biggest challenges in DigRF^{SM} v4 is EMI control. EMI was not a major concern for DigRF v2 and v3, because these interface data rates are well below the mobile's radio operating frequencies. With all the bands now under consideration for LTE mobile device use, the internal radio Low Noise Amplifier (LNA) section is sensitive to energy across frequencies from 700 MHz to nearly 6 GHz. The DigRFv4^{SM} data rates and common mode spectrum emissions now exceed the radio frequencies of several mobile operating bands. This is a huge problem, clearly seen in Figure 11.29(a).

Before addressing EMI mitigation techniques, it is essential to understand how much signal isolation is available from the physical packaging of the RF IC. Adopting the vocabulary of EMI engineering, the LNA input is called the 'victim', while the interface is called the 'aggressor'. As we can see in Figure 11.27 (left), there are many possible paths for energy on the interface to couple into the 'victim' LNA. Figure 11.27 (right – plain dots) shows one example of practically available isolation.

The upper limit to the maximum amount of aggressor noise PSD allowed at the victim's input pins is established with the set of characteristic curves shown in Figure 11.28. Considering worst cases, a cellular LNA with an intrinsic NF of 3 dB may be degraded to a 3.5 dB NF. According to these charts, the interfering noise must be at least 6 dB below the kT floor, i.e. at or below −180 dBm/Hz. Similarly, a more sensitive GPS LNA with an intrinsic NF of 2 dB may be degraded by at most 0.25 dB. Evaluating these two cases, we conclude that the interference must be below −180 dBm/Hz and −184 dBm/Hz for a cellular and a GPS RX respectively.

The limit curve shown in Figure 11.29 (a–c) is based on combining the isolation model with the cellular LNA noise tolerance and shows violation in Figure 11.29(a). Mitigation techniques are therefore required to ensure product success. First on the mitigation techniques list is to

Table 11.15 DigRF version evolutions

DigRF version	Standard	Interface bitrate (Mbps)
v2: 2G	GSM/GPRS/EDGE	26
v3: 3G	2G + HSPA	312
v4: 4G	3G + LTE	1248, 1456, 2498, 2912

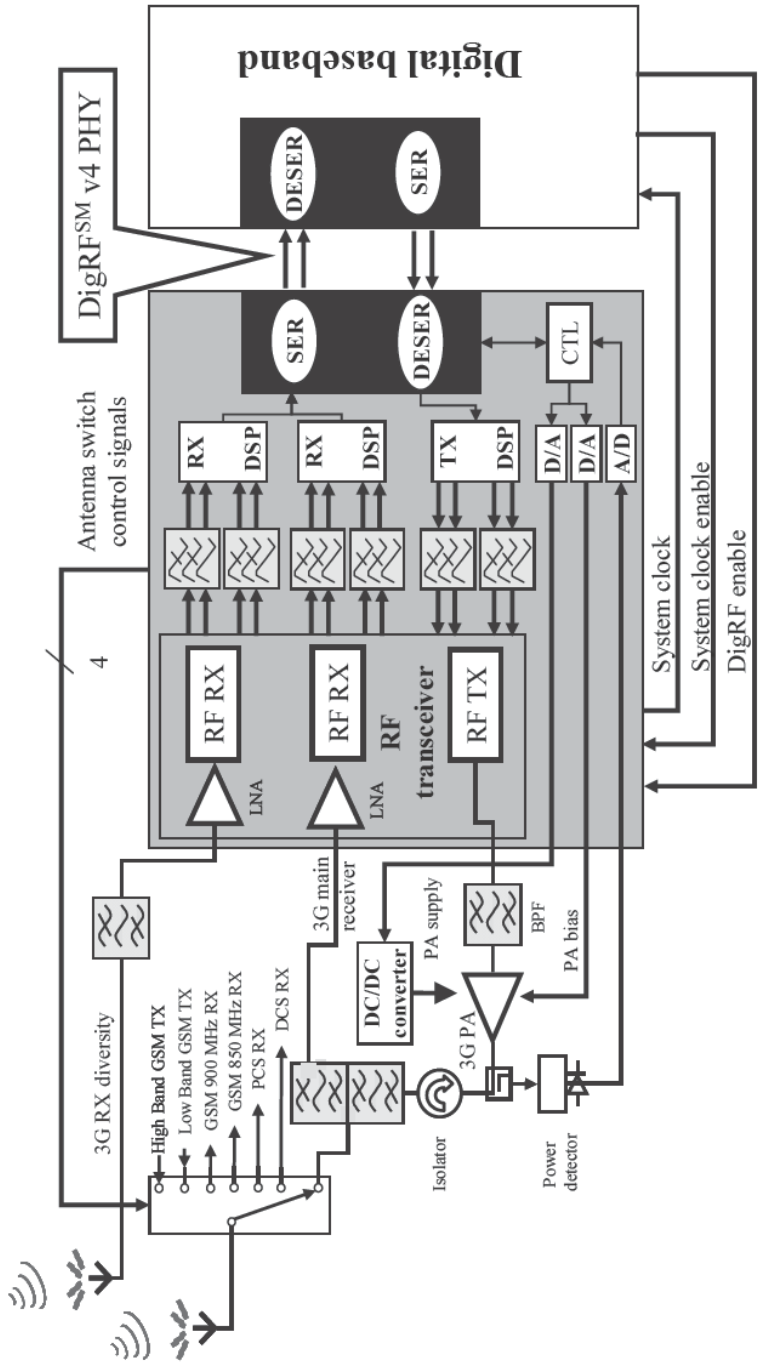


Figure 11.26 Example of digital RF–BB interface application using DigRF^SM v4 with RX diversity. A/D = analog to digital converter, D/A = digital to analog converter, SER = serialize, DESER = deserialize. DSP functions include A/D, D/A, digital filtering and decimation (specific to RX)

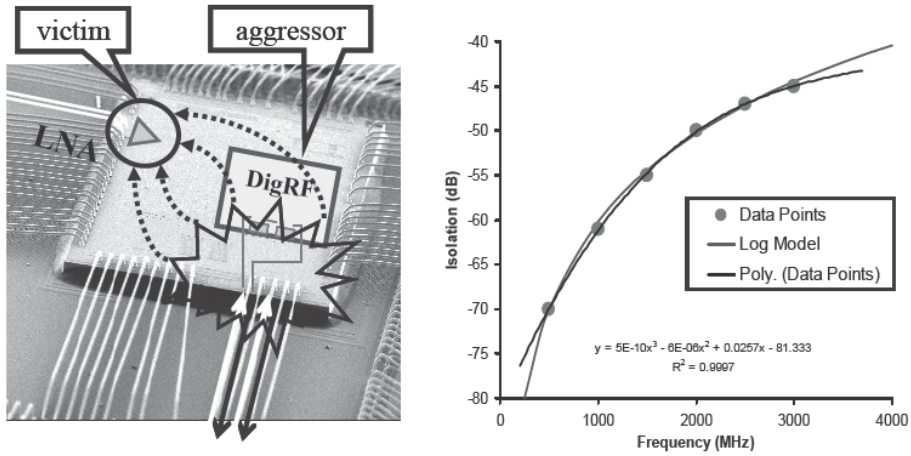


Figure 11.27 Left: examples of EMI coupling paths within an RF IC. Right: in-package isolation example at several frequencies

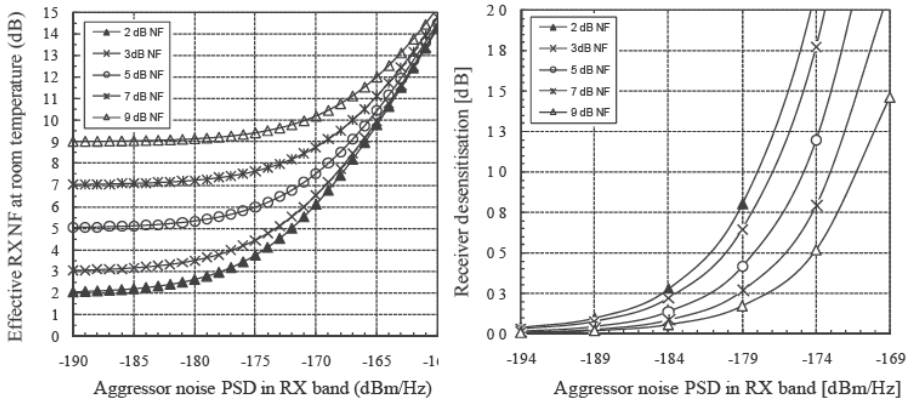


Figure 11.28 Response of a 'victim' LNA to additional noise provided at its input. Left: effective noise rise of a victim NF referred to its LNA input (room temperature). Right: corresponding victim's desense vs aggressor noise PSD

use control of the bit edge slew rate, and the situation is greatly improved as shown in Figure 11.29(b). Secondly, an alternate interface frequency is provided, which can be switched to if a mobile product is experiencing spurious interference from the interface. This alternate interface data rate, 1456 Mbps, is also particularly useful in protecting a local GPS RX input by ensuring the LNA is now operating close to a frequency null of the interface spectrum, as shown in Figure 11.29 (c).

As Figure 11.30 shows at 1248 Mbps, however, a single antenna 20 MHz LTE application requires 70% duty cycle of the interface. Adding the LTE diversity RX, the requirement exceeds 100%! One solution consists in doubling the number of 1248 Mbps lanes. Alternatively, for

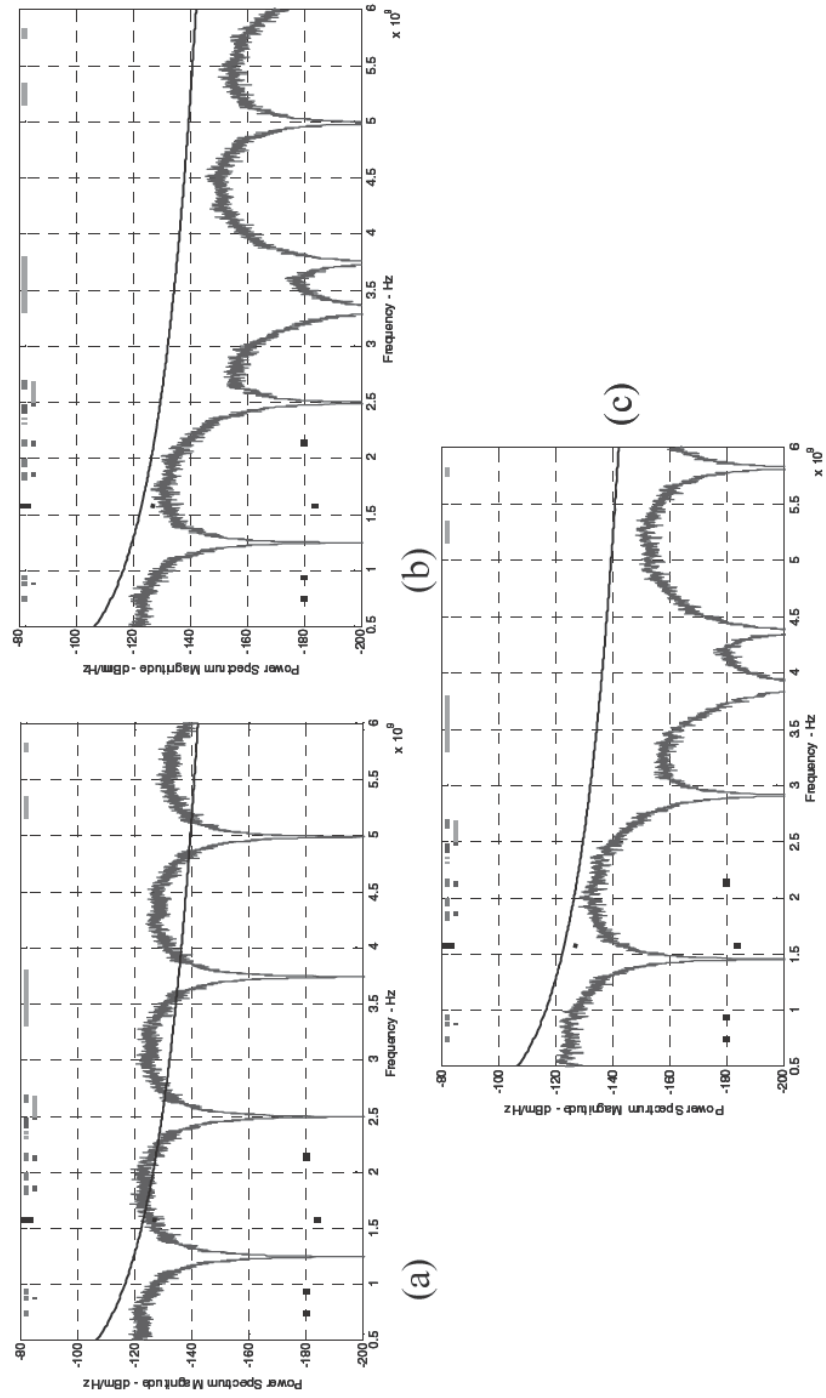


Figure 11.29 Effects from adding EMI mitigation features to the DigRF/M–PHY interface: (a) imperfect differential signaling, (b) slew rate control, and (c) alternate frequency selection of 1456 Mbit/s. Horizontal bars at top of each diagram indicate the location of telecommunication standards victim's frequency bands from 700 MHz to 6 GHz.

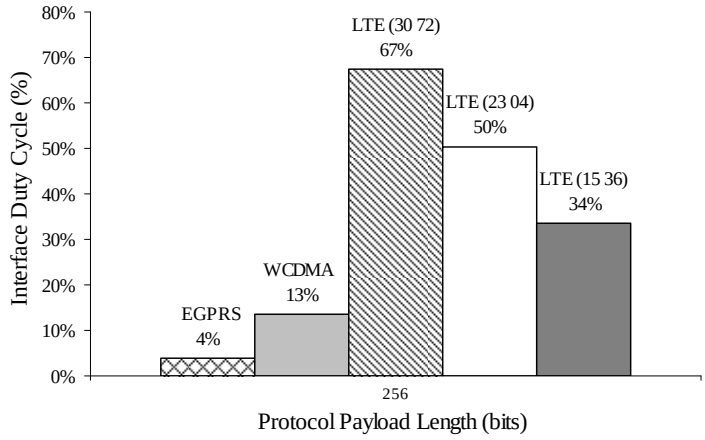


Figure 11.30 Interface duty-cycles for DigRF^{SM} v4 with 256 bit payload field size at 1.248 Gbps

applications which do not have significant EMI sensitivity, doubling the interface bit-rate to 2496 Mbps solves the capacity issue, but now generates a main lobe of the interface spectrum that spans over all bands, including the super-sensitive GPS band.

11.6.4 LTE vs HSDPA Baseband Design Complexity

11.6.4.1 Equalization

LTE aims at reducing the cost per bit, among other challenging goals such as increasing spectral efficiency and allowing flexible bandwidth deployment. From the receiver point of view, one key measure is the required complexity and in particular the comparison with previous releases such as WCDMA and HSDPA. Figure 11.31 shows the estimated complexity based on the baseline receiver for all transmission modes, as introduced in section 11.9, excluding channel decoding operation complexity. Note that the complexity of the LTE receiver grows linearly with respect to system bandwidth and the corresponding maximum nominal throughput. Interestingly, MIMO mode requires less than double the SIMO mode complexity. Comparing the estimated complexity of a Release 6 optimistic HSDPA receiver assuming a low complexity chip-level equalizer category 10 device with 13 Mbps throughput, LTE is shown to be an attractive technology at least from an inner receiver complexity point of view because of OFDM choice. Assuming 5 MHz bandwidth and 16QAM modulation, LTE offers the same throughput with nearly half of the required complexity of HSDPA.

Nevertheless, despite this advantage in the relative comparison for the 5 MHz bandwidth case, the LTE receiver requires a considerable complexity increase compared to HSDPA since the UE must support a system bandwidth of 20 MHz as a minimum requirement and therefore still constitutes a challenge for mobile and chip-set manufacturers. The LTE class 4 device with 150 Mbps capability has a complexity approximately four times higher than a HSDPA category 10 device with 13 Mbps capability.

Further details of the repartition of complexity among inner receiver stages are also given in Figure 11.32. Along the supported modes, the complexity associated with FFT operations – always

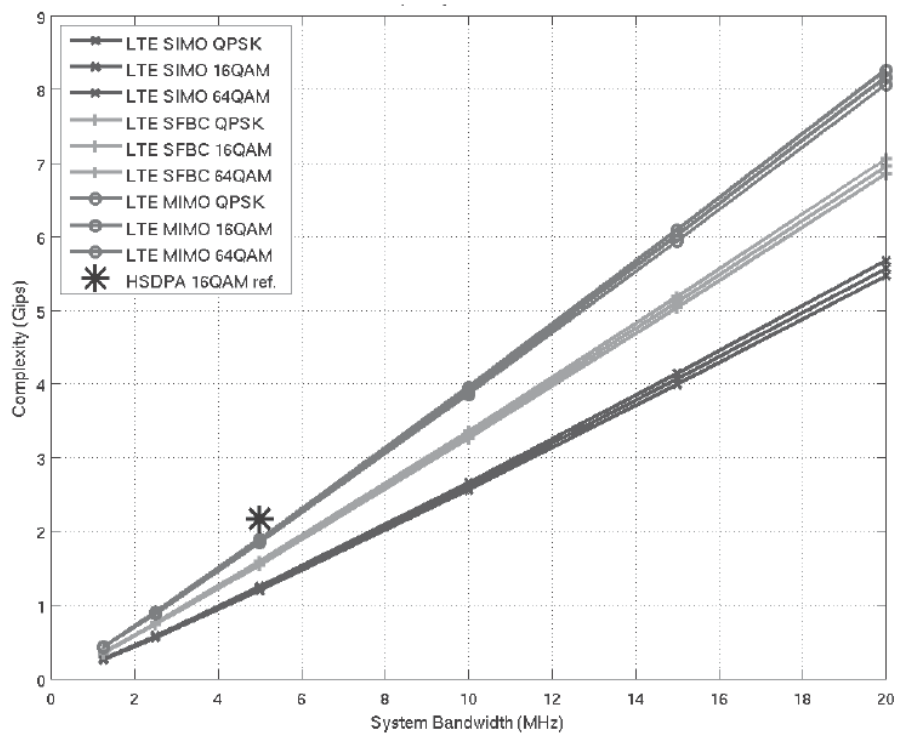


Figure 11.31 Complexity of LTE receiver

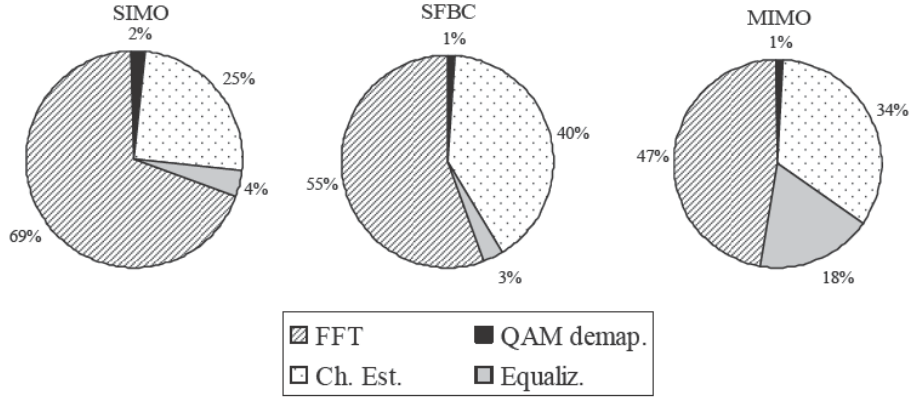


Figure 11.32 Complexity repartition in LTE receiver

equal to two as the number of receiving antennas – becomes less important for overall complexity because of the increased needs of channel estimation and equalization. With MIMO the operation FFT takes 47% of the calculations while channel estimation takes 34% and equalization 18%.

11.6.4.2 Turbo Decoder

In LTE, the efficiency and the complexity of the channel decoding operation grows considerably as the maximum nominal data rate increases compared to previous releases. The Turbo decoder must support rates up to 150 Mbps in a category 4 device. The Turbo code algorithm in LTE is similar to the Turbo code in HSDPA. The Turbo decoder processing delay, and hence its throughput, is roughly linearly proportional to the number of turbo decoding iterations and the size of the transport block. As a consequence, we can describe the Turbo decoder efficiency as

$$\eta = \frac{N_{it} \cdot r_{max}}{f_{clock}} \tag{11.12}$$

where N_{it} is the number of iterations, r_{max} is the maximum data rate and f_{clock} is the working clock frequency of the turbo decoder. Table 11.16 and Table 11.17 show the efficiency/clock-frequency tradeoff for the HSDPA and LTE cases assuming a number of iterations equal to 8 and a maximum data rate of 13 Mbps and 150 Mbps respectively.

Efficiency can be seen as a measure of the parallelization required within the Turbo decoding operation. Comparing the two tables, it seems evident that to support a LTE data rate within reasonable clock frequencies – strictly dictated by power consumption and hardware technology constraints – a high level of parallelization is imposed.

Table 11.16 Turbo decoder efficiency/clock frequency tradeoff for 13 Mbps

Efficiency η (b/s)	Clock frequency (MHz)
5.5	20
2.7	40
1.8	60
1.4	80
1.1	100
0.9	120
0.8	140
0.7	160
0.6	180

Table 11.17 Turbo decoder efficiency/clock frequency tradeoff for 150 Mbps

Efficiency η (b/s)	Clock frequency (MHz)
12.6	100
10.5	120
9.0	140
7.9	160
7.0	180
6.3	200
5.7	220
5.2	240
4.8	260

It is worth mentioning that Turbo decoding parallelization was not attainable given the construction of the interleaver used for the encoding and decoding procedure of previous releases of the 3GPP standard. This problem was solved for the LTE specification and a contention-free interleaver has been built to allow any level of parallelization expressed as a power of two (i.e. 2, 4, 8, 16).

Nevertheless, parallelization itself does not come for free and also constitutes a challenge for manufacturers: it demands an in-depth change in the Turbo decoding algorithm, increases the surface and gate count of an amount proportional to the parallelization factor and, in principle, an operational higher clock frequency.

As a final consideration, it is worth noting that Turbo Decoder complexity is in principle much higher than the complexity of the rest of the mobile receiver signal processing: complexity grows linearly with the target throughput and number of iterations. It can be approximated by [48]:

$$C_{TD} = 200 \cdot N_{it} \cdot r_{max} \text{(MIPS)} \tag{11.13}$$

For LTE, with a maximum data rate of 150 Mbps, the complexity approaches 240 Gips while the FFT, equalization and channel estimation complexity together are in the order of 9 Gips.

The Turbo Decoder then seems to require nearly 96% of the overall complexity in the receiver but this is a rather misleading conclusion. Signal processing involved in the Turbo Decoding is mainly addition operations in the max-log-MAP case. Equalization and associated functions require multiplication operations instead and in principle larger fixed-point sizes. For the implementation of choices making a difference, signal processing complexity is a valuable measure of the challenge required among standard evolutions for the same functionality but cannot allow for more general comparisons.

11.7 UE RF Transmitter

11.7.1 LTE UE Transmitter Requirement

11.7.1.1 Transmitter Output Power

The LTE specified maximum output power window is the same as in WCDMA: 23 dBm with a tolerance of ±2 dB. The earlier WCDMA specifications used 24 dBm with a tolerance of +1/−3 dB. The SC-OFDMA [13] has a higher Peak-to-Average Ratio (PAR) than the HPSK modulation of WCDMA. Figure 11.33 shows the ACLR performance of a WCDMA PA under QPSK modulated SC-OFDMA signal. The WCDMA operation assumes 24 dBm and LTE operation at 23 dBm output power. The main difference is related to the spectral shape and the fact that the occupied BW is slightly higher in LTE (4.5 MHz) than that of WCDMA (99% energy in 4.2 MHz), and consequently the ACLR is slightly degraded.

In a similar manner to HSDPA and HSUPA, a Maximum Power Reduction (MPR) has been introduced in LTE to take into account the higher PAR of 16QAM modulation and some resource block allocation. Again this ensures a proper ACLR under a complex set of TX modulations. Compared to WCDMA, where the only direct interference falling in the RX band is related to spurs and the OOB noise, the LTE TX linearity should also be considered. It is particularly important for 10 MHz BW in the 700 MHz bands where the duplex distance is only 30 MHz

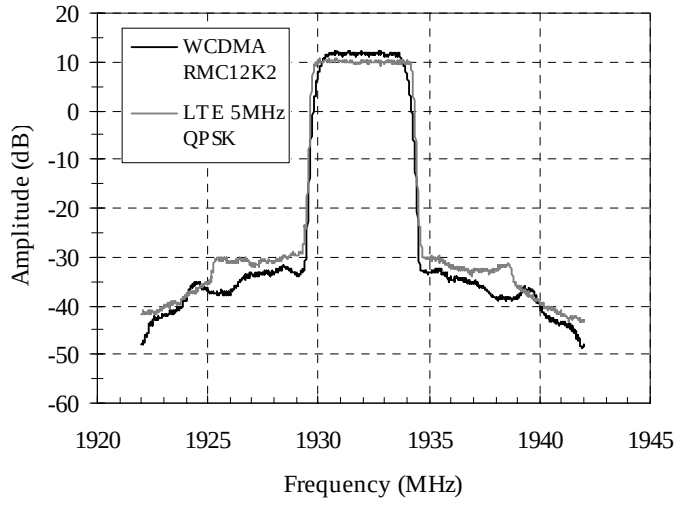


Figure 11.33 24 dBm WCDMA and 23 dBm LTE 5 MHz QPSK spectrums

and where 5th and 7th order intermodulation products of the TX overlap with the RX channel. This phenomenon is discussed in section 11.8.2.

LTE power control ranges from −40 dBm to +23 dBm. WCDMA transmitters offer −50 dBm to +23 dBm of Transmit Power Control (TPC) with 1 dB resolution. The same TPC techniques can be applied in LTE taking into account all MPR cases.

11.7.2 LTE Transmit Modulation Accuracy, EVM

The LTE 16QAM modulation case places stringent requirements on the TX imperfections to meet the Error Vector Magnitude (EVM) budget. Overall the errors should be further minimized for LTE. Each contributor can be analyzed separately to meet the 17.5% EVM budget in QPSK and 12.5% in 16QAM. EVM measurements are done after Zero-Forcing (ZF) equalization in the test equipment. For comparison, the WCDMA HPSK EVM budget is equal to 17.5%, but does not assume any equalization in the test instrument. In practice the design target is around 8%. The TX imperfections contributing to the EVM budget are considered in the next section.

- Carrier rejection: As the measurement is made after ZF-equalization the test equipment partially removes this contribution. LTE standardization has set a separate specification for carrier rejection to ensure that the carrier level stays within reasonable limits for the ZF algorithm. At low output power small DC offsets in the TX chain generate a high carrier leakage to a point where TPC accuracy may not be met anymore. This problem is already severe in WCDMA where the TPC range is 10 dB larger at the bottom end. For this reason carrier leakage compensation is implemented in RF TRXs and this technique generally achieves 40 dB of carrier rejection throughout the TPC range making this contribution negligible in the EVM budget.
- Even order distortion: Even order non-linearity contributes mainly to ACLR as the main effect is to enlarge the transmitted spectrum. As ACLR requirements are 33 dB and 43 dB in adjacent and alternate channels, these contributions are usually negligible in terms of EVM. AM/AM in the PA can be considered similarly.

- LO phase noise: The induced jitter generates phase error in the modulation constellation thus contributing to EVM. As the SC-OFDMA used in transmit is single carrier the phase noise has a contribution similar to the WCDMA case.
- PA distortion: AM/PM distortion (together with AM/AM) has a contribution to EVM and also generates asymmetrical spectral energy in the adjacent channels. AM/PM and AM/AM distortion increases as the transmit power is increased. Hence, the LTE MPR scheme ensures that PAs dimensioned for WCDMA will not dominate the overall EVM budget.
- Image: The signal image generated by the quadrature imperfections in the up-mixing process can be considered as band noise contributing to the TX signal SNR. The use of 16QAM requires better control of such imperfections relative to WCDMA [14].
- Group delay distortion: Compared to WCDMA where the I/Q BB BW is 2 MHz for a minimum duplex distance of 45 MHz, LTE has an I/Q BW of 5 MHz for 30 MHz duplex distance or 10 MHz for 80 MHz duplex distance. Significant BB filtering is required to ensure that BB I/Q noise at the duplex frequency offset is very low (to allow SAW-less transmit architecture). This means that stop-band attenuation for LTE BB filter requires a more stringent filter specification than for WCDMA, potentially introducing more in-band distortion contributing to EVM.

Overall, the LTE EVM specification requires a similar effort to WCDMA for RF imperfection but special attention needs to be paid to the higher bandwidth and smaller duplex distance. For the same reasons, the TX out-of-band noise is more difficult to achieve in LTE.

11.7.3 Desensitization for Band and Bandwidth Combinations (Desense)

Although TX out-of-band noise in RX band is not an explicit requirement in the 3GPP WCDMA specifications, reference sensitivity measurements are made in the presence of the full TX power. To meet the reference sensitivity the TX noise leakage levels must stay below the thermal noise floor. Recent efforts in TX architecture have allowed the removal of the filter between the TRX IC and the PA. Interstage filters were used to clean the TRX noise before further amplification. Removal of the filter is made feasible by careful design of every noise sources in the RF TRX IC.

As discussed in section 11.6, it is essential that the addition of the LTE functionality and the support of new bands do not require reintroduction of these filters. Two new challenges must be solved for this to be feasible: the smaller duplex separation of certain new band configurations (12/13/14) or a wider channel bandwidth. In some specific cases these two challenges are combined. This issue has been recognized by the standardization body, as shown in Table 11.4 where the relaxed sensitivity requirements have been defined for certain combinations of band and bandwidth. The severity of the UE self-desense is further described in section 11.8.2.

11.7.4 Transmitter Architecture

11.7.4.1 Transmit RF Modulator

Direct up conversion is the obvious choice for a 2G/3G/LTE multi-mode TX. It is the de facto standard for WCDMA. The large BW requirements of LTE would pose further challenges to alternative architectures such as polar modulation or other non-Cartesian approaches [15], [16].

This is especially true if the external filter requirement is to be relaxed for FDD. As briefly discussed in the previous sections, thanks to a pragmatic approach in the LTE standardization there is a minimum number of modifications needed to provide LTE transmit capability from a Cartesian transmitter already covering 2G and WCDMA. The main modifications lie in the higher bandwidth requirement on the BB DAC and filter to cover all the different BWs with low out-of-band noise. In addition, extra RF bands need to be supported, which requires extension of the PLLs' tuning range and RF buffer bandwidths.

11.7.4.2 Multi-mode Power Amplifier

As discussed in section 11.6, one essential simplification of the worldwide RF FE is the use of a single PA line-up covering multiple bands and multiple modes. The band coverage can be clustered in the following way:

- one Low Band (LB) PA line-up covering all bands between 698 MHz and 915 MHz;
- one High band (HB) PA line-up covering all bands between 1710 MHz and 2025 MHz;
- one Higher band PA line-up covering all bands between 2300 MHz and 2620 MHz.

The only band that is not covered is the Japanese Band 11, which can be added for this specific phone configuration or even replace one of the other wideband PAs depending on the overall band support.

Each of these line-ups has to support different modulation schemes and maximum output power depending on band–mode combinations. These combinations can be found in Table 11.18, where 2 dB and 3 dB PA to antenna losses are considered for TDD and FDD modes respectively.

Taking into account the different PAR inherent to each modulation scheme and the required back-off to meet ACLR requirements, a given line-up has to meet a range of saturated output power (P_{outsat}) capabilities to achieve best efficiency/linearity tradeoffs. For example the LB PA has to achieve 35 dBm P_{outsat} capability for GSM, GMSK having only phase modulation the PA can be driven into saturation and achieve best efficiency. In WCDMA mode, it needs close to 31 dBm P_{outsat} capability to allow sufficient back-off. A GMSK capable PA would have a very low efficiency in the WCDMA mode if nothing was done to decrease its output power capability. This would prove the multi-mode PA architecture to be uncompetitive in terms of performance especially in 3G modes, which are already challenging for talk time. The only

Table 11.18 Modulation scheme and maximum output power per band configurations for a multi-mode PA

Modes	Sub-bands		
	Low band 698–915 MHz	High band 1710–2025 MHz	Higher band 2300–2620 MHz
GSM (GMSK)	35 dBm	32 dBm	n-a
EDGE (8PSK)	29 dBm	28 dBm	n-a
WCDMA (HPSK)	27 dBm	27 dBm	27 dBm
LTE (QPSK)	26 dBm	26 dBm	26 dBm

way to reach the best PA efficiency in all modes is to tune the output stage load line. This can be achieved in two ways:

- Tuning the output matching to transform the load impedance (usually 50 Ω) into the desired load line for every mode. This technique can be achieved for a small set of impedances and usually results in a lower Q output matching.
- Tuning the PA supply: In this case the saturated output power capability is proportional to the square of the supply voltage. If the supply is varied using a DC/DC converter then efficiency can be optimized for every mode. This technique is becoming more and more popular [17] and has the benefit of allowing optimum efficiency for every mode.

11.7.4.3 Conclusion

Although this section does not provide a detailed analysis of the LTE transmitter requirement it discusses how these requirements can be achieved by simple extrapolation from GSM/EDGE/WCDMA architecture and performance. It is shown that all three modes can be supported with a single transmitter architecture with the minimum of extra hardware when techniques such as co-banding and DC/DC controlled multi-mode PA are introduced. This ensures easy migration towards LTE for mobile equipment manufacturers.

11.8 UE RF Receiver Requirements

The purpose of this section is to highlight the main differences between Frequency Division Duplex (FDD) WCDMA and Full Duplex (FD) – FDD LTE UE RF receiver (RX) system requirements. Throughout this section, WCDMA and LTE are used to refer to UTRA and E-UTRA respectively. The objective of the LTE UE RF test requirements listed in section 7 of [1] is to quantify RF impairments that have an impact on the network performance. These impairments include Noise Figure (NF), receiver Error Vector Magnitude (EVM), selectivity at different frequencies, including adjacent channel, etc. The test requirements have been derived to ensure the industry can make the best possible re-use of IPs developed for WCDMA UEs. This is highlighted in the description of the UE reference sensitivity level and the Adjacent Channel Selectivity (ACS) system requirements. For this reason, the chapter focuses on some of the novel design challenges that are specific to the LTE downlink modulation schemes and its associated new frequency bands. In this respect, the following challenges are discussed: RX self-desensitization, ADC design challenges, and the impact of RX EVM contributors in OFDM vs single carrier.

11.8.1 Reference Sensitivity Level

The reference sensitivity power level is the minimum mean power applied to both UE antenna ports at which a minimum throughput requirement will be fulfilled. The throughput will equal or exceed 95% of the maximum throughput for a specified Fixed Reference Channel (FRC). FRCs are similar to the WCDMA reference measurement channel and in the sensitivity test case the downlink carrier uses QPSK modulation and 1/3 coding rate.

The sensitivity requirement verifies the UE RX NF, which for FDD operation may include noise contributions due to the presence of the UE uplink modulated carrier as described in the next section. Other receiver impairments such as EVM are included within the demodulation performance requirements where a higher Signal-to-Noise Ratio (SNR) is applied. Therefore, the selected FRC provides a reference maximum throughput defined for low SNR operation. Beyond this purpose, the UE reference sensitivity is of primary importance in 36.101 since it also serves as a baseline to set the downlink carrier power for ACS and blocker test requirements.

With LTE, the reference sensitivity requirements present several major differences compared to the WCDMA system requirements:

- LTE flexible bandwidth requires the RF receiver to implement reconfigurability of its channel select filters to support 1.4, 3, 5, 10, 15 and 20 MHz bandwidths.
- The reference sensitivity must be tested on two antenna ports: main and diversity receiver.
- New frequency bands with small Duplex Gap (DG), such as Bands 5, 6, 8 and 11 introduce novel UE self-desense considerations when UE uplink transmission bandwidths are greater than 5 MHz. Also Bands 12, 13, 14 and the more recent Band 17, all falling under the terminology UMTS 700, are impacted by these limitations. This is not a major concern in, for example, WCDMA Band I devices because the UE self-desense is primarily dominated by the UE transmitter chain noise floor at large frequency offsets, as can be seen in Figure 11.34 (left). Adding bands with small Duplex Distance (DD) and small DG for which large BW deployment is planned, now places the UE receiver directly inside the bandwidth of the transmitter chain adjacent channel leakage shoulders, as shown in Figure 11.34 (right). The test requirement therefore includes several relaxations, which can be found in section 11.8.2.

The reference sensitivity level calculation is given by Equation 11.14. The differences with the eNodeB equation (11.6) reside in the UE dimensioning assumptions: the NF is set to 9 dB, the Implementation Margin (IM) is equal to 2.5 dB, and a 3 dB correction factor is applied to account for the dual antenna reception gain. In addition, Equation 11.14 includes a frequency band specific relaxation factor, D_{FB}. Note that D_{FB} is not official 3GPP terminology and is equal to 1, 2 and 3 dB for bands in which the DD/DG ratio is greater than 1.5, 2 and 4 dB respectively, and 0 otherwise [18].

$$P_{REFSENS}[\text{dBm}] = -174[\text{dBm/Hz}] + 10\log_{10}(N_{RB} \cdot 180k) + NF + SNR + IM - 3[\text{dB}] + D_{FB} \quad (11.14)$$

D_{FB} is a metric which reflects design challenges for front-end components such as the duplexer. For a given technology resonator quality factor (Q factor), the higher the DD/DG ratio:

- The higher the Insertion Loss (IL) of each of the duplexer Band-Pass Filters (BPFs). In RX, every decibel (dB) lost directly adds to the UE NF, therefore justifying a sensitivity relaxation. In TX, every dB lost causes heat dissipation in the duplexer.
- The sharper the TX BPF roll-off requirements to prevent TX noise from polluting the RX band. Due to heat dissipation, and mass production process variations, TX BPF must be designed with a slightly increased 3 dB cut-off frequency 'F_c' to ensure carriers located at

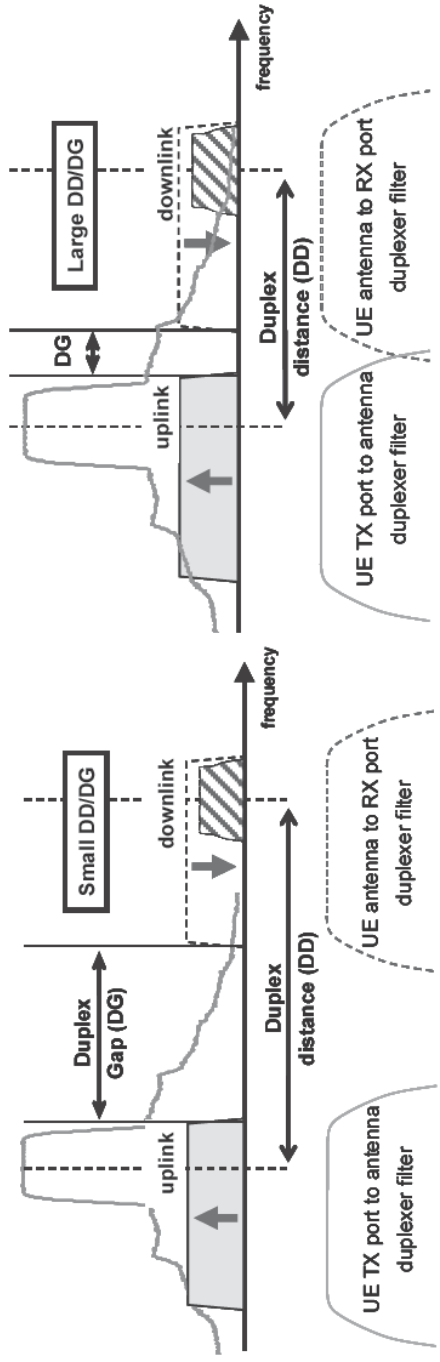


Figure 11.34 Illustration of large duplex gap (left) vs small duplex gap (right) frequency bands

the edge of the band do not suffer from insertion losses that are too high. In these bands the RX is more vulnerable to TX noise leakage, thereby justifying an extra relaxation.

With these assumptions listed, with the exception of the grey shaded values for which UE self-desense applies (cf. section 11.8.2), most of the FDD mode sensitivity levels listed in Table 11.19 can be computed.

For example, Figure 11.35 shows that for a 5 MHz channel bandwidth ($N_{RB} = 25$), the reference sensitivity level is equal to −100 dBm. The simulated SNR for 95% of the maximum throughput is assumed to be equal to −1.0 dB. Recent link level simulations [19], [20], indicate slight variations around this nominal with SNRs of −1.4 and −0.9 dB respectively (FRC A1–3). The throughput is approximately 2.1 Mbps for 5 MHz channel in a sensitivity test.

Note that the 9 dB NF assumption used in LTE is very similar to the assumptions made for WCDMA UEs.

Depending on the balance of the link budget, the UE NF is a relevant parameter when planning cell coverage area. The fact that LTE NF requirements are similar to WCDMA commercial

Table 11.19 UE reference sensitivity levels applied to each antenna port for QPSK modulation. Values denoted with * in grey shaded cells are combinations for which a relaxation in N_{RB} and maximum output power is allowed to prevent UE self-desense.

	Channel bandwidth								
Band	1.4 MHz (dBm)	3 MHz (dBm)	5 MHz (dBm)	10 MHz (dBm)	15 MHz (dBm)	20 MHz (dBm)	Duplex Mode	DD/DG	D_{FB} (dB)
1	–	–	–100	–97	–95.2	–94	FDD	1.46	0
2	–104.2	–100.2	–98	–95	–93.2*	–92*	FDD	4	2
3	–103.2	–99.2	–97	–94	–92.2*	–91*	FDD	4.75	3
4	–106.2	–102.2	–100	–97	–95.2	–94	FDD	1.13	0
5	–104.2	–100.2	–98	–95*			FDD	2.25	2
6	–	–	–100	–97*			FDD	1.29	0
7	–	–	–98	–95	–93.2*	–92*	FDD	2.4	2
8	–103.2	–99.2	–97	–94*			FDD	4.5	3
9	–	–	–99	–96	–94*	–93*	FDD	1.58	1
10	–	–	–100	–97	–95.2	–94	FDD	1.18	0
11	–	–	–98	–95*	–93.2*	–92*	FDD	2.09	2
12	–103.2	–99.2	–97	–94*			FDD	2.5	3*
13	–103.2	–99.2	–97	–94*			FDD	1.48	3*
14									
…									
17	–104.2	–100.2	–98	–95*			FDD	1.67	1*
…									
33	–	–	–100	–97	–95.2	–94	TDD		
34	–	–	–100	–97	–95.2	–94	TDD		
35	–106.2	–102.2	–100	–97	–95.2	–94	TDD		
36	–106.2	–102.2	–100	–97	–95.2	–94	TDD		
37	–	–	–100	–97	–95.2	–94	TDD		
38	–	–	–100	–97	–95.2	–94	TDD		
39	–	–	–100	–97	–95.2	–94	TDD		
40	–	–	–100	–97	–95.2	–94	TDD		

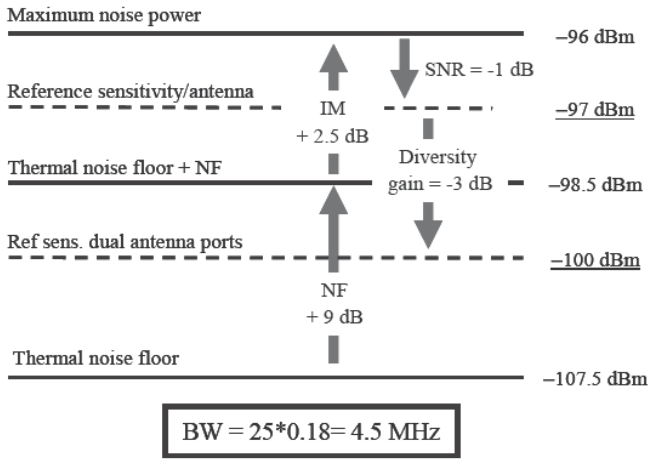


Figure 11.35 Reference sensitivity budget for LTE 5 MHz QPSK (N_{RB} = 25)

devices eases early delivery of UEs. In this way, the LTE standard provides a NF requirement small enough to guarantee good cell coverage, but not too small to allow system solution designers to deliver devices with better performance than the minimum requirement. This last point is important since sensitivity is most often a parameter used as a key selling argument. In this respect, LTE commercial devices from different vendors will deliver different sensitivity levels just like their WCDMA and GSM predecessors. An example of variability in WCDMA commercial reference sensitivity is shown in Figure 11.36 [21].

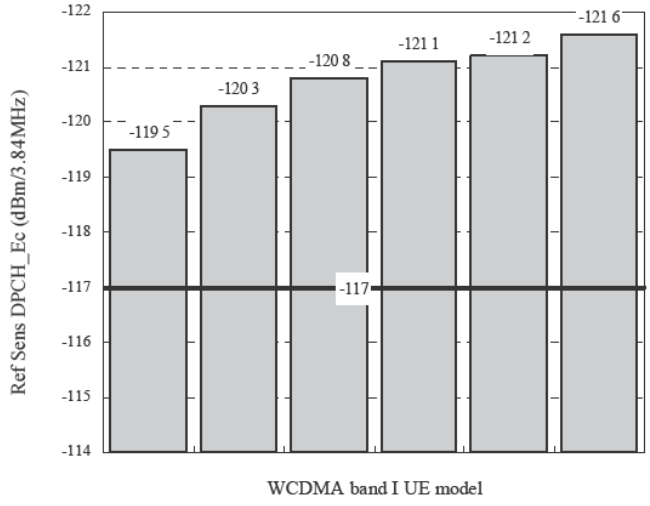


Figure 11.36 Example of class 3 WCDMA band I UE reference sensitivity performance

11.8.2 Introduction to UE Self-desensitization Contributors in FDD UEs

Many of the following items are common to all FDD systems, and in particular, the reader is encouraged to refer to [11] for a detailed discussion on the impact of both TX noise and TX carrier leakage in Direct Conversion Receivers (DCR). This section focuses on the key differences in LTE.

In Figure 11.37, it can be seen that the most sensitive victim is the main receiver LNA, since it is only protected by the duplexer TX to RX port isolation. The diversity LNA benefits from the TX to antenna port duplexer rejection plus the extra antenna to antenna isolation. During conformance tests, this latter is fairly high since the test is performed using coaxial shielded cables, therefore the only coupling mechanisms are those due to PCB cross-talks. Refer to section 11.6.2.2 for details on radiated antenna coupling mechanisms. Both victims operate in the presence of at least two aggressor noise sources: their own PA noise emissions, and the wide-band common mode noise of DigRF^{SM} v4 lines (cf. section 11.6.3).

11.8.2.1 Transmitter Noise Falling into the Receiver Band

Assumptions made in section 11.6.3 are adopted here to illustrate UE self-desense. The victim is a cellular band UE with a 3 dB intrinsic NF and a maximum desense of 0.5 dB is allowed. From section 11.6.3, the maximum aggressor noise PSD must be below −180 dBm/Hz.

11.8.2.2 Large Duplex Distance Frequency Bands

The situation is identical to that experienced in existing WCDMA Band I handsets. Assuming a worst case duplexer isolation in RX of 43 dB, the maximum noise PSD falling in the RX band measured at the PA output port must be less than −180 dBm/Hz + 43 dB = −137 dBm/Hz. Most PAs tested with an ideal signal generator, i.e. a generator which provides a noise floor close to thermal noise at the duplex distance, just barely manage to meet this level [22]. This is one of the reasons why it remains a challenge to design RF IC modulators which can deliver such low noise floors to the PA. The simplest solution is to use an inter-stage BPF, but with the ever-increasing number of bands to be supported, this solution has become unacceptable because of the associated increase of the Bill of Material (BOM). Designing filterless TX RF solutions is a subtle tradeoff exercise between the amount of desense tolerated, the RF modulator current consumption, the BOM cost, and delivering to the customer a competitive reference sensitivity level.

11.8.2.3 Large Transmission Bandwidth in Small Duplex Distance Frequency Bands

Although avoiding receiver desense is not trivial in large DG bands, solutions do exist and this leaves opportunities for innovation. For the small DG, the situation is unfortunately more critical because the aggressor is no longer out-of-band PA noise floor, but the PA ACLR shoulders, as shown in Figure 11.38(a). Therefore, a 3GPP relaxation is the only way to ensure adequate system operation.

An example of ACLR measurements performed using a commercial WCDMA band I PA with a 10 MHz (N_{RB} = 50) QPSK modulated carrier is illustrated in Figure 11.38(b). UMTS 700 MHz front-end IL are emulated with the insertion of a 3 dB resistive attenuator at the PA output port[10].

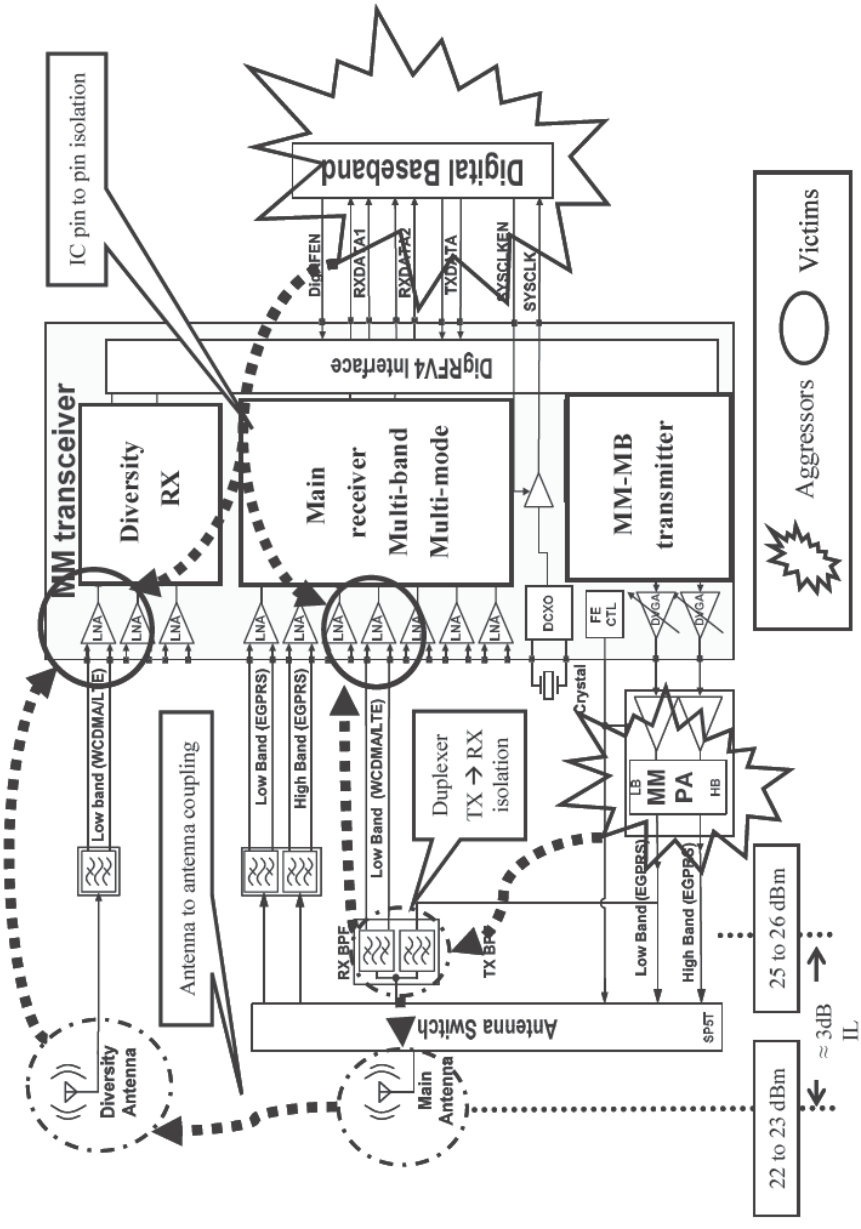


Figure 11.37 Example of aggressors and victims in an optimized quad band 2G – mono band 3G/LTE RF sub-system. Coupling mechanisms are shown in dashed lines. DCXO, digital crystal oscillator

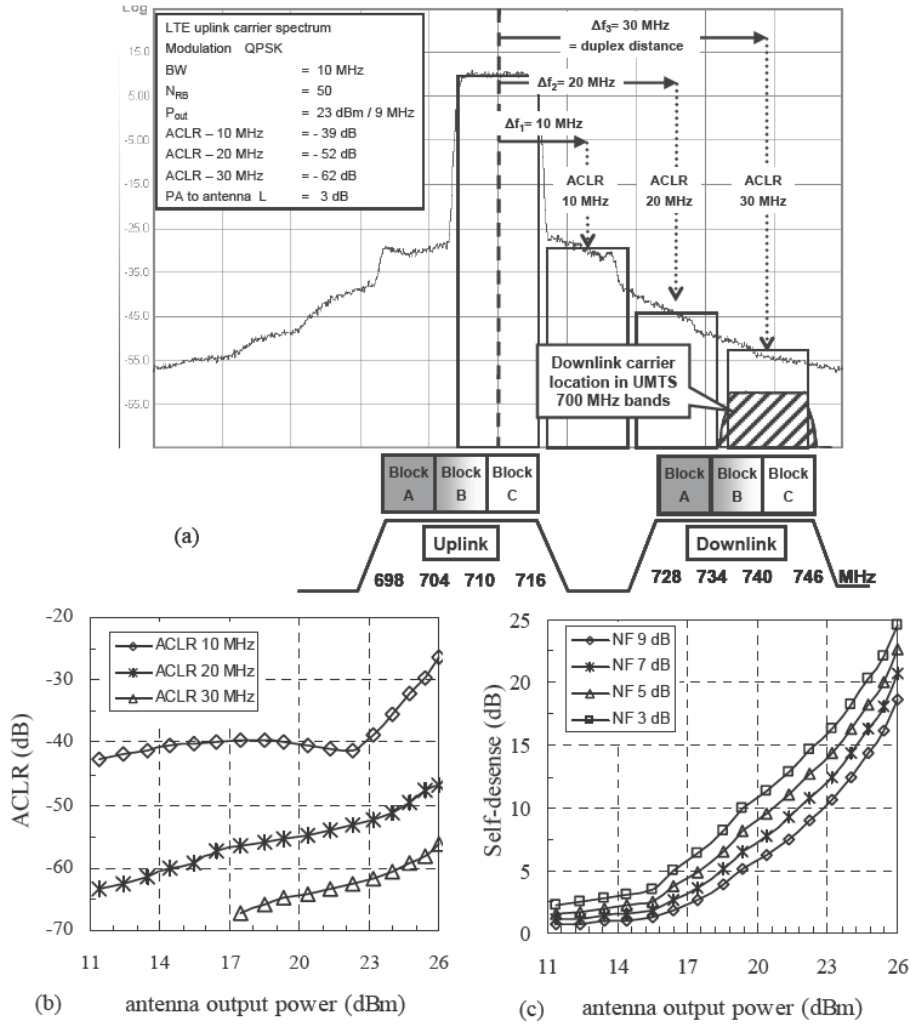


Figure 11.38 Example of LTE 10 MHz QPSK uplink ACLR overlapping the receiver band in the lower UMTS 700 MHz band. (a) ACLR spectral plot; (b) ACLR measured at 10, 20 and 30 MHz offset against antenna output power; (c) self desensitization vs. LNA input referred NF[1]

As can be seen from Figure 11.38(c), at 23 dBm output power, the desensitization reaches in a band 12 example, 16 and 10 dB for 3 and 9 dB NF respectively. To solve this issue, two mitigation techniques have been proposed in RAN4:

[1] Measurements in Figure 11.38 assume a Band XII application: minimum duplexer TX – RX isolation of 43dB, maximum PA to antenna port insertion loss of 3 dB (preliminary measurements on prototypes indicate 2.5 dB maximum in duplexer, 0.5 dB in antenna switch), 23dBm output power at antenna port.

- Maximum Sensitivity Degradation (MSD) [23]: 'the victim's relaxation' technique, which consists of relaxing the reference sensitivity level by an amount similar to those of Figure 11.38(c). The proposal maintains the UE at maximum output power (P_{outmax}) to pass conformance test.
- Point B approach: 'the aggressor relaxation' technique [24], in which the reference sensitivity level is kept intact. This technique maintains the UE at P_{outmax} for a number Resource Block (RB) limited by a point called 'B'. Then, for N_{RB} > point 'B', a progressive back-off of the UE output power is allowed to prevent UE self-desense. Thus, point 'B' corresponds to the maximum number of RBs at which P_{outmax} can be maintained, while point A corresponds to an output power back-off 'X' at which the maximum number of RBs can be supported, as shown in Figure 11.39.

At the time of writing, the MSD approach is adopted. Initial MSD values are proposed for certain bands [26]. Finally, it is worth noting that Half Duplex (HD)-FDD operation has been accepted in RAN 1 [27]. HD-FDD is an alternative solution to the self-interference problem since the transmitter and the receiver do not operate simultaneously. In HD-FDD, the duplexer is no longer needed. This mode of operation can significantly simplify the UE RF front-end architecture as shown in [28].

11.8.2.4 Impact of Transmitter Carrier Leakage Presence at the Receiver LNA Input

In DCRs, differential structure imbalances and self-mixing are known as some of the mechanisms generating second order intermodulation distortion (IMD_2) products [29]. Self-mixing is due to finite isolation between the RF and the LO port of the down-conversion mixer. Under these conditions, the mixer behavior may be approximated as that of a squarer and therefore generates IMD_2 products. The squaring of CW blockers is a simple DC term which can be easily rejected using High Pass Filters (HPF). Squaring an AM modulated blocker, however, generates a wideband noise like an IMD_2 product, which can degrade the wanted signal SNR. In a RX filter-less architecture, the mobile's own TX leakage places the most stringent requirements on the mixer IIp2, which must receive a weak input signal (≈ −85 dBm), in the presence of a

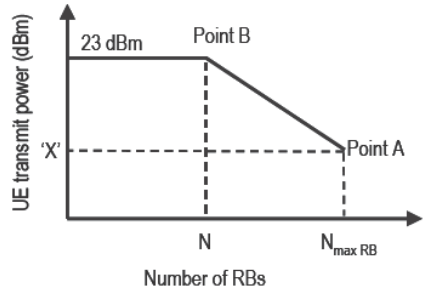


Figure 11.39 Point B approach to prevent UE self-desense

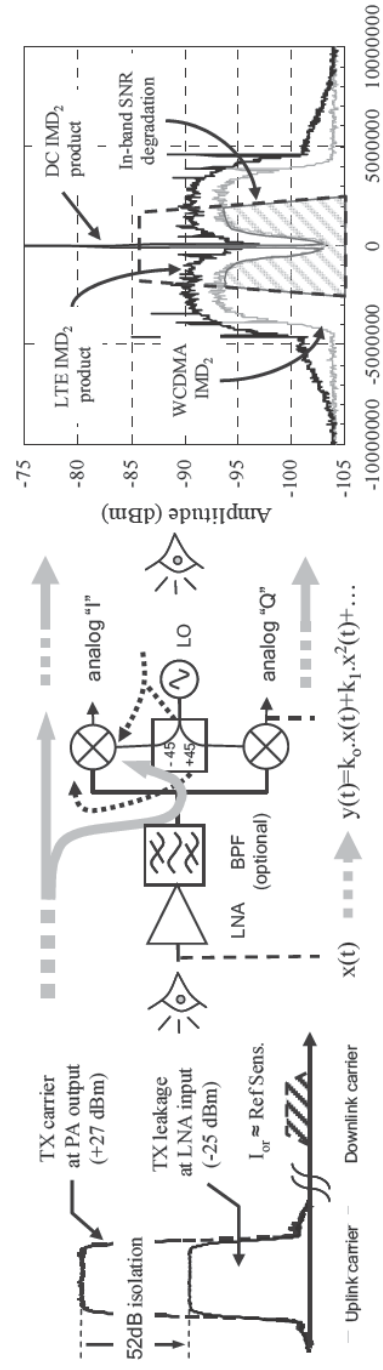


Figure 11.40 Self mixing in direct conversion receivers: Left: TX leakage at LNA input; right: I/Q spectrum observed at mixer output; dashed lines represent the wanted channel

TX leakage mean power of approximately[2] −10.5 dBm at mixer input. The simplest solution consists of using an inter-stage BPF, but this is not the preferred option for the reasons explained in section 11.6.2. A comparison between WCDMA and LTE QPSK uplink modulated carrier IMD_2 products is shown in Figure 11.40.

In a WCDMA RX filter-less application, the mixer IIp2 requirement to ensure a small SNR degradation is in the range of 70 dBm [11], a figure that is extremely challenging to achieve. From Figure 11.40, it can be seen that the higher LTE IMD_2 PSD sets slightly higher mixer IIp2 requirements than for WCDMA receivers.

11.8.3 ACS, Narrowband Blockers and ADC Design Challenges

Both ACS and Narrow-band (NB) blocking requirements are a measure of the receiver ability to receive a wanted signal at its assigned channel frequency in presence of an interfering Adjacent Channel Interferer (ACI), at which a throughput requirement will be fulfilled. The throughput will equal or exceed 95% of the maximum throughput, for a specified reference measurement channel.

The intention of this requirement is to verify the ACI Rejection (ACIR). Both tests are important to avoid UE dropped calls in cells where eNodeBs from adjacent operators are non co-located. In a similar fashion to the WCDMA specifications [43], the LTE requirements are based on a 33 dB ACS budget, which has been derived through extensive coexistence simulations [4]. In order to prevent stringent selectivity requirements for the 15 and 20 MHz BW, a 3 dB and 6 dB ACS relaxation is proposed respectively. The resulting test cases are summarized in Table 11.20[3].

Table 11.20 Relationship between interfering and wanted signals for ACS requirements

Rx Parameter	Units	Channel bandwidth					
		1.4 MHz	3 MHz	5 MHz	10 MHz	15 MHz	20 MHz
ACS test case I							
Wanted signal mean power	dBm	Refsens + 14 dB					
$P_{Interferer}$	dBm	Refsens + 45.5 dB				Refsens + 42.5 dB	Refsens + 39.5 dB
$BW_{Interferer}$	MHz	1.4	3	5	5	5	5
$F_{Interferer}$ (offset)	MHz	1.4	3	5	7.5	10	12.5
Assumed ACS	dB	33	33	33	33	30	27
ACS test case II							
Wanted signal mean power	dBm	–56.5	–56.5	–56.5	–56.5	–53.5	–50.5
$P_{Interferer}$	dBm	–25					

[2] Transmit leakage mean input power at mixer input ≈ PA output power (+27dBm) – isolator/coupler losses (0.5dB) – duplexer isolation (52dB) + LNA gain (15 dB) = –10.5 dBm.

[3] For ACS test case I, the transmitter will be set to 4 dB below the supported maximum output power. For ACS test case II, the transmitter will be set to 24 dB below the supported maximum output power. At the time of publication, there are ongoing discussions to add a small frequency offset to the $F_{Interferer}$ (offset) value to prevent the interfering signal from being orthogonal to the wanted signal [44].

- The ACS test case I is performed with a mean signal power set to 14 dB above the reference sensitivity level, and a variable interferer power for each wanted channel bandwidth as shown in Figure 11.41.
- The ACS test case II is the test which stresses the UE dynamic range by setting the interferer power constant to −25 dBm and variable interferer power so that the 33 dB ACS test condition is met. For example, in a 5 MHz BW (N_{RB} = 25), the Carrier to Interferer power Ratio (CIR) is also equal to −56.5 dBm – (−25 dBm) = −31.5 dB.

One of the UE DCR blocks affected by the ACS II is the LNA, for which the gain must be sufficiently low to prevent overloading of the I/Q mixer inputs, and thereby relaxing the mixer linearity requirements, but also sufficiently high to prevent the UE NF from failing the minimum SNR requirements imposed by the highest MCS test requirements. Additionally, the presence of such a strong interferer, for which the PAPR in LTE is higher than that for WCDMA, sets a linearity requirement on the LNA. LNA non-linearities may generate Adjacent Channel Leakage (ACL), which would overlap the wanted signal and thereby directly degrade the wanted carrier SNR. Figure 11.42 illustrates this challenge.

The NB blocking test is illustrated in Figure 11.43 for 5 MHz BW (N_{RB} = 25). This test ensures LTE UEs can be deployed in regions where other telecommunication standards, such as GSM/EDGE, are in service. To minimize the cellular capacity loss, it is important to minimize the guard bands. For this reason, the NB blocker test places the UE wanted channel in the presence of a blocker located at small frequency offsets. This test only differs slightly from its WCDMA equivalent: in LTE, the blocker is CW (vs GMSK in 3G), the frequency offset[4]

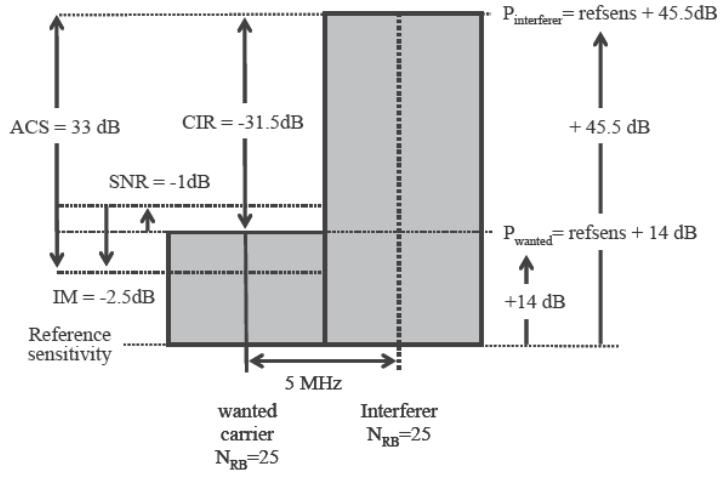


Figure 11.41 Adjacent channel selectivity test case I requirements for 5 MHz LTE

[4] In LTE, the introduction of a small frequency offset has been proposed to ensure that the interferer does not fall in the spectral nulls of the receiver's FFT operation. The offset is an odd multiple of half the tone spacing (2k+1).7.5 kHz, refer to Figure 11.42.

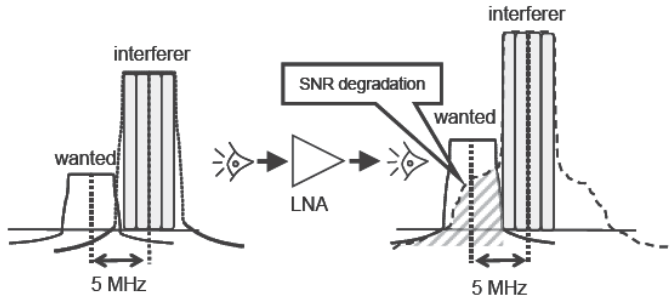


Figure 11.42 Example of LNA non-linearities in the presence of the -25 dBm ACS II interferer. Left: spectrum at LNA input; right: spectrum at the LNA output in the presence of LNA ACL

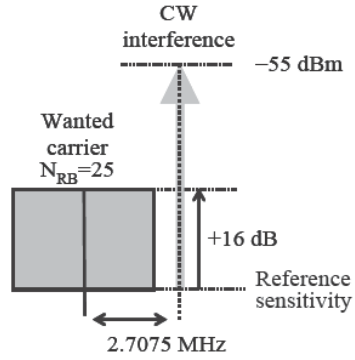


Figure 11.43 Narrowband blocking measurement for 5 MHz LTE

is set to 2.7075 MHz (vs 2.7 MHz in 3G) and the wanted channel benefits from 16 dB desense in 5 MHz BW (vs 10 dB in 3G).

It can be seen that in 5 MHz BW operation, LTE filtering requirements are similar to the existing WCDMA devices. In flexible bandwidth systems the filter design strategy is a compromise articulated around two extreme scenarios: at one end, a receiver with infinite ACIR minimizes the ADC resolution and power consumption at the expense of a sharp analog filter, which may distort the wanted signal. At the other extreme, a receiver which provides no or little ACIR imposes an ADC with a *Dynamic Range (DR)* large enough to cover the worst case 3GPP DR requirements. The following discussion highlights this tradeoff by firstly introducing the impact of bandwidth flexibility on *Analog Channel Filters (ACF)*, and secondly on the ADC.

11.8.3.1 Impact of Flexible Bandwidth Requirements on the Analog Channel Filter Design Strategy

OFDM systems overcome Inter-Symbol and Inter-Carrier Interference (ISI and ICI) due to propagation through time-dispersive channels by introducing a Cyclic-Prefix (CP). The

CP acts in a similar fashion to time guard bands between successive OFDM symbols. Therefore, the longer the CP, the better the resilience to large delay spreads at the expense of an energy loss. The CP length must also be selected so as to avoid the signal smearing caused by the Group Delay Distortion (GDD) of analog filters [30]. Yet, selecting a filter family for the worst case delay spread foreseen in the standard (such as the ETU model, for example) is perhaps not the best strategy. For example, [31] suggests that in most cases the delay spread experienced by the UE is less than the CP length, and therefore the estimation of delay spread by the BB channel estimator can be used to dynamically re-program the transfer function of a post ADC digital FIR filter. This elegant solution provides enhanced ACS performance for a given pre-ADC ACF. So, what is the limit to scaling the 3 dB cut-off frequency 'F_c' of the ACF?

Figure 11.44 illustrates the impact of scaling a baseline[5] filter's F_c optimized for WCDMA ACS and NB blocking onto the experienced GDD. The filter's F_c is either stretched or shrunk by a factor proportional to the LTE BW ratios.

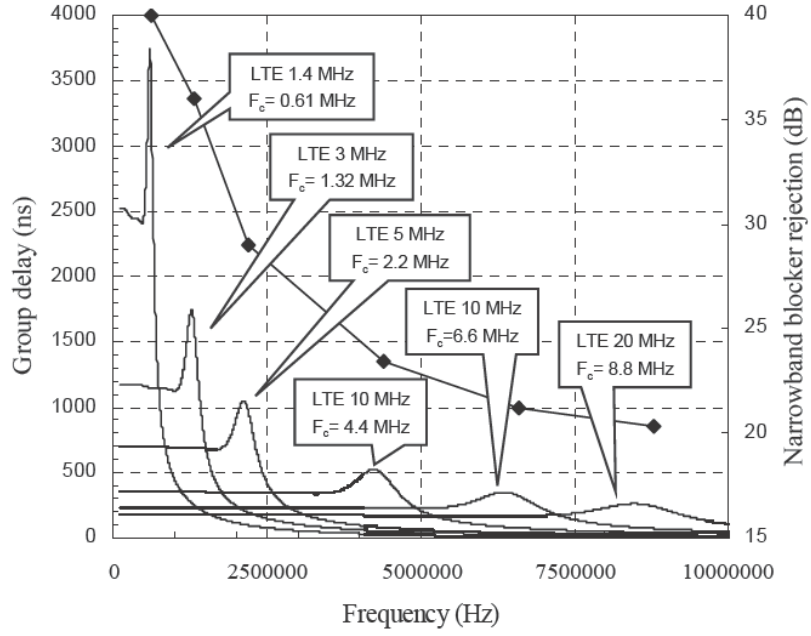


Figure 11.44 Impact of scaling 'Fc' of a 5 MHz baseline ACF optimized for WCDMA proportional to the LTE's operating BW. Left y-axis: filter's group delay; right y-axis: NB blocker filter's rejection performance (diamonds)

[5]The filter used is similar to that in [31]: it has 4 zeros, 8 poles, with a 3dB cut-off frequency "F_c" of 2.2 MHz in 5 MHz BW, and a notch located at 2.8 MHz offset. To reduce the group delay overshoot associated with the presence of a notch so close to the edge of the wanted signal, some of the poles and zeros are used to implement an analog group delay equalizer.

From Figure 11.44, it can be seen that the lower the filter's 'F_c', the higher the filter's delay, GDD, and NB blocker ACIR, so much so that in the 1.4 MHz BW of operation the GDD is slightly greater than 1 µs. This latter case 'consumes' a large amount of the normal CP length of 4.7 µs and could impact ISI in large delay spread channels. From this example, it can be concluded that proportional scaling of the ACF's F_c is probably not the best tradeoff. One alternative takes advantage of the 3GPP relaxations to tailor the ACF ACIR to just meet the ADC DR requirements at 15 and 20 MHz operation. The ADC DR enhancements in small operating BW can then be used to relax the filter's sharpness so as to take full benefit of the CP length.

11.8.3.2 Impact of Flexible Bandwidth Requirements on the ADC DR

For the sake of simplicity[6], the minimum required ADC resolution is estimated by assessing the ADC EVM (EVM_{ADC}) based upon the pseudo-quantization noise (PQN) model. Figure 11.45 (left) shows the optimum Signal to Distortion Quantization Noise Ratio (SDQNR) at the ADC output is met for ADC clipping ratios, also denoted 'ADC back-off' (ADC BO), ranging from 10 to 14 dB. The resulting EVM_{ADC} is plotted in Figure 11.45 (right).

The UE EVM budget is estimated from the required SNR to achieve the highest SIMO LTE throughput which corresponds to the 64QAM 8/9 MCS. From section 11.3.4, 5% throughput loss is met if the total composite EVM is less than 6.3%. Assuming that each EVM impairment

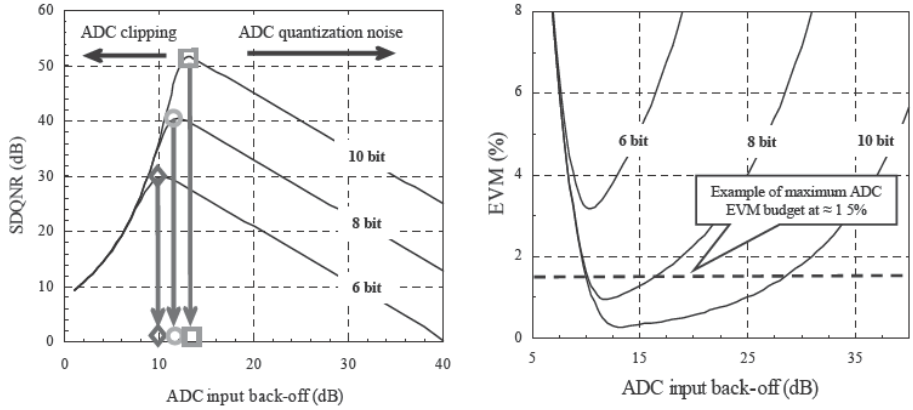


Figure 11.45 Left: SDQNR (in dB) at ADC output against ADC resolution and ADC input back-off (in dB) or ADC clipping ratio. The optimal ADC BO is respectively highlighted with a diamond, circle and square for 6, 8 and 10 bits ADC resolution. Right: corresponding ADC output EVM performance

[6]The PQN model considers the quantization error as an additive noise, which has a white and a uniformly distributed spectrum and which is uncorrelated with the input signal. The reality is more subtle: the recent work published in [32] shows that the PQN approach is not entirely sufficient to model the quantization errors due to the much larger peak to average power ratio of OFDM signals.

is AWGN like, taking an example where the eNodeB EVM is equal to 4.5%, and the UE RF RX EVM performance is equal to 4%, this leaves an EVM_{ADC} budget of 1.5%[7].

Let us first assume an ideal RF–BB chain, which provides an infinite ACIR, and an ideal Analog Gain Control (AGC) loop, so that the optimal BO is always exactly met. The situation is captured in Figure 11.46 (left) and indicates that the lowest acceptable ADC resolution is 8 bit. In a real AGC loop system, the ADC BO over the duration of a call is no longer that of a discrete point, but a spread of BO statistically distributed as shown by the histogram in Figure 11.46. Taking one particular example of AGC loop inaccuracies[8] into account, it can be seen that 10 bit is the minimum ADC resolution, which provides approximately 12 dB headroom (Δ_{RF}) for RF IC imperfections, such as imperfect DC offset cancellation and ACIR (Figure 11.46 right). This requirement is equivalent to a CW DR of 60 dB. One of the differences in LTE is that the UE AGC loop must also cope with time-varying amplitude of in-band signals due to dynamically scheduled users with a varying number of active RBs transmitted at different power levels in the downlink. With 10-bit resolution, Δ_{RF} is sufficiently large to

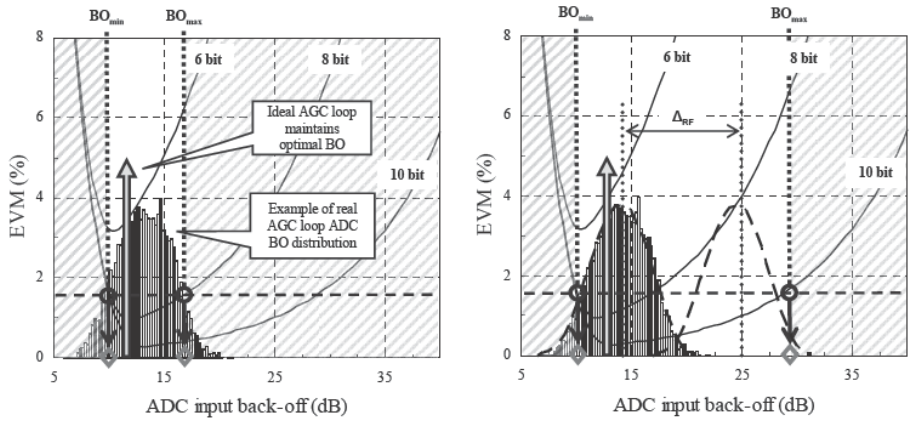


Figure 11.46 Impact of AGC loop imperfections. Left: 8 bit ADC; right: 10 bit ADC. Dashed areas indicate regions which do not fulfill the MCS EVM ADC requirements. The dashed histogram envelope illustrates the ADC BO margin due to RF imperfections (Δ_{RF}) present at ADC input

[7] Assuming AWGN like behavior of each EVM contributor, $EVM_{ADC} = \sqrt{6.3\%^2 - (EVM^2_{eNobeB} + EVM^2_{RFRX})}$. Note that in conformance tests, the eNodeB emulator EVM is extremely low (typically 1% or less). Also note that the recent WiMax (802.16e) RF transceiver design in [33] can achieve ≈1.5% RX EVM, thereby relaxing the overall EVM downlink budget.

[8] The histogram shows a recorded ADC BO distribution captured over a 10 minute long WCDMA BLER measurement performed in a fading test case-1, 3 km/h. The AGC loop updates the analog gain at 10 ms intervals. An AGC loop for LTE is likely to operate at a faster update rate to provide better control accuracy of the ADC BO. The resulting histogram would present a smaller spread of BO, thereby relaxing the margins for RF imperfections.

accommodate one example of RF imperfection margins[9]. Note that each of the assumptions[9] is listed as an example of minimum requirements since these are implementation specific.

In conclusion, the most challenging mode of operation for a multiple-standard ADC is the LTE 20 MHz operation, for which a minimum of 60 dB DR must be delivered. Sigma delta ADCs represent an attractive solution to these requirements [34]. These converters shape the Quantization Noise (QN) floor via a high pass transfer function, thereby digging a notch into the QN PSD within the BW of the wanted carrier, and rejecting the noise out-of-band. The BW of the QN notch can be optimized for each BW of operation by either changing the sampling frequency and/or by reconfiguring the noise shaping filter's transfer function. An example of sigma-delta ADC flexibility is shown in Figure 11.47. It can be seen that by ensuring the DR requirements are met for LTE 20 MHz, the ADC offers a 20 dB improvement for operation at LTE 1.4 MHz. Every decibel gained in the DR can be used as an extra relaxation of the ACF filter design. In particular, the large DR performance at low operating BW relaxes significantly the ACF rejection in GSM mode for which the design of a sharp filter can be expensive in terms of die area.

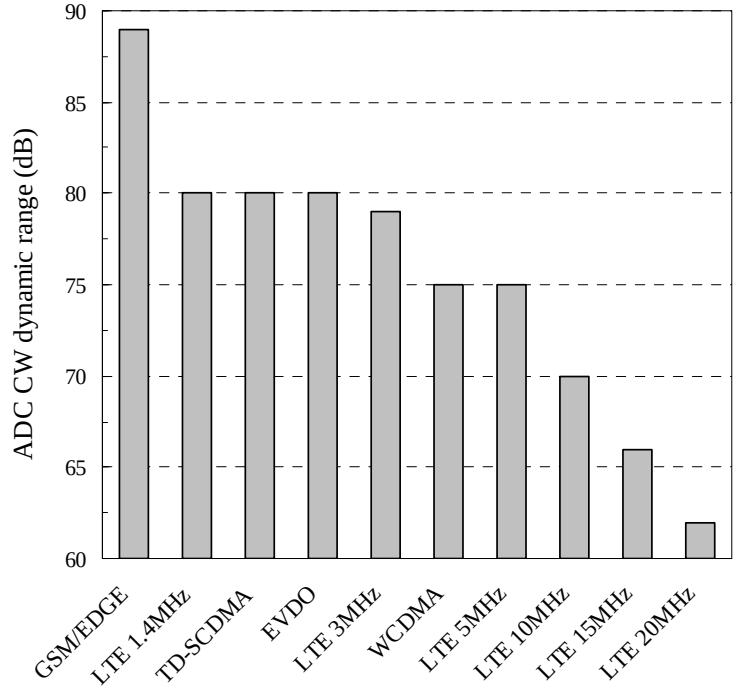


Figure 11.47 Achievable DR with sigma delta ADCs over a wide range of system's BW [35]

9 Following margins are assumed: 1) 2 dB margin for imperfect RF receiver DC offset removal at ADC input, 2) 4 dB to account for relative power difference between a 64QAM user and a QPSK user in the cell [2] - this is a minimum requirement, 3) Imperfect RF RX ACIR forces the AGC loop to lower the effective wanted signal target total ADC BO. With the previous assumptions 1 & 2, the DCR is allowed to present at the ADC input a CIR of ≈4 dB (eNodeB TPC DR) + 2 dB (leaked DC offset) – 12 dB (Δ_{RF}) = –6 dB.

11.8.4 EVM Contributors: A Comparison Between LTE and WCDMA Receivers

In section 11.3.4, the downlink LTE budget is set to approximately 8%. Compared to WCDMA, where 10–12% is acceptable even for HSDPA operation, this requirement appears tougher. However, the novelty in LTE is that EVM measurements make use of a zero-forcing equalizer (cf. section 11.3.4). Thus, a distinction must be made between AWGN like contributors, and contributors that can be equalized, and therefore become transparent to the UE receiver. This is an important difference with WCDMA where, for large enough SF, each EVM contributor behaves like AWGN [36]. This section aims at illustrating these differences with only a few selected impairments: I/Q gain and phase imbalance, distortions due to ACF, and DCR local oscillator phase noise.

11.8.4.1 Impact of Finite Image Rejection due to I/Q Amplitude and Phase Mismatches

In real world analog/RF designs, it is nearly impossible to design DCRs with perfectly identical gain and phase response between I and Q branches. Therefore DCRs come with two natural impairments: amplitude and phase mismatches, denoted ΔA and $\Delta\Phi$[10] respectively, leading to a finite Image Rejection (IR). Finite IR results in each sub-carrier being overlapped with the sub-carrier which is located at its frequency image mirror, as shown in Figure 11.48[11]. The power ratio between a sub-carrier and its image is by definition the IR. Assuming the symbols carried by each sub-carrier are uncorrelated, the impact of IR on LTE is no different to that of a single carrier system and can be considered an AWGN source [37].

Figure 11.49 shows EVM measurements performed with an Agilent™ 89600 Vector Signal Analyzer, which delivers the Error Vector Spectrum (EVS). EVS is a tool which plots the magnitude of the error vector against each sub-carrier. Figure 11.49(b) shows the composite LTE EVS, i.e. it is an overlay of each physical channel's EVS, where the darker line shows the

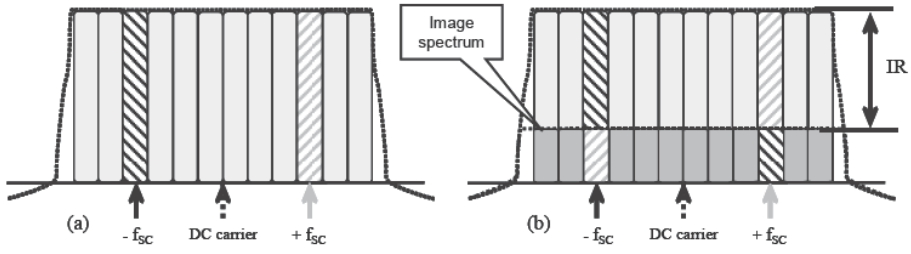


Figure 11.48 (a) Transmitted OFDM carrier; (b) equivalent i+jQ baseband complex spectrum at ADC input in presence of a finite image rejection equal across all sub-carriers

[10] $\Delta\Phi$ may originate from either a local oscillator which does not drive each I and Q mixer in exact quadrature, or from a tiny mismatch in the 3 dB cut-off frequency of each I/Q LPF. In the latter case, $\Delta\Phi$ is not identical for all sub-carriers. The net result is a frequency dependent image rejection (IR). Note that IR can be calibrated in modern DCR designs as shown in [33].

[11] In field trials, incoming sub-carriers are notched due to frequency selective fading and image rejection is not constant across the entire BW of the DCR I/Q bandwidth.

Figure 11.49 Impact of $\Delta\Phi$ and ΔA on WCDMA and OFDM downlink carriers. Impact of $\Delta A = 1.75$ dB on: (a) LTE 64QAM state spreading; (b) LTE carrier EVS; (d) 16QAM HS-DSCH WCDMA; (e) WCDMA EVM spectrum. (c) and (f): comparison of LTE (filled circles) and WCDMA (empty diamonds) EVM performance against ΔA and $\Delta\Phi$

average EVS. The EVS spectral flatness in Figure 11.49 confirms the AWGN-like behavior of EVM due to IR for both standards.

11.8.4.2 Impact of In-band Analog Filter Amplitude and Group Delay Distortions

Zero-IF Low Pass Filter Contribution

Figure 11.50 shows the measured[12] impact of a prototype I/Q channel filter[13] similar to that presented in [31] on a 5 MHz 16QAM LTE and a WCDMA downlink carrier. The equal spreading of each constellation point in Figure 11.50(c) confirms the AWGN-like behavior of EVM_{LPF} for WCDMA, and in this example, results in 8% EVM performance. Figure 11.50(b) shows that without equalization the outermost sub-carriers are severely impacted, while sub-carriers located in the middle are less vulnerable. The use of the ZF-equalizer flattens EVS and brings the composite EVM down to ≈1.2%.

It can be concluded that LTE relaxes the LPF impairment budget compared to WCDMA modulation.

Zero-IF High Pass Filter Contribution

IMD_2 products described in section 11.8.1.2 generate a strong DC term which can lead to saturation through the several decibels of I/Q amplification required to meet the ADC target BO. In WCDMA the DC term can be cancelled with High Pass Filters (HPF), and [36] has shown that its impact on EVM is AWGN-like. HPF design is a compromise between EVM, capacitor size, die area, and DC settling time[14]. A passive 4.5 kHz first order HPF with group delay distortion in excess of the CP length has been deliberately chosen to illustrate the impact on LTE. In contrast to the LPF or BPF test case, the sub-carriers located close to the center of the carrier are the most vulnerable as can be seen in Figure 11.51. For example, the Primary Synchronization Signal (PSS), which occupies the center of the carrier over a 1.4 MHz BW, experiences a 7.5% EVM, while the QPSK user EVM improves as the allocated BW is increased (Figure 11.51(a) and (c)). The distortion impacts so few RS that the ZF-equalizer cannot flatten the impact of the HPF. LTE therefore calls for a careful design of the DC offset compensation scheme.

11.8.4.3 Impact of Phase Locked Loop Phase Noise

If the downlink OFDM signal was just a set of unmodulated closely spaced CW tones, the resulting I or Q output of the DCR mixer would be that of each CW tone multiplied with LO Phase Noise (PN) profile as shown in Figure 11.52(a). Clearly, any PN exceeding the sub-carrier spacing causes SNR degradation. In most PN studies [38, 39], a distinction is made

[12] All EVM measurements performed with an Agilent™ 89600 VSA.

[13] Both I and Q filters are nearly perfectly matched and deliver an IR across the entire receiver bandwidth better than –40 dBc. This guarantees that the EVM performance is dominated by the LPF non-linear phase and amplitude distortions.

[14] In BiCMOS designs, it is difficult to implement analog HPF with cut-off frequencies less than 3 kHz due to cost constraints in die area. The use of RFCMOS allows designing this loop partly in the digital domain with a much smaller cut-off frequency at nearly no impact on die area.

Figure 11.50 Measured impact of an I/Q ACF on the composite EVM of a 5 MHz LTE (N_{RB} = 25) and a WCDMA downlink carrier. (a) LTE composite EVS and 16QAM user data constellation with equalizer 'ON'; (b) LTE reference signals EVS and constellation with equalizer 'OFF'; (c) WCDMA HS-DPCH constellation, EVM ≈8%

Figure 11.51 Measured[12] impact of a 4.5 kHz HPF prototype with ZF equalizer 'ON'. (a) EVS of PSS in 5 MHz (25 RB); (b) EVS of a 5 MHz (25 RB) QPSK user; (c) EVM vs carrier's BW. Stars = average PSS EVM; diamonds = average composite EVM; squares = QPSK user average EVM

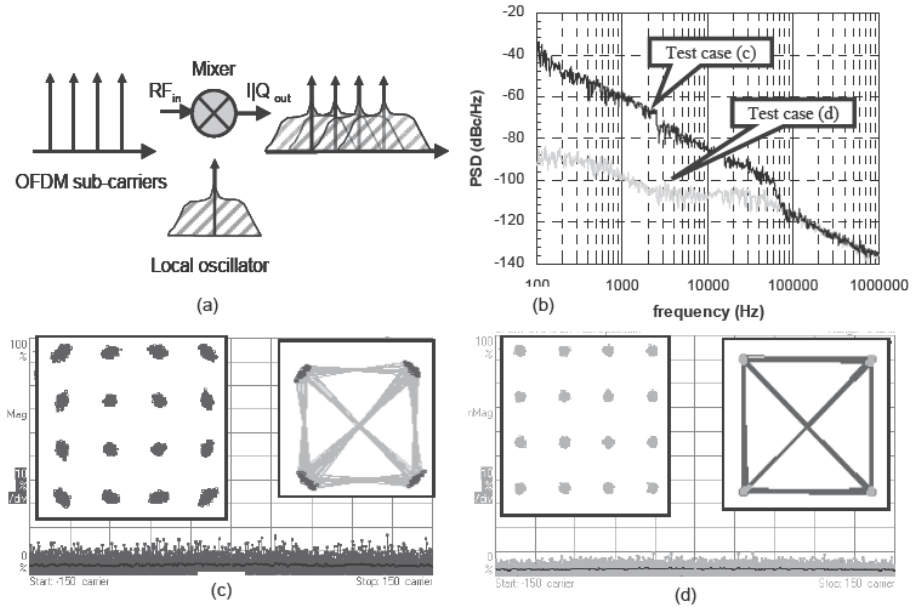


Figure 11.52 (a) Illustration of LO PN multiplication on OFDM sub-carriers; (b) LO PN profiles: degraded PN in black ('test case "c"), 'near ideal' in grey (test case 'd'); (c) and (d) LTE 16QAM and WCDMA QPSK constellations are overlaid on the LTE EVS; (c): easurements with degraded PN; (d) measurements in near ideal PN profile

between PN contributions at high and low frequency offsets from the carrier. The close-in PN produces an identical rotation of all sub-carriers and is also often referred to Common Phase Error (CPE). CPE can be estimated and therefore can be corrected for. The high frequency offset PN components generate ICI. This contribution can be modeled as AWGN and cannot be corrected. An illustration of both CPE and ICI due to a degraded PN profile is presented in Figure 11.52(c) where the LTE 16QAM user data suffer from both constellation rotation and state spreading.

In conclusion, the OFDM sub-carrier spacing used in LTE introduces a new LO phase noise requirement for frequency offsets ≥ 15 kHz offset. This differs from WCDMA where EVM for this impairment only depends on the integrated LO phase error.

11.9 UE Demodulation Performance

11.9.1 Transmission Modes

OFDM modulation in LTE downlink allows optimization of the transmission in time, frequency and antenna domains. The UE 3GPP standard compliancy must be ensured by satisfying the requirements covering a wide range of modes comprising Transmit/Receive diversity and spatial multiplexing. We examine in this section the principles of each mode and the corresponding receiver schemes for each case.

11.9.1.1 Single Input Multiple Output (SIMO) Mode

Plain OFDM Single Input Single Output (SISO) transmission is not supported by LTE as UEs are required to have at least two receiving antennas. The SIMO case constitutes then the simplest scheme and it is a straightforward extension of the SISO case. Let us assume a transmission bandwidth of B = 20 MHz and a sampling frequency of f_s = 30.72 MHz and normal cyclic prefix mode. At the receiver side, for each receiving antenna, base-band digital signal samples after RF and analog-to-digital conversion are buffered over an OFDM symbol of duration Ts = 71.3 μs (or 71.9 μs for the first OFDM symbol of a slot since the cyclic prefix samples are then slightly more). The cyclic prefix samples are discarded, assuming timing offset recovery is safely done, and N = 2048 samples are FFT transformed into a frequency domain equivalent signal. Due to LTE standard assumptions, at the output of the FFT, only K = 1200 samples are kept and the others are discarded. Depending on the physical channel that needs to be modulated and its specific allocation over the sub-carriers, an operation of demultiplexing and re-ordering of the complex digital samples takes place.

Even for multi-path propagation, providing that the channel coherence time is much larger than the OFDM symbol duration, it is well known that in OFDM the signal at the output of the FFT at each sub-carrier position can be considered as affected by flat fading. In this case the optimal demodulation scheme is the simple Matched Filter and the effect of the channel is undone by multiplication by the complex conjugate of the channel coefficient at any given sub-carrier. The QAM symbols are obtained by combining the derotated symbols on the same sub-carriers across the two receiving antenna paths. This operation is also generally known as Maximum-Ratio-Combining and allows optimal benefit from the additional antenna diversity.

After MRC operation, the QAM symbols are demapped and the sequence of soft-bits for channel decoding operation is buffered until the total expected amount is available after demodulation of several OFDM symbols. The soft-bits are rate-matched for the parameters specific to the physical channel under consideration; eventually an operation of Hybrid-Retransmission-Request (HARQ) combining is performed and the channel decoding operation is invoked. At its output, the sequence of hard-decisions is verified against the transported Cyclic-Redundancy-Check (CRC) bits to decide whether the decoded bits are correct or not. The channel decoder to be used depends on the nature of the decoded physical channel: dedicated channels, e.g. PDSCH, are always turbo-encoded while channels carrying control information, e.g. PDCCH, are convolutionally encoded, and thus decoded by means of a Viterbi decoder.

For channels supporting the HARQ protocol, the result of the redundancy check operation is fed back to the BS. The receiver performance is computed upon the success rate of the CRC in terms of throughput, which is the measure of net successfully decoded bits after the HARQ process.

11.9.1.2 Transmit Diversity Mode

Transmit diversity is implemented in LTE by means of Spatial-Frequency-Block-Coding (SFBC). SFBC is the Alamouti encoding of two QAM symbols lying in neighboring sub-carriers. For transmit-diversity transmissions then, at the receiver side, symbols at the output of the FFT need to be re-ordered in pairs according to the SFBC encoding carried out at the transmitter and the Alamouti scheme is undone via a linear operation.

11.9.1.3 MIMO Mode

MIMO transmission mode is the key enabler of a high data rate of LTE (up to 150 Mbps for 20 MHz bandwidth) and allows the transmission of one or two independent data streams depending on the channel conditions experienced by the UE. In MIMO mode, the channel at each sub-carrier is represented by a matrix whose size is given by the number of transmitting (N_{Tx}) and receiving (N_{Rx}) antennas. If the propagation conditions result from a rich scattering environment, the rank of the channel matrix is full and in these conditions spatial multiplexing of two data streams can be supported. If instead the channel matrix rank is not full, i.e. a rank is equal to one, only one codeword is transmitted. As for HARQ acknowledgements, MIMO closed loop mode requires continuous feedback from the UE to the BS on a sub-frame basis. Together with the channel rank information, the UE also provides the BS with the indexes of the pre-coding codebook vectors to be used at the transmitter side. The closed loop MIMO pre-coding mechanism, at the expense of additional signaling overhead, is the method used in LTE to effectively exploit the MIMO channel diversity. This is because the pre-coding vector indexes, requested by the UE, are chosen such that the SNR is maximized, therefore also maximizing the throughput. The SNR is computed on the overall equivalent channel constituted by the cascade of the pre-coding matrix and the instantaneous propagation channel matrix.

The standard does not mandate for a particular detection scheme but instead assumes a linear Minimum-Mean-Squared-Error (MMSE) equalizer as a baseline detector to establish the minimum performance requirement. The MIMO transceiver scheme is shown in Figure 11.53. The equalizer coefficients are adjusted depending on the channel matrix coefficients, the pre-coding vector and the interference power. It can equivalently be regarded as a 2×2 matrix multiplication of the 2×1 vector constituted by the complex signal at each sub-carrier at the output of the FFT of the two receiving antennas as follows:

$$\underline{\hat{x}}_i = \underline{\underline{G}}_{i\,MMSE}\underline{y}_i = \underline{\underline{G}}_{i\,MMSE}(\underline{\underline{\tilde{H}}}_i\underline{x}_i + \underline{n}_i) = \underline{\underline{G}}_{i\,MMSE}(\underline{\underline{H}}_i\underline{\underline{P}}_i\underline{\hat{x}}_i + \underline{n}_i) \tag{11.15}$$

Where:

- $\underline{\hat{x}}_i$ is the 2×1 detected symbol vector at sub-carrier i;
- $\underline{\underline{G}}_{i\,MMSE}$ is the MMSE equalizer 2×2 matrix at sub-carrier i;
- $\underline{y}_i$ is the 2×1 received signal vector at sub-carrier i;
- $\underline{\underline{\tilde{H}}}_i$ is the 2×2 equivalent channel matrix resulting from the cascade of the 2×2 pre-coding matrix $\underline{\underline{P}}_i$ and the actual 2×2 channel matrix $\underline{\underline{H}}_i$ at sub-carrier i;
- $\underline{n}_i$ is the 2×1 interference vector received signal at sub-carrier i.

11.9.2 Channel Modeling and Estimation

11.9.2.1 3GPP Guidelines for Channel Estimation

The coherent detection schemes mentioned earlier require the availability of a reliable channel estimate for each sub-carrier, for each OFDM symbol and for each link between transmitting and receiving antennas. For this purpose, LTE systems provide a Reference Signal (RS) whose resource elements are disposed in the time–frequency plane in a diamond-shaped uniform grid.

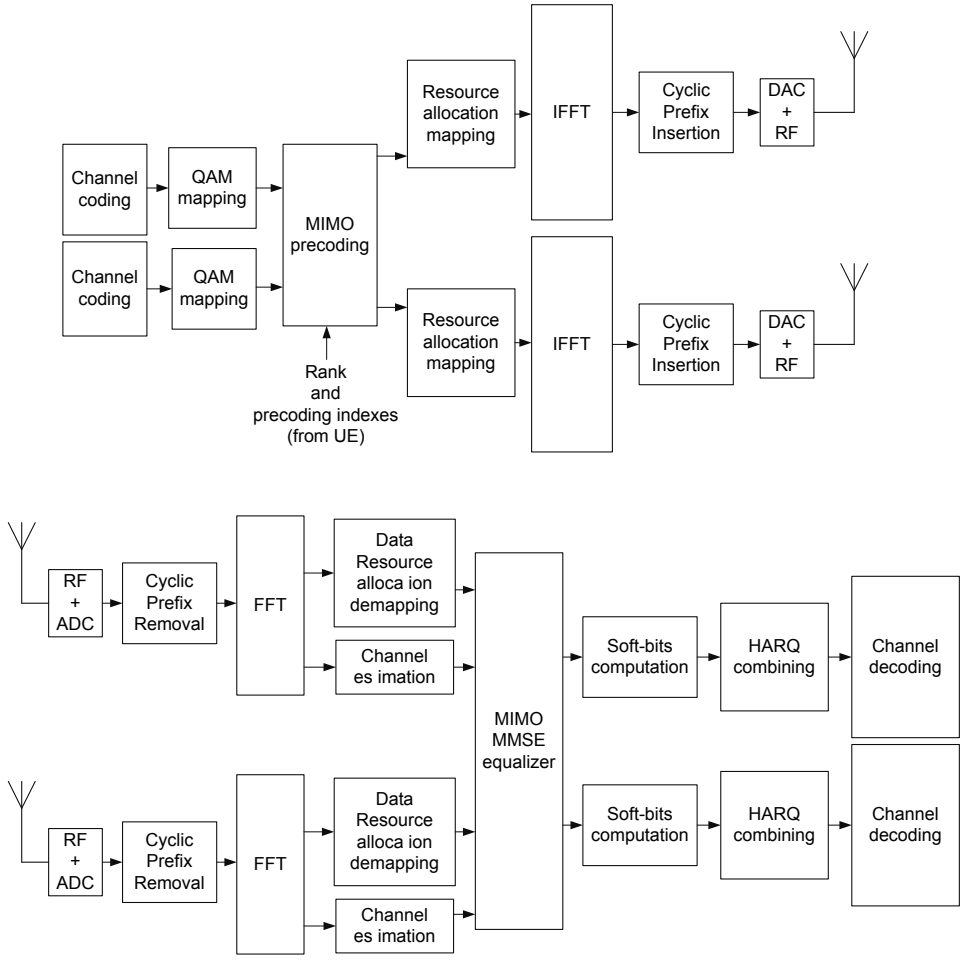


Figure 11.53 MIMO transceiver

For multiple antennas transmissions, RS are interleaved and zeroed out in correspondence to the RS of other antennas to minimize mutual interference. Hence, thanks to this particular structure introduced specifically in LTE, the channel estimation over antennas can be performed independently for each link between each transmitting antenna and each receiving antenna. This pilot signaling scheme considerably simplifies the channel estimation task in MIMO applications.

As for the receiver scheme, the standard gives freedom in the implementation of the frequency–time channel estimation, even if some companies provided practical guidelines along the standardization discussions for the definition of the performance requirements. These guidelines proposed the channel estimation to be implemented as the cascade of two 1D filtering processes: the first dimension consists of Wiener/MMSE filtering of the OFDM symbols containing RS in the frequency direction. The second dimension follows as a Wiener/MMSE

filter in the time domain to obtain the full channel transfer function for all sub-carrier indexes and OFDM symbols within a sub-frame.

Only the reference signals belonging to the current sub-frame are used for time domain interpolation. The coefficients used for the frequency domain filtering process are chosen from a pre-computed set as a function of signal-to-noise ratio only [45].

It is worth noting that since channel estimation is performed after the FFT operation, the actual channel being estimated is indeed the convolution of several filters' impulse responses, amongst which the time variant component is that of the physical air interface propagation channel, and the other filters are either the eNodeB channel filters, or those of the UE RF front end section. In that respect, in-band distortions introduced by RF filters are naturally compensated for as long as the total delay spread does not exceed the CP length.

11.9.3 Demodulation Performance

11.9.3.1 PDSCH Fixed Reference Channels

The performance requirements in [1] include a set of Fixed Reference Channels for each transmission mode – namely SIMO, transmit diversity and MIMO modes. The FRCs so far agreed, for the SIMO case and cell specific RS signals, fall into three categories involving a restricted set of Modulation and Coding Schemes: QPSK with coding rate 1/3, 16QAM with coding rate ½ and 64QAM with coding rate ¾.

The choice of such FRC categories was made so as to reduce to a minimum the number of tests while having representative MCS in the overall set of 29 MCS combinations.

Within each FRC category, one test is specified for a given system bandwidth and therefore characterized by a specific transport block length, code-block segmentation parameters assuming an allocation spanning the entire available bandwidth.

An additional FRC category is also defined for a single resource block allocation happening on the band-edge and making use of QPSK modulation.

11.9.3.2 PDSCH Demodulation Performance Requirements

The performance requirements are expressed as the required SNR $\hat{I}_{or}/N_{oc}$ in dB to obtain a given fraction of the nominal throughput of a given FRC and in given propagation conditions. The fraction of the nominal throughput is always chosen to be 30% or 70%. It is worth noting that for RF section performance tests, the metric has been set to a throughput greater or equal to 95% of the maximum throughput of the reference measurement channel. An example of the performance requirement for 64QAM as stated in the 3GPP standard is presented in Table 11.21.

The required SNR is agreed by consensus among the companies participating to the standard. The value is computed upon the decoding performances of a full-floating point receiver chain where an implementation margin is taken to account for the degradation induced by fixed point or other signal-processing non-idealities. The implementation margin is in the range of 1.5–2.0 dB. The principle just explained for the derivation of the performance requirement is graphically depicted in Figure 11.54.

Table 11.21 Minimum performance 64QAM (FRC)

Bandwidth	Propagation Condition	Correlation Matrix	Reference value	
			Fraction of Maximum Throughput (%)	SNR $\hat{I}_{or}/N_{oc}$ (dB)
10 MHz	EVA5	Low	70	17.7
	ETU70	Low	70	19.0
	EVA5	High	70	19.1

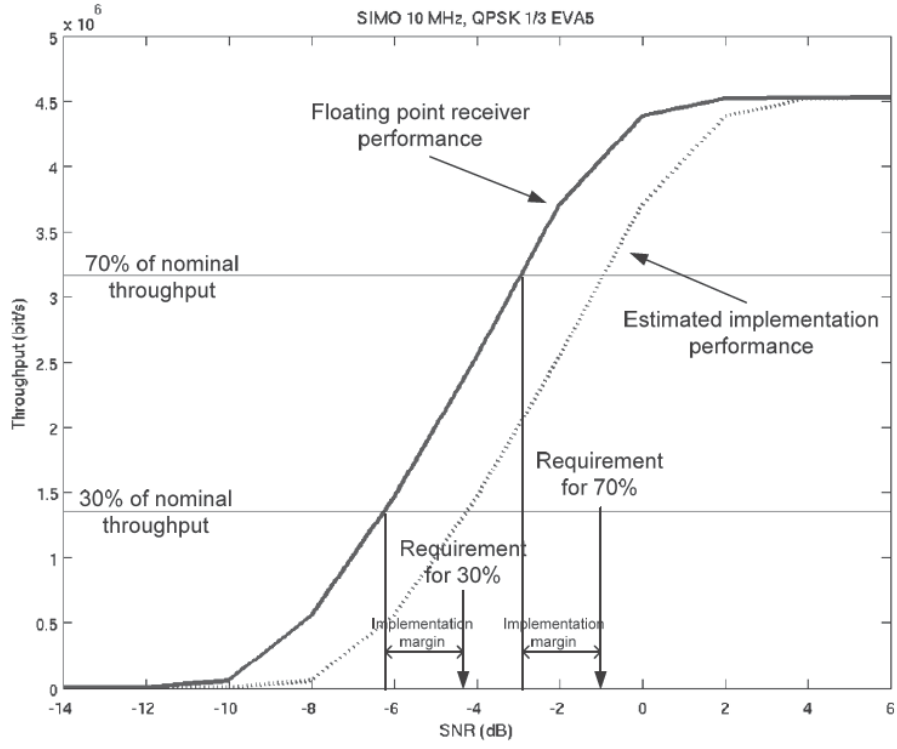


Figure 11.54 Principle of performance requirement derivation

11.9.3.3 EVM Modeling and Impact onto Performance

With the aim of deriving a realistic performance requirement, LTE assumes that the transmitter is modeled as a non-ideal signal source characterized by transmitter Error Vector Magnitude. The non-ideality in general results from non-linearities happening because of OFDM signal high Peak-to-Average Power Ratio and limited dynamic of the RF transmitter amplifier.

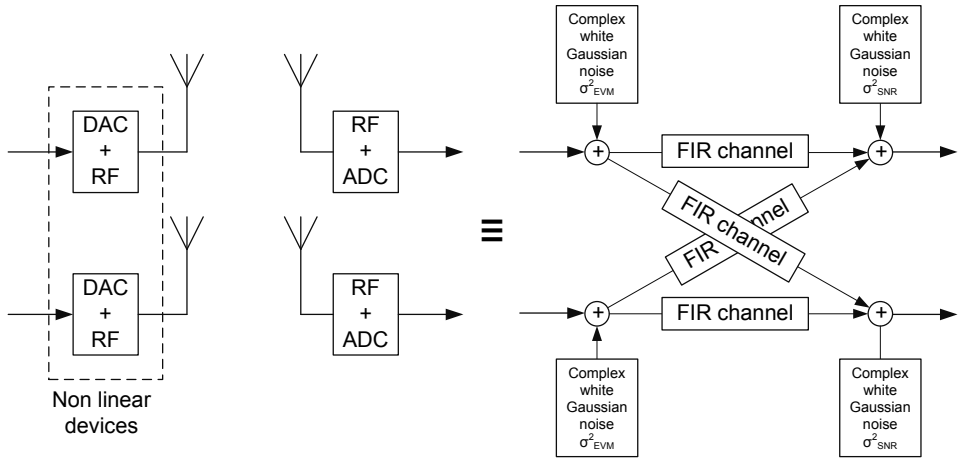


Figure 11.55 Equivalent EVM model used for performance simulations

Table 11.22 Error Vector Magnitude assumption for performance requirements

Modulation	Level	Equivalent per Tx antenna noise variance σ^2_{EVM} (dB)
QPSK	17.5%	–15.13924
16QAM	12.5%	–18.0618
64QAM	8%	–21.9382

Precise descriptions of the EVM modeling for LTE performance requirement derivation were provided, for example, in [47]. The model counts non-linearities as an additive distortion source at the transmitter thanks to the Bussgang theorem [48], as shown in Figure 11.55.

Table 11.22 presents the EVM levels to be assumed in performance simulation. The effect of the additive distortion source is to limit the attainable capacity as the receiver effective SNR cannot increase as other cell interference vanishes because of the irreducible distortion term, as Figure 11.56 shows. As an additional consequence, the MCS set gets limited and upper-bounded as the general system capacity.

11.10 Requirements for Radio Resource Management

The mobility procedures are described in general in Chapter 7. This section focuses on the performance requirements for mobility. The performance requirements for radio resource management (RRM) for E-UTRA are defined in [3] and follow a similar basic structure to those defined for UTRA in [40]. There are requirements supporting mobility in both E-UTRA RRC_Idle, and RRC_Connected states and mobility for E-UTRA intra-frequency, and E-UTRA inter-frequency as well as performance requirements for handover and reselection to other Radio Access Technologies (RATs) including UTRA FDD, UTRA TDD and GSM. Performance is also specified for mobility to non-3GPP RATs such as CDMA2000 1× and High Rate Packet Data (HRPD), also known as CDMA EV-DO.

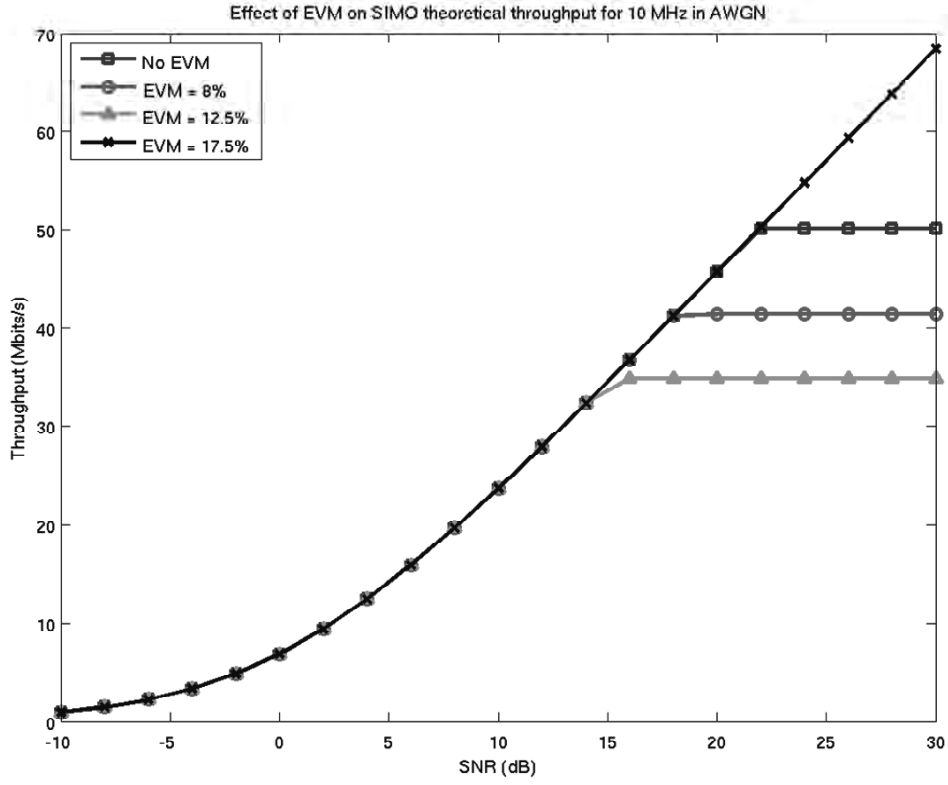


Figure 11.56 Effect of EVM on theoretical SIMO throughput for 10 MHz in AWGN

In RRC_Idle state, mobility is based on UE autonomous reselection procedures, whereas in RRC_Connected state measurement reports are made by the UE, to which the E-UTRA network may respond by issuing handover commands. Due to the autonomous nature of the idle state reselection procedures, and the relative importance of power saving in RRC_Idle state, typical idle performance requirements allow for more implementation freedom, whereas in RRC_Connected state, consistent UE measurement reporting is of prime importance, and so more details of measurement performance are specified to ensure that network based mobility algorithms can work with all UEs. One important difference in E-UTRA compared to UTRA is the specification of large DRX cycles in RRC_Connected state. The DRX cycle can be, for example, up to 2.56 s. The large DRX cycles are similar in length to idle state DRX cycles, but measurements are reported to the E-UTRA network for handover decision making.

As well as the performance requirements for supporting reselection and measurements for handover, [3] also includes performance requirements for radio related RRC procedures including RRC re-establishment and random access, and additionally contains requirements for UE transmit timing, and UE timer accuracy. This section gives an overview of the mobility related requirements.

11.10.1 Idle State Mobility

From a minimum performance perspective, [3] specifies how quickly the UE should be able to detect and evaluate the possibility of reselection to newly detectable target cells. In this context, an E-UTRA cell is considered detectable if it has signal to noise and interference ratio SCH_RP/$I_{ot} \geq -3$ dB. For cells that have previously been detected, [3] also specifies the minimum rate at which additional measurements should be made, and the minimum acceptable filtering and spacing of the samples used to support reselection. Maximum allowed evaluation times for cells that have been previously detected are also given. Largely, the requirements for measurement rate and maximum evaluation time are copied from those used previously for UTRA since basic needs for mobility should be similar.

Additionally, [3] specifies that reselection criteria will be re-evaluated at least every DRX cycle, since a new serving cell measurement, and probably also new neighbor cell measurements must be performed at least every DRX cycle. Also, once the reselection decision has been made, it is expected that the UE will make a short interruption in receiving the downlink paging channels when performing reselections. For reselections to E-UTRA and UTRA target cells, this is specified as the time needed to read system information of the target cell and an additional 50 ms implementation allowance.

11.10.2 Connected State Mobility when DRX is Not Active

In RRC_Connected state, the UE continuously searches for, and measures intra-frequency cells, and may additionally search for, and measure inter-frequency and inter-RAT cells if certain conditions are fulfilled, including the configuration of a measurement gap sequence if one is needed by the UE. Both periodic and event triggered reporting mechanisms are defined in RRC specifications to provide measurement results to the E-UTRA network. The relevant measurement quantities are shown in Table 11.23 for different radio technologies.

In general terms, [3] defines minimum performance requirements which indicate how quickly newly detectable cells should be identified and reported, measurement periods for cells that have been detected, and the applicable accuracy requirements and measurement report mapping for each of the measurement quantities in a measurement report. Furthermore, there are requirements in chapter 5 of [3] defining some aspects of how quickly the handover should be performed by the UE, once network mobility management algorithms have made the decision to perform a handover, and transmitted an RRC message to the UE to initiate the handover.

Table 11.23 Measurement quantities for different radio access technologies

Radio technology	Applicable measurement quantities
E-UTRA	Reference Signal Received Power (RSRP). Reference Signal Received Quality (RSRQ)
UTRA FDD	CPICH Received Symbol Code Power (RSCP). CPICH Ec/Io
UTRA TDD	PCCPCH received symbol code power (PCCPCH RSCP)
GSM	RSSI Note: BSIC confirmation may be requested for GSM measurements

11.10.2.1 Cell Identification

When in RRC_Connected state, the UE continuously attempts to search for, and identify new cells. Unlike UTRA, there is no explicit neighbor cell list containing the physical identities of E-UTRA neighbor cells. For both E-UTRA FDD and E-UTRA TDD intra-frequency cell identification, there are rather similar requirements and the UE is required to identify a newly detectable cell within no more than 800 ms. This should be considered a general requirement applicable in a wide range of propagation conditions when the SCH_RP/$I_{ot} \geq -6$ dB and the other side conditions given in [3] are met. The 800 ms requirement was agreed after a simulation campaign in 3GPP where cell identification was simulated by a number of companies at different SCH_RP/I_{ot} ratios.

It is also important to note that cell identification includes a 200 ms measurement period after the cell has been detected. Therefore, to comply with the requirement to identify cells within 800 ms, the UE needs to be able to detect cells internally in a shorter time to allow for the measurement period..

When less than 100% of the time is available for intra-frequency cell identification purposes – because, for example, the intra-frequency reception time is punctured by measurement gaps – the 800 ms time requirement is scaled to reflect the reduced time available.

For E-UTRA inter-frequency measurements, a similar approach is taken. Since inter-frequency cell identification is performed in measurement gaps, the basic identification time $T_{basic\_identify}$ is scaled according to the configured gap density so that

$$T_{Identify\_Inter} = T_{Basic\_Identify\_Inter} \cdot \frac{T_{measurement\_Period\_Inter\_FDD}}{T_{Inter}} \text{ [ms]} \quad (11.16)$$

The $T_{measurement\_Period\_Inter}$ is multiplied by the number of E-UTRA carriers which the UE is monitoring (denoted as N_{freq}), which in turn means that the identification time $T_{Identify\_Inter}$ is also multiplied by N_{freq} and so the requirement for inter-frequency cell identification becomes longer the more frequency layers are configured.

Similar requirements are also defined in [3] for UTRA cell identification when in E-UTRA RRC_Connected state.

11.10.2.2 Measurement of Identified Cells

Once cells are identified, the UE performs measurements on them, over a defined measurement period. 3GPP specifications do not define the sample rate at which the UE Layer 1 (L1) is required to make measurement (or even that uniform sampling is necessary) but the measurements specified are filtered over a standardized measurement period to ensure consistent UE behavior. The UE is also expected to meet the accuracy requirements (discussed further in section 11.10.2.3) over the measurement period, which places constraints on how many L1 samples of the measurement need to be taken and filtered during the measurement period.

For intra-frequency cells, minimum capabilities for measurement are defined in [3]. In summary, the UE is required to have the capability to measure 8 intra-frequency cells when 100% of the time is available for intra-frequency measurements. When less time is available, for example due to inter-frequency measurement gaps, then the requirement is scaled down accordingly. The period for intra-frequency measurements is specified as 200 ms, although this would again be scaled up if less time is available for intra-frequency measurements.

For inter-frequency cells, there are two measurement period requirements defined in [3], one of which is a mandatory requirement, and one of which may be optionally supported by the UE. The mandatory requirement is based on a measurement bandwidth of 6 resource blocks, in which case the measurement period is $480 \times N_{freq}$ ms. When it is indicated by signaling that a bandwidth of at least 50 resource blocks is in use throughout a particular frequency layer, the UE may optionally support a measurement period of $240 \times N_{freq}$ ms. In this case, the UE knows from the signaling that it is safe to make the measurement over the wider bandwidth, and may therefore be able to achieve the required accuracy while filtering fewer measurement samples in the time domain, making possible a reduced measurement period.

For inter-frequency measurements, the minimum requirement is that the UE is capable of supporting three E-UTRA carrier frequencies (in addition to the intra-frequency layer) and on each of these three carrier frequencies it should be capable of measuring at least four cells.

11.10.2.3 Accuracy Requirements and Report Mapping

For both RSRP and RSRQ, absolute and relative accuracy requirements are defined in [3] chapter 9. Absolute accuracy considers the difference between the actual and the ideal measurements for a single cell, whereas relative accuracy considers how much error can be expected when comparing the levels of two cells.

For RSRP, both intra-frequency and inter-frequency absolute and relative accuracy requirements are defined. For comparison of two cells measured on the same frequency, the main sources of inaccuracy can be studied in link level simulations of the measurements, and this was the approach used to define the accuracy requirements. For absolute measurements of RSRP, uncertainty in the RF gain setting is a significant additional source of error, and this is reflected in the accuracy requirements, especially in extreme temperature and voltage conditions. When considering inter-frequency relative RSRP accuracy, some of the uncertainties in RF gain setting cancel out, since they apply to both the cells which are being compared, so the performance requirement for inter-frequency RSRP relative accuracy is somewhat tighter than the inter-frequency absolute accuracy requirement, but not as tight as the intra-frequency relative accuracy requirement.

When considering the accuracy of RSRQ, it should be noted that comparison of two RSRQ measurements on the same frequency is not particularly useful. The reason is that the RSSI component of the RSRQ in both measurements will be similar because for intra-frequency both cells must be measured on the same frequency layer. For this reason, only absolute accuracy requirements are defined for intra-frequency RSRQ measurements. Since RF gain setting uncertainties to an extent affect both the measurement of the RSRP and RSSI components in RSRQ, uncertainties will somewhat cancel out, and as such, RSRQ absolute accuracy is required to be better than RSRP absolute accuracy.

[3] also defines RSRP and RSRQ report mapping, which defines how a UE measured quantity value should be mapped to a signaled information element. This defines the range and resolution of the values which may be reported by the UE.

11.10.3 Connected State Mobility when DRX is Active

One new feature of E-UTRA compared to UTRA which is important from a mobility aspect is that rather large DRX cycles (e.g. up to 2.56 s) may be used when the UE is in RRC_Connected

state. This means that inactive UEs may still be kept in the connected state, and will use connected state mobility procedures such as handover. To allow power saving opportunities for such devices, [3] specifies different measurement performance requirements which are applicable when large DRX cycles (80 ms and greater) are active. The basis of the large DRX measurement performance requirements is to ensure that the UE would be able to perform the mobility related measurements at or around the time when the receiver is active in the DRX cycle anyway. Power saving can be performed for the remainder of the DRX. To facilitate this, measurement periods and cell identification times are specified as a function of the configured DRX cycle length.

11.10.4 Handover Execution Performance Requirements

Similarly to UTRA specifications, handover performance is characterized by two different delays, which are illustrated in Figure 11.57. Total handover delay is denoted by $D_{handover}$ and is the total time between the end of transmission of the handover command on the source cell, and the start of UE transmissions to the target cell. In addition, maximum allowable interruption time is separately specified, and represents the time for which the UE is no longer receiving and transmitting to the source cell, and has not yet started transmission to the target cell.

RRC procedure delays are specified in [42] section 11.2 and the interruption time is given by

$$T_{interrupt} = T_{search} + T_{IU} + 20\,\text{ms} \tag{11.17}$$

where T_{search} is the time taken to search for the target cell. This will be 0 ms when the target cell has previously been measured by the UE within the last 5 s, and is only not zero for a blind handover to a cell not known to the UE. T_{IU} relates to the uncertainty in timing between the physical channel structure of the old cell. An additional 20 ms implementation margin is also included in the requirement.

As indicated in Figure 11.57, the total handover delay is either $D_{handover}$ = RRC procedure delay + $T_{interrupt}$ or activation time + $T_{interrupt}$ if an activation time is given which is greater than the RRC procedure delay in the future.

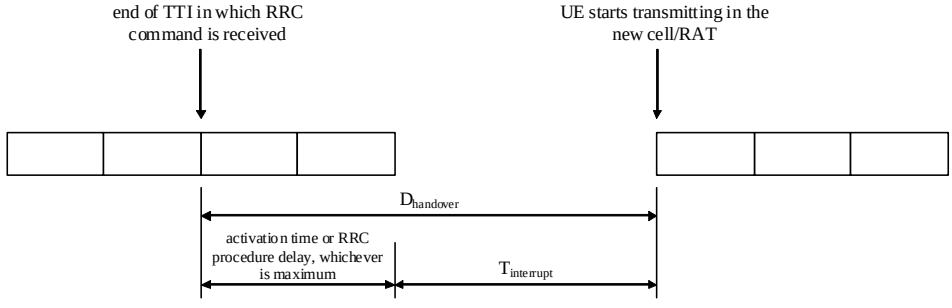


Figure 11.57 Handover delay and interruption time

11.11 Summary

Clearly defined performance requirements are an essential part of an open well-functioning standard. The requirements are needed to provide predictable performance within an operator's own band in terms of data rates, system capacity, coverage and mobility with different terminals and different radio network vendors. The requirements are also needed to facilitate the coexistence of LTE in the presence of other systems as well as the coexistence of adjacent LTE operators in the same area. The regulatory requirements are considered when defining the emission limits. 3GPP has defined minimum RF performance requirements for LTE terminals (UE) and for base stations (eNodeB) facilitating a consistent and predictable system performance in a multi-vendor environment.

3GPP LTE frequency variants cover all the relevant cellular bands. The first version of the specifications covers 17 different variants for FDD and 8 for TDD bands. More frequency variants will be added in the future. The frequency variants are independent of the other content of the releases.

3GPP LTE requirements are defined in such a way that it enables efficient multimode GSM/WCDMA/LTE device implementation from the RF requirements perspective. Yet, the need to support receive diversity required for MIMO operation, and the multiple frequency band combinations present a significant challenge for the front-end components optimization of such mobile devices. The SC-FDMA transmission in uplink has similar requirements as the current HSUPA transmission. The baseband processing requirements are naturally increased due to the high data rates in LTE but the simple front-end structure in OFDMA and the Turbo decoding parallelization simplifies the practical implementations.

References

[1] 3GPP Technical Specification 36.101 'User Equipment (UE) radio transmission and reception', v. 8.3.0.
[2] 3GPP Technical Specification 36.104 'Base Station (BS) radio transmission and reception', v. 8.3.0.
[3] 3GPP Technical Specification 36.133 'Requirements for support of radio resource management', v. 8.3.0.
[4] 3GPP Technical Report 36.942 'Radio Frequency (RF) system scenarios', v. 8.0.0.
[5] 3GPP Technical Specifications 36.141 'Base Station (BS) conformance testing', v. 8.0.0.
[6] 3GPP Technical Specifications 36.108 'Common test environments for User Equipment (UE); Conformance testing', v. 8.0.0.
[7] 3GPP Technical Specifications 36.114 'User Equipment (UE)/Mobile Station (MS) Over The Air (OTA) antenna performance; Conformance testing', v. 8.0.0.
[8] 3GPP R4–070342 'On E-UTRA spectrum emission mask and ACLR requirements', April 2007.
[9] European Communications Committee (ECC)/European Conference of Postal and Telecommunications Administrations (CEPT) Report #19, 'Report from CEPT to the European Commission in response to the Mandate to develop least restrictive technical conditions for frequency bands addressed in the context of WAPECS', December 2007.
[10] 3GPP R4–070124 'System simulation results for derivation of E-UTRA BS EVM requirements', February 2007.
[11] Holma, H., Toskala, A., 'WCDMA for UMTS: HSPA Evolution and LTE', 4th edition, Wiley, 2007.
[12] The Mobile Industry Processor Interface (MIPI℠) Alliance, www.mipi.org
[13] Myung, H.G., Junsung Lim, Goodman, D.J., 'Peak-To-Average Power Ratio of Single Carrier FDMA Signals with Pulse Shaping', Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on September 2006: 1–5.
[14] Valkama, M., Anttila, L., Renfors, M., 'Some radio implementation challenges in 3G-LTE context', Signal Processing Advances in Wireless Communications, 2007. SPAWC 2007. IEEE 8th Workshop on 17–20 June 2007: 1–5.

[15] Priyanto, B.E., Sorensen, T.B., Jensen, O.K., Larsen, T., Kolding, T., Mogensen, P., 'Impact of polar transmitter imperfections on UTRA LTE uplink performance', Norchip, 2007, 19–20 November 2007: 1–4.
[16] Talonen, M., Lindfors, S., 'System requirements for OFDM polar transmitter', Circuit Theory and Design, 2005. Proceedings of the 2005 European Conference, Volume 3, 28 August–2 September 2005: III/69–III/72.
[17] Chow, Y.H , Yong, C.K., Lee, Joan, Lee, H.K., Rajendran, J., Khoo, S.H., Soo, M.L., Chan, C.F., 'A variable supply, (2.3–2.7) GHz linear power amplifier module for IEEE 802.16e and LTE applications using E-mode pHEMT technology'. Microwave Symposium Digest, 2008 IEEE MTT-S International 15–20 June 2008: 871–874.
[18] Ericsson, ' TP 36.101: REFSENS and associated requirements', TSG-RAN Working Group 4 (Radio) meeting #46bis, R4–080696, Shenzhen, China, 31 March–4 April, 2008.
[19] Ericsson, 'Simulation results for reference sensitivity and dynamic range with updated TBS sizes', TSG-RAN Working Group 4 (Radio) meeting #47bis, R4–081612, Munich, 16th–20th June 2008.
[20] Nokia Siemens Networks, 'Ideal simulation results for RF receiver requirements', 3GPP TSG-RAN WG4 Meeting #48, R4–081841, Jeju, Korea, 18th–22nd August, 2008.
[21] STN-wireless internal WCDMA handset benchmarking test report.
[22] Anadigics AWT6241 WCDMA Power Amplifier datasheet, http://www.anadigics.com/products/handsets_data-cards/wcdma_power_amplifiers/awt6241
[23] Ericsson, 'Introduction of MSD (Maximum Sensitivity Degradation)', TSG-RAN Working Group 4 (Radio) meeting #49, R4–082972, Prague, Czech Republic, 10–14 November 2008.
[24] NTT DoCoMo, Fujitsu, Panasonic, ' Performance requirements on Self interference due to transmitter noise', TSG-RAN Working Group 4 Meeting #47, R4–080873, Kansas, USA, 5–9 May 2008.
[25] Ericsson, ' Introduction of MSD (Maximum Sensitivity Degradation)', 36.101 Change Request # 90, R4–083164, 3GPP TSG-RAN WG4 Meeting #49, Prague, Czech Republic, 10–14 November 2008.
[26] NTT DOCOMO, ' Maximum Sensitivity Reduction values for Band 6/9/11', R4–083043, TSG-RAN Working Group 4 Meeting #49, Prague, Czech Republic, Nov. 10–14, 2008.
[27] 'Half Duplex FDD Operation in LTE', 3GPP TSG RAN WG4 Meeting #46, R4–080255, Sorrento (Italy), February 11th to 15th 2008.[28] Ericsson, 'HD-FDD from a UE perspective', TSG-RAN Working Group 4 (Radio) meeting #46bis,R4–080695 Shenzhen, China, 31 March – 4 April 2008.
[29] Manstretta, D , Brandolini, M., Svelto, F., 'Second-Order Intermodulation Mechanisms in CMOS Downconverters', IEEE Journal of Solid-State Circuits, Vol. 38, No. 3, March 2003, 394–406.
[30] Faulkner, M., 'The Effect of Filtering on the Performance of OFDM Systems', IEEE Transactions on Vehicular Technology, vol. 49, no. 5, September 2000: 1877–1883.
[31] Lindoff, B., Wilhelmsson, L., 'On selectivity filter design for 3G long-term evolution mobile terminals', Signal Processing Advances in Wireless Communications, 2007. SPAWC 2007. IEEE 8th Workshop on Volume, Issue, 17–20 June 2007: :1–5.
[32] Dardari, D., 'Joint Clip and Quantization Effects Characterization in OFDM Receivers', IEEE Transactions On Circuits And Systems – I: Regular Papers, vol. 53, no. 8, August 2006: 1741–1748.
[33] Locher, M. *et al*., 'A Versatile, Low Power, High Performance BiCMOS, MIMO/Diversity Direct Conversion Transceiver IC for Wi-Bro/WiMax (802.16e)', IEEE Journal of Solid State Circuits, vol. 43, no. 8, August 2008: 1–10.
[34] Norsworthy, S.R., Schreier, R., Temes, G.C., 'Delta-Sigma data converters, theory, design and simulation', Wiley Interscience, 1996.
[35] Internal STN-wireless data extracted with simulation parameter modifications, From: 'Ouzounov, S. van Veldhoven, R. Bastiaansen, C. Vongehr, K. van Wegberg, R. Geelen, G. Breems, L. van Roermund, A., "A 1.2V 121-Mode ΣΔModulator for Wireless Receivers in 90nm CMOS", Solid-State Circuits Conference, 2007. ISSCC 2007. Digest of Technical Papers. IEEE International Publication Date: 11–15 February 2007: 242–600.
[36] Martel, P., Lossois, G., Danchesi, C., Brunel, D., Noël, L., 'Experimental EVM budget investigations in Zero-IF WCDMA receivers', submitted to Electronics Letters.
[37] Windisch, M., Fettweis, G., 'Performance Degradation due to I/Q Imbalance in Multi-Carrier Direct Conversion Receivers: A Theoretical Analysis', Communications, 2006. ICC apos;06. IEEE International Conference on Volume 1, June 2006: 257–262.
[38] Robertson, P., Kaiser, S., 'Analysis of the effects of phase-noise in orthogonal frequency division multiplex (OFDM) systems', Communications, 1995. ICC 95 Seattle, Gateway to Globalization, 1995 IEEE International Conference on Volume 3, 18–22 June 1995: 1652–1657.
[39] Pollet, T., Moeneclaey, M., Jeanclaude, I., Sari, H., 'Effect of carrier phase jitter on single-carrier and multi-carrier QAM systems', Communications, 1995. ICC 95 Seattle, Gateway to Globalization, 1995 IEEE International Conference on Volume 2, 18–22 June 1995:1046–1050.

[40] 3GPP Technical Specification 25.133 'Requirements for support of radio resource management (UTRA)', v. 8.3.0.
[41] 3GPP Technical Specification 36.213 'Physical layer procedures', v. 8.3.0.
[42] 3GPP Technical Specification 36.331 'Radio Resource Control (RRC); Protocol specification', v. 8.3.0.
[43] 3GPP Technical Specification 25.101 'User Equipment (UE) radio transmission and reception (UTRA)', v. 8.3.0.
[44] Qualcomm Europe, Change Request 'R4–082669: UE ACS frequency offset', TSG-RAN Working Group 4 (Radio) meeting #49, Prague, CZ, 10–14 November, 2008.
[45] Motorola, 'Reference Channel and Noise Estimators', 3GPP E-mail reflector document for TSG-RAN Working Group 4 Meeting #46, Sorrento, Italy, 11–15 February, 2008.
[46] Almers, P. *et al*., 'Survey of channel and radio propagation models for wireless MIMO systems'. EURASIP Journal on Wireless Communications and Networking, 957–1000, July 2007.
[47] Ericsson, 'R4–071814: Clarification of TX EVM model', TSG-RAN Working Group 4 (Radio) meeting #44, Athens, Greece, 20–24 August 2007.
[48] Bannelli, P., Cacopardi, S., 'Theoretical Analysis and Performance of OFDM Signals in Nonlinear AWGN Channels,' IEEE trans. on Communications, vol. 48, no. 3, March 2000.
[49] Berkmann J. *et al*., 'On 3G LTE Terminal Implementation – Standard, Algorithms, Complexities and Challenges', Wireless Communications and Mobile Computing International Conference, 2008. IWCMC '08.

12

LTE TDD Mode

Che Xiangguang, Troels Kolding, Peter Skov, Wang Haiming and Antti Toskala

12.1 Introduction

With full coverage in the 3GPP Release 8 specifications of both Time Division Duplex (TDD) and Frequency Division Duplex (FDD) modes of operation, LTE can effectively be deployed in both the paired and unpaired spectrum. LTE TDD and FDD modes have been greatly harmonized in the sense that both modes share the same underlying framework, including radio access schemes OFDMA in downlink and SC-FDMA in uplink, basic subframe formats, configuration protocols, etc. As clear indication of the harmonization, the TDD mode is included together with the FDD mode in the same set of specifications, including the physical layer where there are just a few differences due to the uplink/downlink switching operation. In terms of architecture there are no differences between FDD and TDD and the very few differences in the MAC and higher layer protocols relate to TDD specific physical layer parameters. Procedures are kept the same. Thus there will be high implementation synergies between the two modes allowing for efficient support of both TDD and FDD in the same network or user device. Coexistence would of course still require careful analysis.

Another key feature of the LTE TDD mode (known also as TD-LTE) is the commonality with TD-SCDMA. This is an advantage as in, e.g. China, where the Release 4 based TD-SCDMA (including enhancements from later releases) has opened up a large-scale TDD system deployment, paving the way for further deployment of 3GPP based LTE TDD using the available unpaired spectrum. As presented in Chapter 11, there is a global trend to reserve significant unpaired spectrum allocations.

In this chapter, the detailed aspects of LTE TDD that differ from the FDD mode are introduced. Further, information related to both the link and system performance of the LTE TDD mode of operation is given.

LTE for UMTS: OFDMA and SC-FDMA Based Radio Access Edited by Harri Holma and Antti Toskala
 ISBN: 978-0-470-99401-6

12.2 LTE TDD Fundamentals

The basic principle of TDD is to use the same frequency band for transmission and reception but to alternate the transmission direction in time. As shown in Figure 12.1 this is a fundamental difference compared to FDD, where different frequencies are used for continuous UE reception and transmission. Like FDD, LTE TDD supports bandwidths from 1.4 MHz up to 20 MHz but depending on the frequency band, the number of supported bandwidths may be less than the full range. For example, for the 2.5 GHz band, it is not likely that the smallest bandwidths will be supported. Since the bandwidth is shared between uplink and downlink and the maximum bandwidth is specified to be 20 MHz in Release 8, the maximum achievable data rates are lower than in LTE FDD. This way the same receiver and transmitter processing capability can be used with both TDD and FDD modes enabling faster deployment of LTE.

The TDD system can be implemented on an unpaired band (or in two paired bands separately) while the FDD system always requires a pair of bands with a reasonable separation between uplink and downlink directions, known as the duplex separation. In a FDD UE implementation this normally requires a duplex filter when simultaneous transmission and reception is facilitated. In a TDD system the UE does not need such a duplex filter. The complexity of the duplex filter increases when the uplink and downlink frequency bands are placed in closer proximity. In some of the future spectrum allocations it is foreseen that it will be easier to find new unpaired allocations than paired allocations with sensible duplex separation thereby increasing further the scope of applicability for TDD.

However, since uplink and downlink share the same frequency band, the signals in these two transmission directions can interfere with each other. This is illustrated in Figure 12.2

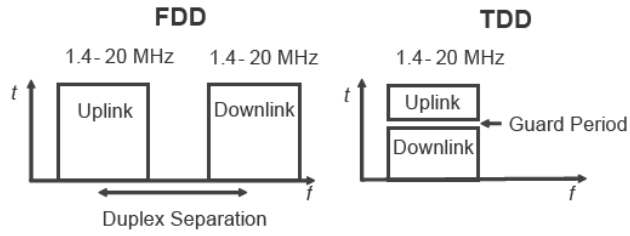


Figure 12.1 Principles of FDD and TDD modes of operation

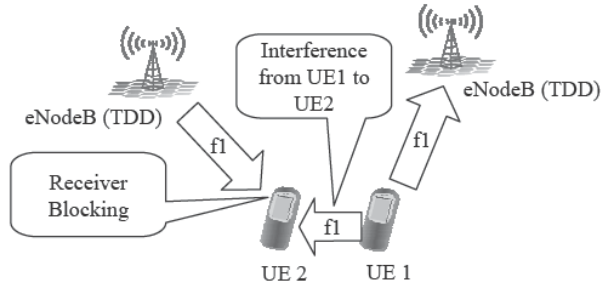


Figure 12.2 Interference from uplink to downlink in uncoordinated TDD operation

with the use of TDD on the same frequency without coordination and synchronization between sites in the same coverage area. For uncoordinated deployment (unsynchronized) on the same frequency band, the devices connected to the cells with different timing and/or different uplink/downlink allocation may cause blocking for other users. In LTE TDD the base stations need to be synchronized to each other at frame level in the same coverage area to avoid this interference. This can be typically done by using, for example, satellite based solutions like GPS or Galileo or by having another external timing reference shared by the LTE TDD base stations within the same coverage area. LTE FDD does not need the base station synchronization. There is no interference between uplink and downlink in FDD due to the duplex separation of the carriers.

Two adjacent LTE TDD operators (on adjacent carriers) should preferably synchronize the base stations and allocate the same asymmetry between uplink and downlink to avoid potentially detrimental interference to the system reliability. If the two operators do not coordinate the deployments, there is instead a need for guard bands and additional filtering. These requirements are discussed in Chapter 11.

12.2.1 LTE TDD Frame Structure

As the single frequency block is shared in the time domain between uplink and downlink (and also between users), the transmission in LTE TDD is not continuous. While often also being the case for data transmissions towards a certain user in LTE FDD mode, the level of discontinuity then depends entirely on the scheduling function (except for half-duplex FDD terminals). For control channels, e.g. the PDCCH and the PHICH, the transmission for FDD is continuous. For LTE TDD all uplink transmissions need to be on hold while any downlink resource is used and, conversely, the downlink needs to be totally silent when any of the UE is transmitting in the uplink direction. Switching between transmission directions has a small hardware delay (for both UE and eNodeB) and must be compensated. To control the resulting switching transients a Guard Period (GP) is allocated which compensates for the maximum propagation delay of interfering components (e.g. depends on cell size and level of available cell isolation). The impact of discontinuous uplink and downlink on the link budget in LTE TDD is covered specifically in section 12.5.

To explain the exact implementation of the mechanism for switching between downlink and uplink and vice versa, consider the example setup of the LTE TDD frame structure in Figure 12.3.

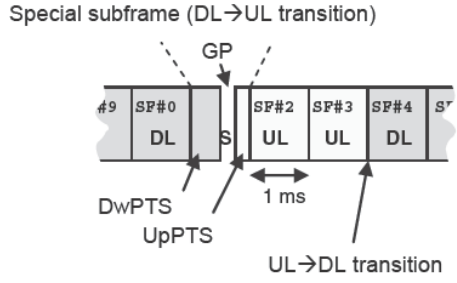


Figure 12.3 Simple illustration of the DL→UL and UL→DL transitions implemented in downlink (DL), uplink (UL), and special (S) subframes (SF)

The subframe denoted by either uplink (UL) or downlink (DL) has a design in common with LTE FDD with some minor but significant differences related to common control channels. In LTE TDD there is maximally one DL→UL and one UL→DL transition per 5 ms period (half-frame). The UL→DL transition is carried out for all intra-cell UE by the process of time alignment. The eNodeB instructs each UE to use a specific time offset so that all UE signals are aligned when they arrive at the eNodeB. Hence, uplink is synchronous as is the case for FDD. To ensure that the UE has sufficient time to shut down its transmission and switch to listening mode, the UE does not transmit a signal during the last 10–20 ms of subframe #3 in Figure 12.3. This procedure ensures that there is no UE transmission power from the own cell that spills over into the downlink transmission. Although eNodeBs in different cells are fully synchronized, this method does not prevent UE from other cells spilling their interference into the downlink transmission of the current sector. However, in practice this is less of a problem since any individual UE has limited transmission power.

While the UL→DL switching is merely an intra-cell method, the DL→UL switching method ensures that the high-power downlink transmissions from eNodeBs from other neighbor cells do not interfere when the eNodeB UL reception is ongoing in the current cell. Adopting the methodology of TD-SCDMA, LTE TDD introduces a special (S) subframe that is divided into three parts; the Downlink Pilot Time Slot (DwPTS), the GP, and the Uplink Pilot Time Slot (UpPTS). The special subframe replaces what would have been a normal subframe #1. The individual time duration in OFDM symbols of the special subframe parts are to some extent adjustable and the exact configuration of the special time slot will impact the performance. The GP implements the DL→UL transition and the GP has to be sufficiently long to cover the propagation delay of all critical downlink interferers on the same or adjacent carriers as well as the hardware switch-off time. Hence, the correct setting of the GP depends on network topology, antenna configurations, etc. To fit into the general LTE frame numerology, the total duration of DwPTS, GP, and UpPTS is always 1 ms.

DwPTS is considered as a 'normal' downlink subframe and carries control information as well as data for those cases when sufficient duration is configured. High commonality is achieved by rescaling the transport block size according to its length. In this way the effective coding rate for a certain selection of payload and transmission bandwidth will stay the same. UpPTS is primarily intended for sounding reference signal (SRS) transmissions from the UE, but LTE TDD also introduces the short RACH concept so that this space may be used for access purposes as well. The flexibility for the different components of the special subframe is summarized in Table 12.1 when using the normal cyclic prefix. The GP can be up to about 700 μs thus supporting a cell range up to 100 km. For more details related to the many possible configurations, including the extended cyclic prefix, the reader is referred to [1]. When discussing coexistence with TD-SCDMA in section 12.5 more details related to special subframe configuration are given.

Table 12.1 Possible durations configurable for special subframe components. Total duration of the three components is always 1 ms (normal cyclic prefix) [1]

Component	Unit	Range of duration
Downlink pilot time slot (DwPTS)	# OFDM symbols (71 μs)	3–12
Guard period (GP)	# OFDM symbols (71 μs)	1–10
Uplink pilot time slot (UpPTS)	# OFDM symbols (71 μs)	1–2

12.2.2 Asymmetric Uplink/Downlink Capacity Allocation

A key advantage to TDD is the ability to adjust the available system resources (time and frequency) to either downlink or uplink to match perfectly the uplink and downlink traffic characteristics of the cell. This is done by changing the duplex switching point and thus moving capacity from uplink to downlink, or vice versa. The LTE TDD frame structure can be adjusted to have either 5 ms or 10 ms uplink-to-downlink switch point periodicity. The resulting resource split is either balanced or between two extreme cases:

- single 1 ms subframe for uplink and 8 times 1 ms subframe for downlink per 10 ms frame;
- if we want to maximally boost the uplink capacity, then we can have three subframes for uplink and one for downlink per 5 ms. Thus the resulting uplink activity factor can be adjusted from 10% to 60% (if we do not consider the UpPTS).

For continuous coverage across several cells, the chosen asymmetry is normally aligned between cells to avoid the interference between transmission directions as described earlier. Hence, synchronization in a LTE TDD system deployed in a wide area is conducted both at frame and uplink–downlink configuration level. In practice it is expected that the uplink–downlink configuration is changed in the network only very seldom and the associated signaling of LTE TDD in Release 8 has been optimized according to this assumption, e.g. a LTE TDD Release 8 cellular system would be characterized as a static or semi-static TDD system. The UE is informed about the active uplink–downlink configuration via the System Information Block (SIB-1), which is broadcast via the Dynamic Broadcast Channel (D-BCH) to the cell with an update rate of 80 ms. The knowledge about which uplink–downlink configuration is active in the cell is essential for the UE to know the location of critical control channels and the timing of adaptation methods such as HARQ.

12.2.3 Co-existence with TD-SCDMA

As mentioned, there is a strong legacy between LTE TDD and TD-SCDMA. The coexistence of these systems on the same or adjacent carriers has therefore been a key discussion point during the standardization process. For sharing a site with TD-SCDMA and using the same frequency band, the systems need to have an aligned uplink/downlink split to avoid the interference between different base station transceivers. As TD-SCDMA slot duration does not match the LTE TDD subframe duration, the LTE subframe parameterization was designed to accommodate coexistence. From the knowledge of the relative uplink/downlink division, the relative timing of TD-SCDMA and LTE TDD can be adjusted to allow coexistence without BTS to BTS interference, as shown in Figure 12.4. Note the duration of fields in LTE TDD subframe with uplink/downlink that vary depending on the configuration, thus the timings shown can take different values.

Besides timing alignment, the exact configuration of the special subframe in LTE TDD also plays an important role when allowing TD-SCDMA/LTE TDD coexistence. Some of the configurations with a good match are listed in Table 12.2, including the required detailed settings for DwPTS, GP and UpPTS. Normal cyclic prefix is assumed. For more options for configuration, the reader is referred to [1].

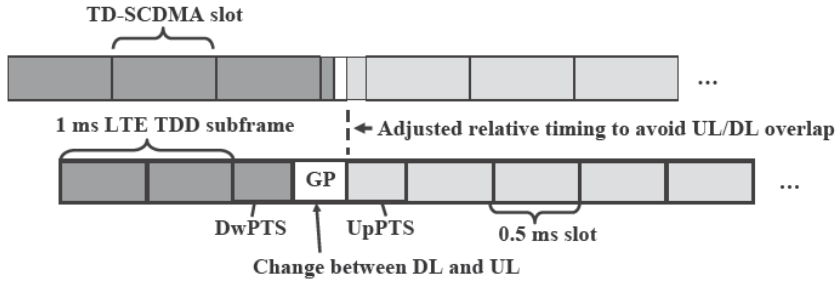


Figure 12.4 Ensuring TD-SCDMA and LTE TDD coexistence with compatible UL/DL timings

Table 12.2 Example modes for coexistence between TD-SCDMA and LTE TDD (normal cyclic prefix)

TD-SCDMA configuration	LTE TDD configuration	Special subframe configuration (OFDM symbols)			
		Configuration	DwPTS	GP	UpPTS
5DL-2UL	#2 (3DL-1UL)	#5	3	9	2
4DL-3UL	#1 (2DL-2UL)	#7	10	2	2
2DL-5UL	#0 (1DL-3UL)	#5	3	9	2

12.2.4 Channel Reciprocity

One interesting aspect for TDD systems is that the signals in uplink and downlink travel through the same frequency band. Assuming symmetric antenna configurations and adequate RF chain calibration, there is a high correlation of the fading on the own signal between uplink and downlink known as channel reciprocity. The key associated benefit is that the measurements in one direction may fully or partially be used to predict the other direction thereby enabling a reduction of the signaling overhead needed for link adaptation, packet scheduling and the use of advanced transmit antenna techniques. As already mentioned antenna configurations need to be reciprocal for full reciprocity to be applicable. The baseline configuration for LTE is two antennas in both eNodeB and UE, so in this case antenna configuration seems to be reciprocal. However, as the UE only has one power amplifier, it can only transmit on one antenna at a time so the only way to measure the full channel in UL is for the UE to switch the sounding transmission between the two antennas. If this is done we can say that both channel and antenna configuration are reciprocal.

An additional practical limitation towards the exploitation of reciprocity is that most adaptation schemes, including link adaptation methods and fast packet scheduling, rely on SINR or throughput assessments. As the interference level may vary significantly between uplink and downlink, the LTE TDD specifications therefore do not rely on the availability of channel reciprocity and allow for a full decoupling of uplink and downlink thus using the very same uplink and downlink reporting methods as are available for LTE FDD. However, within the specifications there is a large degree of freedom available to optimize the total signaling overhead in cases where reciprocity has significant gain and is applicable.

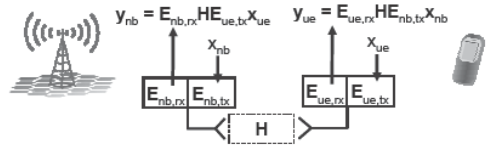


Figure 12.5 Model to illustrate the impact from RF units to channel reciprocity. Capital letters identify matrixes

A simple linear model for the TDD channel including impact from RF-chains is shown in Figure 12.5. From this it is observed that even though the channel represented by **H** is identical for UL and DL then the channel seen by physical layer is different. In the typical application of channel state information we would like to know the channel as seen by physical layer (in DL $\mathbf{E}_{nb,tx}\mathbf{H}\mathbf{E}_{ue,rx}$) because pre-coding and decoding are all done after the signal has passed through the RF parts. This information cannot be estimated from UL sounding. It is possible to improve the channel reciprocity by adding calibration procedures within the Node B and UE but to calibrate the full RF chain so that $\mathbf{x}_{nb} = \mathbf{x}_{ue}$ would imply $\mathbf{y}_{nb} = \mathbf{y}_{ue}$ would require a standardized procedure.

Relying on channel reciprocity for pre-coding is very challenging as reciprocity of the full complex valued channel is assumed. For other purposes, such as reciprocity based fast power control, we only rely on reciprocity of the channel attenuation and this would be less sensitive towards calibration errors [2].

12.2.5 Multiple Access Schemes

In LTE TDD the applied multiple access schemes are the same as with LTE FDD. Hence, OFDMA is used for the downlink and SC-FDMA is used for the uplink direction as shown in Figure 12.6. The reasoning behind the selection was shared by TDD and FDD modes respectively. The selection of a power efficient uplink scheme was very important especially for a TDD device because of the more stringent uplink link budget resulting from the limited

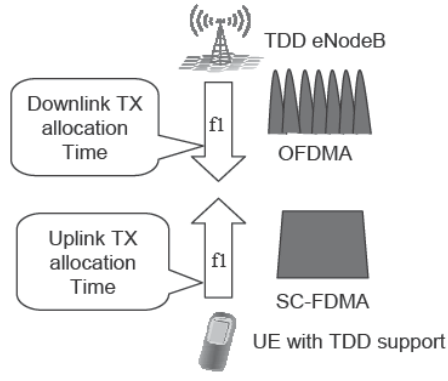


Figure 12.6 LTE TDD multiple access schemes in uplink and downlink

uplink transmission time allocation. From an implementation complexity viewpoint, this is an improvement over, for example, WCDMA, where different multiple access schemes between FDD (with WCDMA) and TDD (with TD/CDMA) modes required different receiver and transmitter solutions (even with the identical chip rate in Release 99) in the baseband implementation. Hence, complexity of any dual mode TDD/FDD implementation is reduced in LTE when compared to UTRAN.

12.3 TDD Control Design

Due to the special frame structure, the placement of critical control channels is different for LTE TDD than it is for LTE FDD. The exact placement of control channels depends on which uplink–downlink configuration is active in the cell and is thus key information for the UE to know to set up its connection with the network. In the following, the placement of the critical channels as well as their special meaning for LTE TDD is described. For general information about the meaning of the control channels, the reader is referred to Chapter 5, where the dynamic use of Physical Downlink Shared Control Channel (PDCCH) for uplink and downlink resource allocation is covered.

For FDD operation, where downlink and uplink are in separate frequencies with continuous transmission and reception on its dedicated frequency, the shared control channel design is straightforward due to one-to-one associated downlink and uplink subframes. However, for TDD operation where downlink and uplink share the same frequency for transmission and reception but alternate the transmission direction in time, physical control channel design is a bit more challenging. This was further complicated by the high available flexibility to adjust the downlink/uplink allocation.

12.3.1 Common Control Channels

To illustrate the placement of the different control channels, TDD uplink–downlink configuration #1 is used as reference. The critical system control channels are shown in Figure 12.7 and described in the following.

The Primary Synchronization Signal (PSS) is placed at the third symbol in subframes #1 and #6, which is different from LTE FDD where the PSS is placed at the last symbol of the first slot in subframes #0 and #5. The Secondary Synchronization Signal (SSS) is placed at the

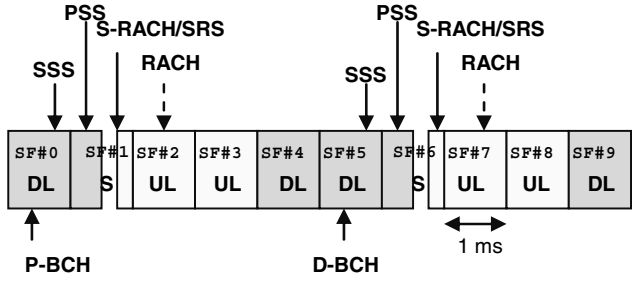


Figure 12.7 Mapping of critical control channels to TDD configuration #1

last symbol in subframes #0 and #5, which is also different from LTE FDD where the SSS is placed at the second last symbol of the first slot in subframes #0 and #5. Synchronization via the PSS is typically a robust process and UE can survive UE-to-UE interference of up to 100 dB for the initial cell search. This high measuring robustness is facilitated by using normalization of samples achieved over a long measuring window.

There are five RACH preamble formats defined for LTE TDD. Four of the RACH preamble formats are common with LTE FDD, while the LTE TDD specific RACH preamble format 4 is known as SHORT RACH (S-RACH) due to the short preamble sequence duration. As shown in Figure 12.7 the S-RACH is transmitted on the UpPTS within the special subframe.

RACH needs to be protected to make access reliable and to allow for the coexistence of multiple LTE TDD users. The RACH channel is fairly flexible regarding the density and placement in time and frequency. The RACH density, similarly to that for LTE FDD, can be 0.5, 1, 2, 3, 4, 5 or 6 in every 10 ms radio frame. The exact placement of RACH in LTE TDD within the available UL resources (i.e. UL subframes and UpPTS) in the configured UL/DL configuration is selected by proper network planning taking into account the requirement for coexistence with TD-SCDMA, traffic load, network topology, etc. All together, there are 58 configurations for RACH placement in time and frequency for each UL/DL configuration with the common principle that RACH channels are first distributed in the time domain among all available UL resources before multiplexed in the frequency domain. This principle is exemplified in Figure 12.8 for UL/DL configuration #0, and the exact RACH configurations can be found in [1]. The maximum available number of RACH channels per UL subframe or UpPTS (for S-RACH only) is six, which is different from FDD where at most one RACH channel is available per UL subframe. This ensures that even with limited UL resources available, sufficient RACH opportunities for high traffic load can be created. As illustrated in Figure 12.9, when there is more than one RACH channel in one UL subframe or UpPTS, the RACH channels are distributed towards both edges of the system bandwidth for RACH in UL subframe or placed from only one edge (either bottom or top) when in UpPTS. For the latter case, the frequency position is altered in the next RACH instance as shown in the figure.

The placement of the Primary Broadcast Channel (P-BCH) and the Dynamic Broadcast Channel (D-BCH) is the same as it is in LTE FDD. However, the detection of D-BCH is slightly different from that in LTE FDD. This is because to detect D-BCH, the size of the control channel in that DL subframe needs to be known to the UE. However, this raises a chicken-and-egg problem since the size of the control channel depends on the active UL/DL configuration and the UE does not yet know the active UL/DL configuration before it correctly detects D-BCH during the initial cell search. For all UL/DL configurations, there are three different sizes for the control channel in any DL subframe, thus the UE will need to perform three hypotheses on control channel size to detect the D-BCH. Once the UE correctly receives the D-BCH, it knows the UL/DL configuration in the cell, after which the hypothesis detection is no longer needed when detecting the D-BCH.

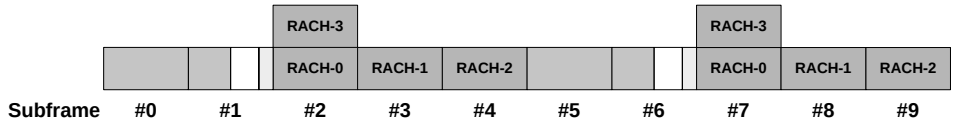


Figure 12.8 RACH distribution/placement in time and frequency in LTE TDD

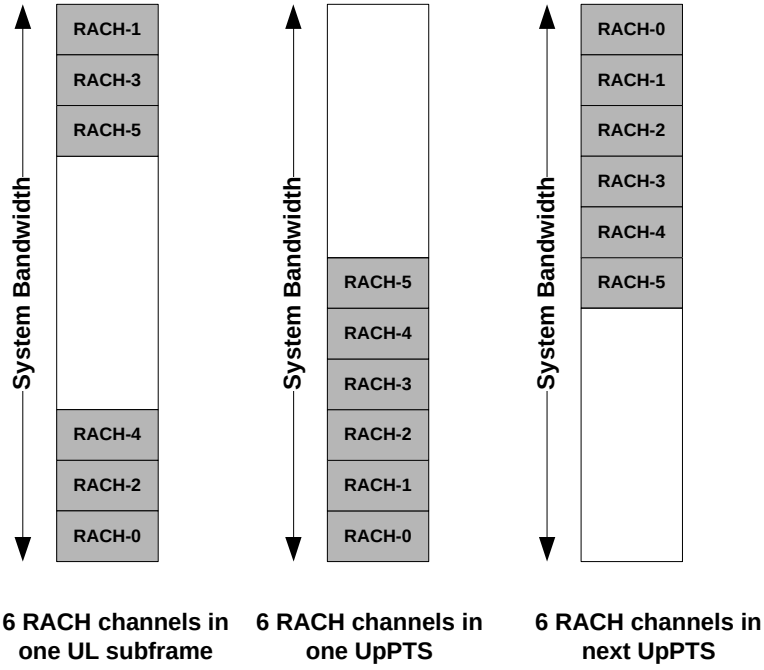


Figure 12.9 RACH placement in the frequency domain in LTE TDD

The paging procedure is the same for LTE TDD and FDD; however, the exact subframes being used for paging are slightly different in LTE TDD. For FDD, the subframes #0, #4, #5 and #9 can be used for paging. For TDD, the subframes #0, #1, #5 and #6 can be used for paging.

12.3.2 Sounding Reference Signal

In LTE TDD the Sounding Reference Signal (SRS) for any UE can be transmitted not only in the last symbol of one UL subframe as in LTE FDD, but also in one or both symbols of UpPTS. Since the UpPTS is anyway available and cannot carry uplink data traffic, it is expected to be the primary location for SRS in LTE TDD. The support of SRS is considered mandatory for all LTE TDD devices due to the installed base of TD-SCDMA base stations, which often have existing beamforming antenna configurations.

12.3.3 HARQ Process and Timing

The key to designing and configuring HARQ for TDD operation is to determine the required processing time of the eNodeB and the UE for Uplink (UL) and Downlink (DL) respectively. The relevant processing times are as follows:

- DL UE: Duration from the time when the last sample of the packet is received in downlink until a HARQ-ACK/NACK is transmitted in UL.
- DL eNodeB: Duration from the time when an HARQ-ACK/NACK is transmitted in UL until the eNodeB can (re)-transmit data on the same HARQ process.
- UL eNodeB: Duration from the time when the last sample of the packet is received in UL until a HARQ-ACK/NACK (or a new allocation on same HARQ process) is transmitted in DL.
- UL UE: Duration from the time when a UL grant (or HARQ-ACK/NACK) is given until the UE is able to transmit the associated packet in UL.

In FDD, the eNodeB and UE processing times for both DL and UL are fixed and assumed to be 3 ms due to invariant DL and UL subframe configuration and continuous DL and UL transmission and reception. In TDD, although the eNodeB and UL (minimum) processing times are the same as for FDD, the actual HARQ timing (i.e. DL UE/eNodeB, UL eNodeB/UE) varies depending on the active DL/UL configuration. This is exemplified in Figure 12.10, where it is seen that compared to the minimum processing time of 4 ms, there is sometimes an additional delay incurred due to unavailable DL or UL subframe after 3 ms.

In addition to the minimum HARQ processing time, a few other decisions were made related to the special subframe to fix the required number of HARQ processes and timing in the specifications:

- There shall be always a PDCCH in the DwPTS at least for UL grant or PHICH. If DwPTS spans more than three OFDM symbols it also contains a PDSCH.
- UpPTS does not contain HARQ control information nor a PUSCH.

With the given assumptions of processing time and special subframe behavior, the number of the DL HARQ process and the UL HARQ process varies with TDD between 4 and 15 for DL and 1 and 7 in UL (in FDD that was always constant). As for FDD, in TDD the DL HARQ is asynchronous and the UL HARQ is synchronous. The resulting delay thus varies depending on the subframe being used as well as on the uplink/downlink split applied, and can take values between 4 and 7 ms for the delay k_1 between the associated UL grant (or an intended PHICH) on the PDCCH and the UL data transmission over the PUSCH. Respectively the delay k_2 between the associated UL data transmission over PUSCH and PHICH (or a UL grant for re-/new transmission) over the PDCCH also varies between 4 and 7 ms.

The multi-TTI scheduling scheme in TDD allows for efficient use of the downlink shared control channel resources (PDCCH) in case less downlink resources are available and further decreases the UE complexity of decoding PDCCH (see Figure 12.11).

The DL HARQ timing delay k_3 between the associated DL data transmission over PDSCH and the UL HARQ-ACK transmission over PUCCH or PUSCH varies between 4 and 13 ms depending again on the uplink/downlink split applied. This is now simplified since TDD DL HARQ is operated in asynchronous mode like LTE FDD. The intended DL HARQ timing is derived by $n + k_3$, where n is the addressed subframe index number.

Due to the discontinuous UL/DL transmission, the TDD frame structure adds some delay to the user plane transmission when HARQ is involved. While the HARQ Round Trip Time (RTT) in LTE FDD is 8 ms (as described in Chapter 11) the corresponding LTE TDD RTT is at least 10 ms and up to 16 ms. The relatively small difference is because even for LTE FDD the core RTT is dominated by the UE and eNodeB HARQ processing times.

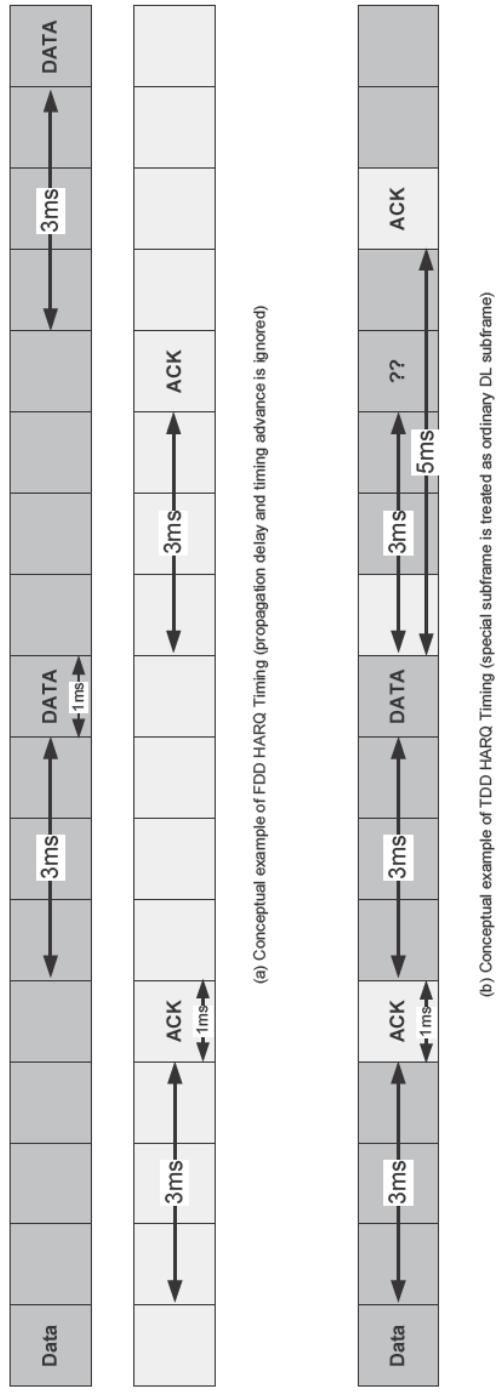


Figure 12.10 HARQ Timing: (a) LTE FDD; (b) LTE TDD

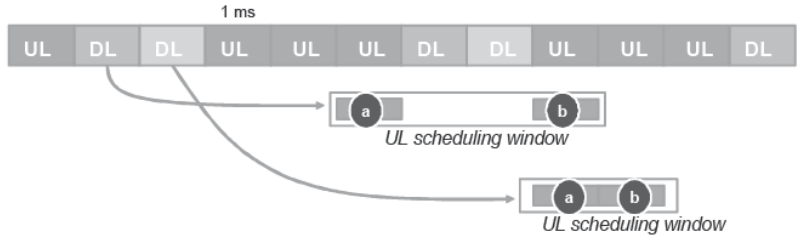


Figure 12.11 Possible uplink multi-TTI scheduling

12.3.4 HARQ Design for UL TTI Bundling

UL TTI bundling is used to improve the UL coverage by combining multiple UL TTIs for one HARQ process transmission. In a system where it is expected that many retransmissions are needed to successfully transmit a packet in uplink, UL TTI bundling provides fast automatic retransmission so that the delay penalty is minor. Although many details of UL TTI bundling are the same for LTE FDD and TDD, specifically the HARQ timing is different due to the inherited special UL HARQ timing in LTE TDD. In terms of the 3GPP discussion, only TDD UL/DL configurations #0, #1 and #6 have bundling fully defined in Release 8. To control the aspect of TTI bundling timing and HARQ process number after TTI bundling, the starting point of UL TTI bundling is limited and the number of bundled HARQ processes is fixed according to the bundle size. Thus the number of bundled HARQ processes is defined by:

$$Bundled\_HARQ\_No = \left\lceil \frac{Original\_HARQ\_No \times Bundled\_HARQ\_RTT}{Bundled\_size \times Original\_HARQ\_RTT} \right\rceil \quad (12.1)$$

In Equation 1, *Original_HARQ_No* and *Original_HARQ_RTT* are fixed for each TDD configuration in the LTE uplink system, so if assuming for instance TDD configuration #1 and a bundle size of 4, the number of bundled HARQ processes is 2, as shown in Figure 12.12.

As to the principle of ACK/NACK timing, it is always tied to the last subframe in a bundle, which is exactly the same as the rule in FDD. The uplink TTI bundling for LTE FDD is described in Chapter 10.

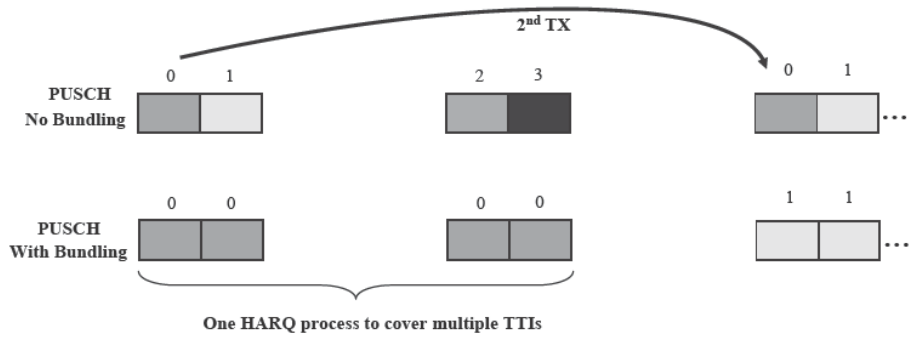


Figure 12.12 TTI bundling with TDD configuration #1

12.3.5 UL HARQ-ACK/NACK Transmission

In the same way as for FDD, the UL HARQ-ACK/NACK for LTE TDD is transmitted over the PHICH on PDCCH. The PHICH mapping and indexing is mostly the same for LTE FDD and TDD, i.e. the PDCCH in one DL subframe only contains the PHICH associated to a single UL subframe PUSCH. The only exception to this rule is the TDD UL/DL configuration #0 where the PHICH associated has specific exception specified.

12.3.6 DL HARQ-ACK/NACK Transmission

For both LTE FDD and TDD, the DL HARQ-ACK/NACK is transmitted on the PUCCH or the PUSCH depending on whether UL has simultaneous data transmission in the same UL subframe or not. In many cases the DL HARQ-ACK/NACK associated from more than one PDSCH, e.g. up to 9, will be mapped into a single UL subframe. This so-called multiple UL ACK/NACK transmission is, however, notably different from the FDD MIMO case in which the DL HARQ-ACK/NACK associated from a single PDSCH (e.g. two codewords) is mapped into a single UL subframe.

The fact that multiple downlink transmissions may need to be acknowledged within a single uplink subframe, makes the design for good UL coverage for control channels in TDD even more challenging. A very special design arrangement has been created to accomplish this task while simultaneously respecting the single carrier property of the UL multiple access scheme when UE has to transmit multiple DL HARQ-ACK/NACKs. There are two DL HARQ ACK/NACK feedback modes supported in TDD operation of LTE which are configured by a higher layer on a per-UE basis:

- ACK/NACK bundling feedback mode (the default mode), where a logical AND operation is performed per codeword's HARQ ACK/NACK across multiple DL subframes PDSCH whose associated HARQ ACK/NACK is mapped into the same UL subframe.
- ACK/NACK multiplexing feedback mode, where a logical AND operation is performed across spatial codewords within a DL HARQ ACK/NACK process. In Release 8 LTE TDD up to 4 bits DL HARQ ACK/NACK is supported per UL subframe, hence for UL/DL configuration #5 this feedback mode is not supported.

The ACK/NACK bundling feedback mode is the most aggressive mode to relieve the coverage problem of multiple UL ACK/NACK transmission in TDD. The allocated DL resources have been decoupled from the required UL feedback channel capability; i.e. only a single DL HARQ-ACK/NACK is transmitted in a single UL subframe regardless of the number of associated DL subframes carrying PDSCH for the user. The single ACK/NACK is then created by performing a logical AND operation over all associated HARQ ACK/NACK per UL subframe. This way, TDD has the same number of HARQ ACK/NACK feedback bits and thus transmission formats on PUCCH as FDD per UL subframe. The ACK/NACK encoding and transmission format in PUSCH is the same as it is in FDD.

Without proper compensation in the link adaptation and packet scheduling functions, the probability of DL HARQ NACK will increase causing more unnecessary DL retransmissions when using ACK/NACK bundling. Thus the control channel reliability is the key with ACK/NACK bundling. The second mode is more attractive for when the UE has sufficient UL cov-

erage to support multiple ACK/NACK bits on PUCCH. When in ACK/NACK multiplexing feedback mode, the status of each DL subframe HARQ-ACK/NACK, i.e. ACK, NACK or DTX (no data received in the DL subframe), one of the QPSK constellation points in certain derived PUCCH channels is selected for transmission on the UE side, and the eNodeB can decode the multi-bit HARQ ACK/NACK feedback by monitoring all constellation points from all associated PUCCH channels. The exact mapping table can be found in [3].

12.3.7 DL HARQ-ACK/NACK Transmission with SRI and/or CQI over PUCCH

When in ACK/NACK bundling mode, if both bundled HARQ-ACK/NACK and SRI are to be transmitted in the same UL subframe, the UE transmits the bundled ACK/NACK on its derived/assigned PUCCH ACK/NACK resource for a negative SRI transmission, or transmits the bundled HARQ-ACK/NACK on its assigned SRI PUCCH resource for a positive SRI transmission. This operation is exactly the same as for FDD.

When in ACK/NACK multiplexing mode, if both multiple HARQ-ACK/NACK and SRI are transmitted in the same UL subframe, the UE transmits the multiple ACK/NACK bits according to section 12.3.5 for a negative SRI transmission, and transmits 2-bit information mapped from multiple ACK/NACK input bits on its assigned SRI PUCCH resource for a positive SR transmission using PUCCH format 1b. The mapping between multiple ACK/NACK input bits and 2-bit information depends on the number of generated HARQ-ACK among the received DL subframe PDSCH within the associated DL subframes set K. The exact mapping table can be found in [3].

12.4 Semi-persistent Scheduling

Semi-persistent Scheduling (SPS) can be used with all the TDD UL/DL configurations. Many details of SPS are the same for LTE FDD and TDD, but in this section some TDD specific aspects related to SPS are detailed. The reader is referred to Chapter 10 for more information about SPS. To match the special frame structure of TDD, the SPS resource interval must be set to equal a multiple of the UL/DL allocation period (i.e. 10 ms) to avoid the conflict of non-matching UL/DL subframes because the UL subframe and DL subframe do not exist simultaneously. Furthermore, LTE UL uses synchronous HARQ and there are some problems for most UL/DL configurations in TDD because the HARQ RTT is 10 ms. When UL SPS is used for VoIP traffic (AMR codec periodicity is 20 ms), the second retransmission of a previous packet will collide with the next SPS allocation, since the period of SPS resource is two times the RTT. The collision case is shown in Figure 12.13. In the figure the numbers 1, 2 and 3 indicate different VoIP packets. For example, at the 20 ms point, the second retransmission of VoIP packet #1 collides with the initial VoIP packet #2. To solve this problem, two solutions are available to the network: dynamic scheduling and two-interval SPS patterns.

Although configured for SPS, the UE will anyway listen to dynamic allocations on the PDCCH. Such allocations will always override an existing persistent allocation. By using dynamic scheduling at the collision point it is possible to mitigate the problem as shown in Figure 12.14: if the UE is asked for a retransmission, the UE will perform a retransmission unless it has an empty buffer. With these definitions, if there are other following idle subframes available, the eNodeB will next schedule the retransmission in the current subframe, and

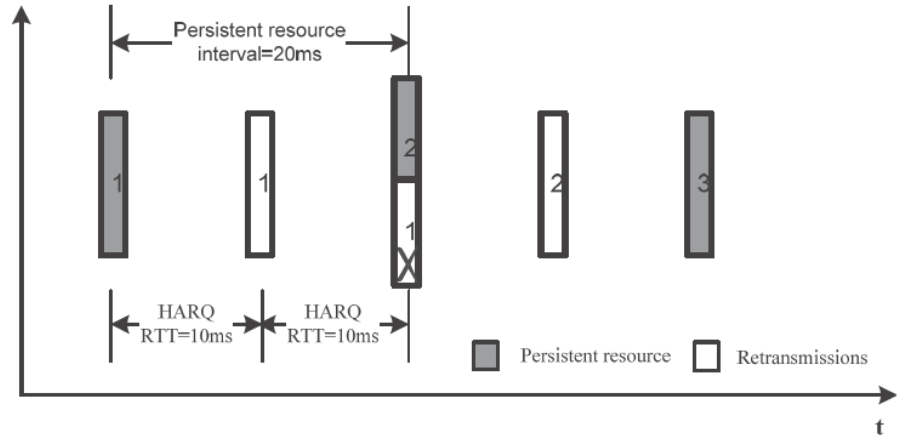


Figure 12.13 Collision between retransmissions and a new transmission

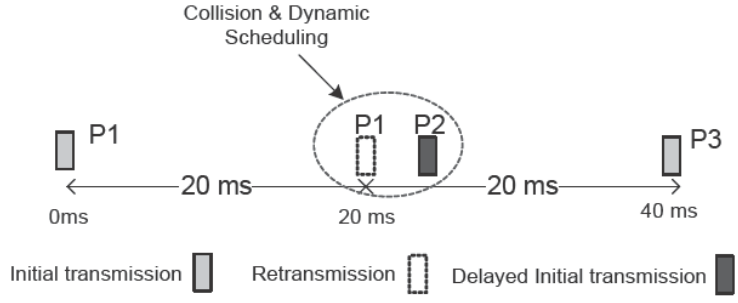


Figure 12.14 Dynamic scheduling at collision point

reschedule the initial transmission that was supposed to take place on the SPS resources in following subframes at the collision point.

The second solution is to use a two-interval SPS pattern. Here a two-interval SPS pattern means that two periods are used for semi-persistent scheduling while only one semi-persistent scheduling period is used and persistent allocation is carried out based on the pre-defined period in the conventional scheme. With a two-interval SPS pattern, a resource pattern with two different intervals (i.e. T1, T2, T1, T2 …) is used to avoid the main collision between the second retransmission of the previous packet and the SPS allocation. The procedure is given in Figure 12.15, in which T1 is not equal to T2 and the combined set of T1 and T2 is always a multiple of 10 ms. The offset between T1 and T2 is several subframes, and a variable subframe offset is used to indicate the offset. The following formulas are used to calculate T1 and T2:

$$T1 = SPS\ periodicity + subframe\_offset \tag{12.2}$$

$$T2 = SPS\ periodicity - subframe\_offset \tag{12.3}$$

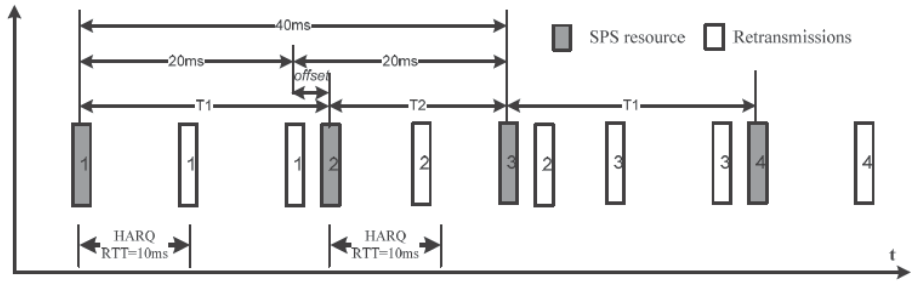


Figure 12.15 Two-interval SPS patterns

'*SPS periodicity*' will be signaled by RRC signaling; '*subframe_offset*' (positive value in Figure 12.15, but can also be negative) is implicitly defined according to the different TDD configurations and the starting point of a two-interval SPS pattern), and then T1 (the first time periodicity) and T2 (the second time periodicity) can be computed in terms of these equations. The allocation period always starts with the first time period T1. However, even configured with a two-interval SPS pattern, some residual collisions might still exist if the number of required retransmissions is large, i.e. 4. Any residual collisions can be avoided by means of dynamic scheduling as described earlier.

12.5 MIMO and Dedicated Reference Signals

LTE supports a number of different MIMO modes in DL, as already described in Chapter 5, covering both closed loop schemes with UE feedback information to the Node B. Together with the information of actually adapted downlink parameters from eNodeB this adds up to a significant amount of signaling to handle the DL closed loop MIMO. For TDD mode the earlier mentioned channel reciprocity can be exploited to mimic closed loop MIMO with a reduced signaling overhead. Under the assumption that the channel is identical for UL and DL we can estimate the channel in UL by using SRSs transmitted by the UE and then apply this channel knowledge for selecting the best DL pre-coding matrix. In this way UE feedback can be reduced or even avoided.

Going a step further, we can also eliminate the pre-coding matrix indication in the DL allocation message by using UE Specific Reference Signals (URS) to transfer this information. Moreover, the use of URS decouples the physical transmit antenna from the UE detection complexity and system overhead resulting from having a cell specific reference signal for each transmit antenna. URS are specified for LTE Release 8, and can be used for both FDD and TDD modes. However, this transmission mode is especially attractive when considered in a TDD setting where channel reciprocity is available.

URS are transmitted on antenna port 5 and they are generated with the same procedure as cell specific reference signals. The only difference is that the UE RNTI impacts the seed of the pseudo-random generator used to generate the code. The pattern for normal cyclic prefix and extended cyclic prefix can be seen in Figure 12.16(a) and Figure 12.16(b), respectively. There are 12 resource elements used for URS per PRB per 1 ms subframe so the additional overhead is rather large. With a normal cyclic prefix, cell specific reference signals on antenna port 0

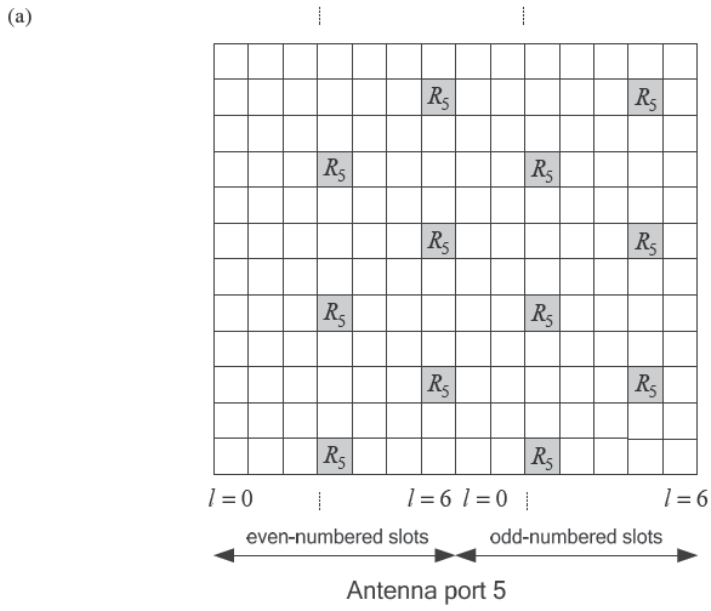


Figure 12.16 (a) URS pattern for normal CP; (b) URS pattern for extended CP

and 1 and a control channel region of three symbols antenna, enabling URS will reduce the number of resources for data transmission by 10%. On the other hand this gives a very robust performance at the cell edge or for high UE velocity.

One advantage of using URS is that pre-coding does not need to be quantified. As data and reference signals are using the same pre-coding matrix, the combined channel can be directly estimated from the UE specific reference signals and then used for demodulating the signal.

Due to the rather large overhead of URS, this mode is mainly expected to be used when Node B deploys more than four antennas. The Release 8 specifications do not allow for more than four different common reference signal ports, so in this case the only way to acheive UE specific pre-coding (beamforming) is to use URS.

An issue which is vendor specific and where eNodeB implementations may differ is how antennas are deployed and pre-coding is determined when URS are used. Two different scenarios will be discussed here. In the typical macro cell with antennas mounted above the roof-top, the azimuth spread of the channel would be low and the optimal solution would be to have antennas with narrow spacing and use traditional angular beamforming to create a narrow beam which directs the power towards the UE while reducing the interference to other UEs. In this case, the system only relies on estimating the Direction of Arrival (DoA) from the UL transmission and this can be done without relying on channel reciprocity. Moreover, the standard allows for reuse of time and frequency resources by other users in the same sector. Sufficient angular separation should be ensured to maintain performance.

In a scenario with many scatterers around the base station, the azimuth spread would be larger and angular beamforming might not work very well. Another solution could then be to rely on pre-coding determined from the eigen-vectors of the complex channel matrix determined from UL sounding. When the azimuth spread is increased, the rank of the channel could become larger than one and UE could potentially benefit from dual stream transmission. Although dual stream transmission is not supported in Release 8, it is a potential candidate for future releases. For more details on MIMO, beamforming and channel modeling see [4].

12.6 LTE TDD Performance

In this section the performance for LTE TDD is analyzed. For FDD mode, extensive analysis of performance has already been addressed in Chapter 9. As many of the observations provided there apply equally well to TDD mode, in the following we try to focus on the areas where TDD mode is different from FDD mode. First we look at the link performance, i.e. how well the receivers in eNodeB and UE can decode the physical channels. In general this is an area where there is little difference between FDD and TDD as reference signal patterns and channel coding are duplexing mode agnostic. After link performance, we discuss the link budget. Link budget for TDD is different from FDD because of the discontinuous transmission, so coverage for different bit rates in TDD will invariably be worse than for FDD. However, there are still a number of details to pay attention to in order to evaluate the TDD link budget properly.

This section ends with a discussion on system performance. First we look at the best effort type of service where a fairly small number of users per sector are assumed to download large amounts of data. Assuming that we are designing a network to deliver a certain bit rate for a certain number of active users, then the most important difference between networks based on FDD and TDD technologies is that UL and DL in TDD mode would need double the system bandwidth compared to FDD. Data transmission would also need to be carried out with a larger

bandwidth in the TDD system to achieve bit rates similar to FDD. Both system and transmission bandwidths affect the operation of RRM algorithms and in this the impact to the system performance is analyzed.

Performance assuming VoIP service is also evaluated and this gives a somewhat different perspective on the system performance as bit rates are low and users are many when we load the system with 100% VoIP traffic. As VoIP is a symmetric service, TDD systems can have a higher VoIP capacity than FDD because the split between UL and DL resources can be adjusted. The increased HARQ round trip time for the TDD system is also shown to have some effect on the VoIP performance where many UEs are coverage limited.

12.6.1 Link Performance

The two most important factors impacting the link performance for data transmission are channel estimation and channel coding. As reference signal design and channel coding are very similar for TDD and FDD, the link performance is also very similar. One source of difference is the discontinuous transmission in TDD. FDD receivers can use the reference signals from the previous subframe to estimate the channel. This is especially important for the DL link performance where the UE receiver should start decoding the control channel region as soon as it is received. In a TDD system, when a UE receiver decodes subframes transmitted right after UL→DL switching, it could not rely on reference signals from previous subframes and this could introduce some degradation to the channel estimation and thus coverage of control channels. The importance of such potential loss will depend on the UE implementation. If, for example, the UE could wait for the second column of reference signals then the performance degradation could be reduced.

Another potential difference in link performance between FDD and TDD is related to the special subframe. As explained earlier the guard period is created by adjusting the number of symbols in DwPTS. When the DwPTS length is reduced we also eliminate some reference signals as the rule is not to move them to new locations. The potential loss is minor as reference signals from a previous subframe could be taken into use to improve channel estimation. In the special case of UE specific reference signals we lose one column of reference signals even with full DwPTS length. In this case the UE cannot use reference signals from a previous subframe as these could have been transmitted with a different pre-coding.

Short RACH performance is clearly worse compared to 1 ms RACH preamble and thus should be used only in environments where link budget is not foreseen to be an issue.

12.6.2 Link Budget and Coverage for TDD System

The link budget calculation aims at estimating the range of different bit rates. A detailed description of how to calculate link budgets for LTE is already given in Chapter 9. Here we focus on the differences between link budgets for TDD and FDD modes. The differences relate mainly to the limited maximum UE transmit power and in the following we therefore focus our attention on UL link budgets.

The TDD UE cannot transmit continuously since the transmission must be switched off during the downlink reception, The UE will thus need to transmit with a larger bandwidth and a lower power density to achieve a similar bit rate to a FDD system. The lower power density is because the UE transmitter is limited on total maximum power, not on power per Hz.

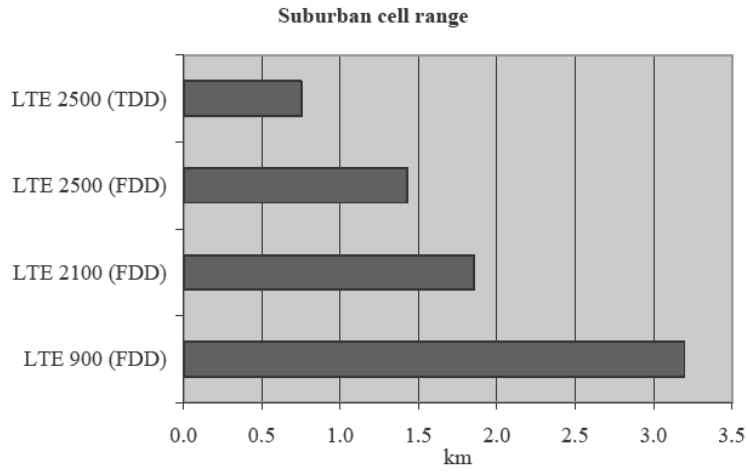


Figure 12.17 Uplink cell range for LTE FDD and TDD systems for 64 kbps

As a simple example, if the downlink : uplink share is 3 : 2, the UE transmission power density is reduced by $10 \times \log10(2/5) = -4$ dB as we need roughly 5/2 times the bandwidth for the TDD UL transmission. Another way of viewing this is that at a fixed distance from the base station the maximum achievable FDD bit rate will roughly be 2½ times larger than the bit rate achieved with maximum UE transmit power in a TDD system. Note that for DL, the power density can be assumed to be similar between FDD and TDD mode as the size of the power amplifier in eNodeB can be adapted to the system bandwidth.

So from a UL coverage perspective, FDD based systems do have an advantage over TDD systems due to the continuous transmission. Moreover, coverage with the TDD system can also be challenging because the TDD spectrum is typically situated at higher frequencies such as 2.3 GHz or 2.5 GHz. A cell range comparison for a suburban propagation environment is shown in Figure 12.17. The best coverage is obtained by using a FDD system at low frequency. The cell range for LTE900 FDD is four times larger (cell area 16 times larger) and LTE2500 FDD is 80% larger than LTE2500 TDD. The assumed data rate is 64 kbps and the cell range is calculated with the Okumura–Hata propagation model with 18 dB indoor penetration loss, 50 m base station antenna height and −5 dB correction factor. For maximum LTE coverage, LTE TDD deployment at high frequency could be combined with LTE FDD deployment at a lower frequency.

12.6.2.1 MCS Selection and UE Transmission Bandwidth for Coverage

For a certain target bit rate different combinations of MCS and transmission bandwidth have different coverage and different spectral efficiency. From Shannon's information theory [5] we know that if we want to maximize coverage under a fixed total transmission power constraint we should increase MCS when physical layer spectral efficiency (SE) is < 1 bit/s/Hz and increase bandwidth when SE is > 1 b/s/Hz. As adjusting BW does not impact SE, but increasing MCS does, the link should be operated with an MCS which achieves a SE of 1 bit/s//Hz (for LTE QPSK 2/3 would do the job). Then bandwidth can be adjusted to achieve the required bit rate with optimal coverage.

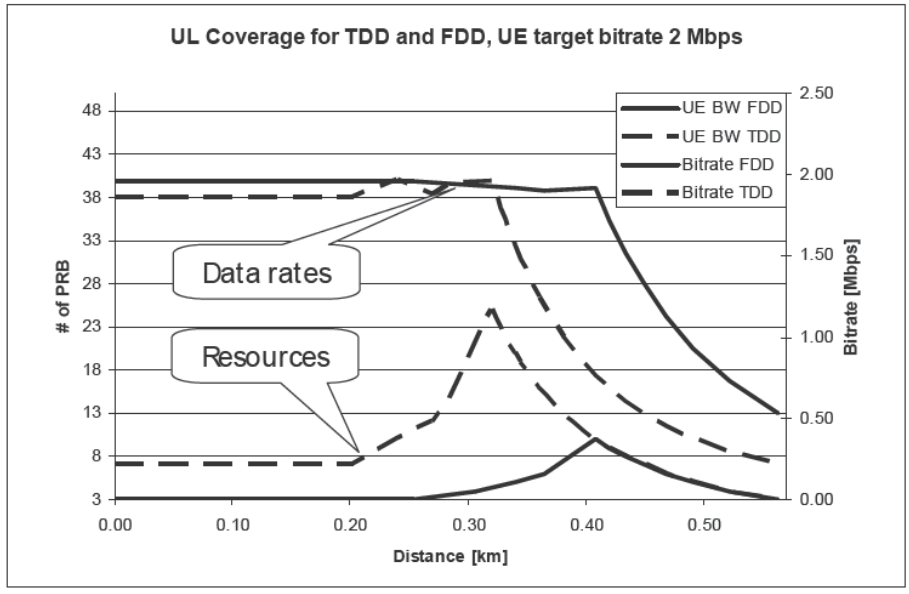


Figure 12.18 Coverage and required number of physical resource blocks for a 2 Mbps target bit rate for FDD and TDD UL. MCS in coverage limited region is QPSK ¾

In Figure 12.18 an example of this process is given. A UE at the cell edge selects MCS QPSK ¾ and sets the transmission bandwidth according to the available transmit power and required SINR. Curves for both TDD and FDD UE are shown and assumptions are as given in [6]. From the figure we can see that when the UE moves towards the Node B, path loss is reduced and the link gain can be used to increase the UE transmission bandwidth. As the transmission bandwidth increases, the UE bit rate also increases. When the UE target bit rate is reached, and if the UE path loss is reduced further, we can start to increase MCS and reduce bandwidth to improve the SE while keeping the bit rate on target. At some point the maximum MCS is reached and only at this point can we start to reduce the UE total transmit power. While the TDD system supports 2 Mbps in UL 300 m from the eNodeB, FDD system increases the coverage for 2 Mbps to 400 m. For more details on the interaction of UE transmit power, MCS and transmission bandwidth see Chapter 9.

12.6.2.2 Coverage for Low Bit Rates

When the data bit rate decreases, the relative overhead from header and CRC will increase. The effect of this is that even in FDD mode it does not make sense to schedule UEs in coverage problems with very narrow bandwidth and low MCS. An example for a data bit rate of 64 kbps is illustrated in Table 12.3. From this we can see that due to excessive overhead, the UE bandwidth increase for UL in a TDD 3DL/2UL configuration is only a factor 1.6, corresponding to a transmit power loss of 2 dB, not 4 dB as in the example given above where overhead was not taken into account.

Table 12.3 Required UE transmission bandwidth to support 64 kbps for TDD and FDD

System	TDD UL with 3DL/2UL	FDD UL
Service bit rate (kbps)	64	64
MCS	QPSK 1/10	QPSK 1/10
Data bits per TTI (bits)	160	64
Header (3 byte)	48	48
CRC (3 byte) (bits)		
Total bits per TTI (bits)	208	112
Payload per PRB in physical layer (2 symbols for DM RS 1 symbol for SRS)	27 bits	27 bits
Required number of PRBs	8	5

Table 12.4 Required SINR for VoIP in TDD and FDD with and without TTI bundling. UE transmission bandwidth assumed to be 360 kHz (2 PRB)

System	TDD (3DL/2UL or 2DL/3UL)		FDD	
Bundling enabled	No	Yes	No	Yes
Number of transmissions	5	8	7	12
Required SINR	−3.3 dB	−5.3 dB	−4.7 dB	−7.04 dB

One way of reducing the overhead for low bit rates is to use TTI bundling, as described in section 12.3.3. When TTI bundling is enabled the maximum number of retransmissions within a certain time limit is increased. In that way we can have more aggressive MCS selection and thus lower the relative overhead from protocol header and CRC. Due to the reduced number of UL TTIs in TDD, the potential link budget improvement from TTI bundling is not as important as for FDD. As shown in Table 12.4, for a VoIP service in 3DL/2UL configuration we can improve the number of transmissions of one VoIP packet within 50 ms from 5 to 8 TTIs, which is a 2 dB improvement to the link budget. We note that the link budget gain from TTI bundling in TDD mode is similar in both 2DL/3UL and 3DL/2ULconfiguration.

Finally, when operating at very low UL bit rates far away from the base station, coverage on UL control channels could also become a limiting factor. The control channel for TDD has been designed so that if ACK/NACK bundling mode is selected then the required receiver sensitivity is similar between TDD and FDD.

12.6.3 System Level Performance

TDD mode is in many aspects similar to FDD and this is also valid when we analyze the systems from a performance point of view. In general, when we compare time and frequency duplexing there are some differences related to the different ways of using the spectrum which make it difficult to make a 100% fair comparison of spectral efficiencies. Whereas a TDD mode system needs a guard period between UL and DL, a FDD system needs a large separation in frequency. Secondly if TDD systems have the same partition of UL and DL resources they can operate in adjacent bands; if not they need to be separated in a similar way to UL and DL for FDD.

Another spectrum related issue is that for the TDD system to provide a similar capacity to a FDD system the DL and UL system bandwidth needs to be double that of a FDD system. This impacts the RRM and users will typically need to operate with larger transmission bandwidths. That can be challenging for UL data transmission due to the limited transmission power of the UE.

One advantage for the TDD RRM solution is the possibility of exploiting channel reciprocity. In the current RRM framework (see Chapter 8) two parallel mechanisms are available for obtaining channel state information. For DL, the UE can be configured to feedback CQI, PMI and RI reports based on measurements of DL reference signals. In UL the UE can transmit SRSs so that Node B can measure the radio channel. For TDD mode, when channel reciprocity is present we ideally need only one of these mechanisms as the DL channel state can be inferred from the UL channel state or inversely. As mentioned earlier, there are challenges before this could work in a practical RRM solution, such as differences in UL/DL interference levels, different UL/DL antenna configurations and lack of UL/DL radio chain calibration, but on the other hand gains could be important. UL sounding, for example, can take up more than 10% of the UL system capacity. See section 12.2.4 for a further discussion of channel reciprocity.

12.6.3.1 Round Trip Time for TDD Systems

Feedback control loops are used for quite a few purposes in the LTE system. As TDD systems do not have continuous transmission and reception we might expect that the round trip time for such control loops would be increased for TDD systems, potentially degrading the system performance. However, due to the need for processing time, i.e. time for the UE or the Node B to decode and encode the control information, the typical round trip times between TDD and FDD are quite similar. Since the resulting delays are usually about 10 ms compared to 8 ms for the FDD RTT, the impact of the TDD frame structure is rather low and unlikely to impact on the performance of TCP/IP regardless of being run over LTE TDD or FDD.

12.6.3.2 Scheduling

One of the key LTE RRM features is channel aware scheduling, which is available for both UL and DL. Under the right conditions this feature can bring gains in spectral efficiency of up to 50% with even more important improvements to the coverage. To maximize the scheduling gain it is important to have frequency selective channel knowledge and flexibility in the control signaling to allocate UEs to the optimal frequency resources. Obtaining detailed frequency selective channel state information and enabling flexible resource allocation in the frequency domain is very costly in terms of control signaling.

For DL a number of different frequency resource allocation schemes are specified in the standard. For channel aware scheduling, the most effective allocation scheme specified gives a bit mask where each bit corresponds to a sub-band consisting of a number of continuous PRBs. To obtain the channel state for different sub-bands the CQI report contains a similar bit mask indicating the best M sub-bands and one indication of supported MCS on these sub-bands. The number of PRBs in one sub-band and the value of M are fixed in the specification. As the size of the bit mask increases with the system bandwidth, it was decided to increase sub-band size as the system bandwidth increases. If we compare a 10 MHz FDD DL and a 20 MHz TDD DL the scheduling gain is most likely to have some loss for 20 MHz especially for channels with a

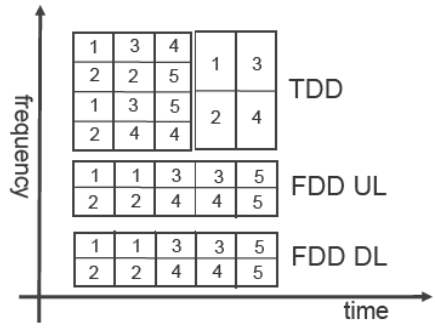


Figure 12.19 TDD and FDD resource allocations in LTE. Due to limitations on the number of users scheduled per TTI, TDD mode UEs are forced to transmit at a higher bandwidth

narrow frequency correlation bandwidth. The increase in system bandwidth could on the other hand give a TDD system an advantage due to the increased frequency diversity.

For UL due to the single carrier constraint, frequency resources have to be allocated contiguously; so here there is no difference for the resource allocation signaling between a lower and higher system bandwidth. For obtaining channel state information, we can enable SRS transmission from the UE. Sounding takes up resources and in contrast to the situation for DL, the larger the bandwidth that needs to be sounded, the more resources are needed. On the other hand it is not necessary to know the channel state information over the full bandwidth to obtain channel aware scheduling gain.

The resource allocation signaling consumes a lot of DL control channel resources and as these are limited the typical number of users that can be scheduled in one TTI is 10–20 (5–10 UL users and 5–10 DL users). As we assume that system bandwidth in TDD is double that of FDD, the TDD UE needs to operate at double bandwidth to have full bandwidth use in the system (see Figure 12.19). The increase in UE bandwidth can especially impact the UL channel aware scheduling gain. As the frequency resources are forced to be allocated contiguously, it is difficult to allocate resources according to channel quality in the frequency domain.

The total amount of DL control channel resources varies with the UL/DL ratio as up to three symbols can be reserved in each DL subframe (two for DwPTS). 2DL/3UL TDD potentially has fewer resources than in FDD while for other configurations there are more. The common case of 3DL/2UL has slightly more DL control channel resources than its FDD counterpart (8 symbols in 20 MHz for TDD compared to 15 symbols in 10 MHz for FDD) so there are no important differences between FDD and TDD. It should be mentioned that the TDD allocation messages (DCI) have a slightly higher payload, both due to higher system bandwidth but also due to additional bits for special features such as multi TTI and ACK/NACK bundling.

12.6.3.3 Uplink Power Control

The uplink power control is another very important RRM function as it has high impact on the distribution of bit rates in the sector. Moreover, fine tuning of power control parameters can improve both cell spectral efficiency and coverage. As we have discussed earlier the typical TDD UE will need to transmit at a higher bandwidth compared to an FDD UE. This puts extra

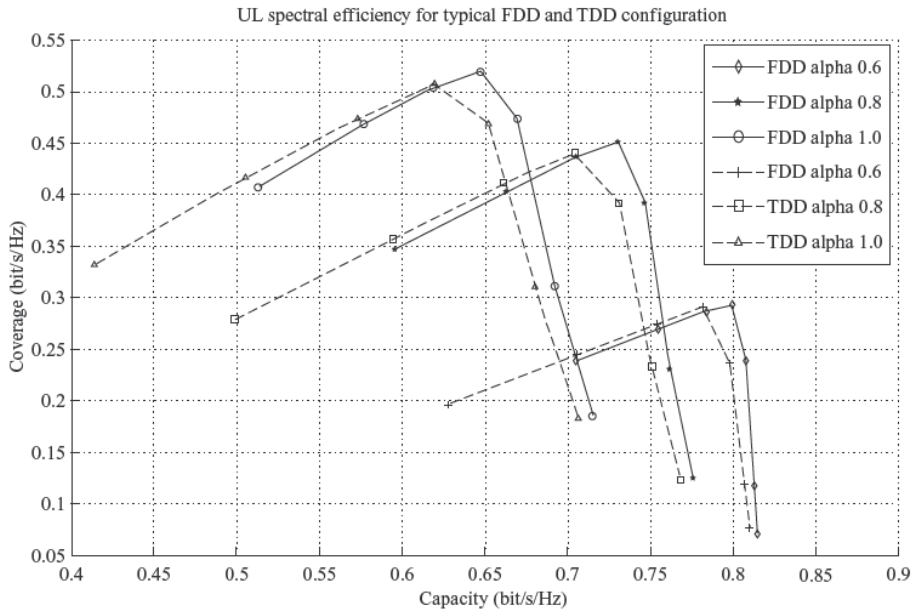


Figure 12.20 Uplink spectral efficiency for TDD and FDD systems with different power control settings. See [6] for assumptions

pressure on uplink transmit power resources, which are already limited, and can lead to poorer coverage for TDD systems. However, where interference limits the system performance, the performance difference is expected to be minor. In Figure 12.20 an uplink spectral efficiency comparison between a 20 MHz TDD system and a 2×10 MHz FDD system is shown where the TDD UEs transmit with a bandwidth of 12 PRB and FDD PRB with 6 PRB. From this figure we can see that both cell and coverage performance are very similar but slightly better for the FDD system. This difference is caused by reduced channel aware scheduling gain due to the higher bandwidth. Note that the optimal uplink interference level in the system depends on the bandwidth; for 12 PRB UE bandwidth it is 3 dB lower than for 6 PRB independent of the value of alpha. When site density decreases, the coverage performance of the TDD system will decrease faster than the FDD system and we will eventually see the effect from the differences in the link budget.

12.6.3.4 Downlink HARQ ACK/NACK Bundling

Finally we mention the effect on performance of the downlink HARQ ACK/NACK bundling feature described in section 12.3.5. As only one ACK/NACK is available for multiple subframes and these subframes are from different HARQ processes at different numbers of retransmissions then there is a risk that excessive retransmission will occur. If we consider particularly the very downlink heavy TDD configurations with aggressive MCS selection, then the resulting high retransmission probability will drag down the downlink system performance.

This effect can be offset by operating HARQ with relatively low BLER for the first transmission; however, this means that downlink HARQ will operate at a suboptimal operating point. Another way to reduce excessive retransmission from ACK/NACK bundling is to schedule users less often but with higher bandwidth. This on the other hand could reduce channel aware scheduling gain. In the most common TDD configurations with 3 downlink and 2 uplink subframes the impact from ACK/NACK bundling will be minor as a maximum of 2 ACK/NACKs will be bundled.

Simulations show that the loss in the case of 4 downlink/1 uplink is less than 10%. However, for users in coverage problems which rely on HARQ for the link budget, the excessive retransmission could give more important performance problems as one NACK typically will trigger retransmission of multiple TTIs.

12.6.3.5 Performance of VoIP

Voice over Internet protocol (VoIP) traffic should be supported as efficiently as possible in LTE TDD mode. However, supporting VoIP in LTE TDD faces very similar challenges to the FDD mode:

1 tight delay requirements with much longer HARQ RTT;
2 various control channel restrictions for different uplink/downlink configurations; and
3 even more serious coverage problems in uplink due to discontinuous transmission – as there are multiple uplink/downlink configurations possible, this can be used to achieve higher total VoIP capacity.

The following part presents the system level performance of VoIP traffic in LTE TDD mode at 5 MHz system bandwidth. The capacity evaluation criterion is the same as FDD: defined as the maximum number of per sector VoIP users that can be supported without exceeding a 5% outage level. VoIP capacity numbers are obtained from system level simulations in macro cell scenario 1 and macro cell scenario 3 [2], the main system simulation parameters being aligned with [7].

Similar to FDD in Chapter 10, the simulation results of VoIP capacity are summarized in Table 12.5 and Table 12.6 for two different AMR codecs, AMR 7.95 and AMR 12.2, and for

Table 12.5 VoIP capacity in LTE TDD at 5 MHz for macro case 1

	Configuration 0 (2DL/3UL)		Configuration 1 (3DL/2UL)	
VoIP codec	AMR 12.2	AMR 7.95	AMR 12.2	AMR 7.95
Downlink capacity				
Dynamic scheduler, without packet bundling	64	64	112	112
Dynamic scheduler, with packet bundling	114	122	194	206
Semi-persistent scheduler	102	140	168	220
Uplink capacity				
Dynamic scheduler	88	92	88	114
Semi-persistent scheduler	134	174	86	110

Table 12.6 VoIP capacity in LTE TDD at 5 MHz for macro case 3

	Configuration 0 (2DL/3UL)		Configuration 1 (3DL/2UL)	
Downlink capacity				
VoIP codec	AMR 12.2	AMR 7.95	AMR 12.2	AMR 7.95
Dynamic scheduler, without packet bundling	54	54	94	94
Dynamic scheduler, with packet bundling	90	94	158	168
Semi-persistent scheduler	84	114	144	194
Uplink capacity				
VoIP codec	AMR 12.2		AMR 12.2	
Dynamic scheduler, without TTI bundling	< 40 (50 ms delay bound)		< 40 (50 ms delay bound)	
Dynamic scheduler, with TTI bundling	< 40 (50 ms delay budget)		< 40 (50 ms delay budget)	
	50 (60 ms delay budget)		50 (60 ms delay budget)	
	72 (70 ms delay budget)		54 (70 ms delay budget)	

both dynamic and semi-persistent schedulers and for both macro case 1 and macro case 3 as defined in [6].

For both UL and DL simulations, the number of control symbols used in the special subframe is set to two, which contains six Control Channel Elements (CCEs). In a normal DL subframe, there are three control symbols available, which provide ten CCEs for UL and DL control.

For DL VoIP capacity for LTE TDD, performance of fully dynamic PS is seriously control channel limited if packet bundling is not allowed for both simulation scenarios. However, with packet bundling the users having good CQI are scheduled less often and hence the capacity can be boosted up to 70–90% because more control channel resources are released to be used by other users. There is no difference in the packet bundling concept between FDD and TDD. The reader is referred to Chapter 10 for more information about packet bundling. Semi-persistent PS performs very well in control channel limited situations, and always outperforms dynamic PS if packet bundling is not allowed. Fully dynamic PS outperforms semi-persistent PS only if packet bundling is allowed and the VoIP packet payload is high enough (e.g. AMR12.2 or higher) so that control channel limitations can be avoided. The performance difference between scenario 1 and scenario 3 is up to 15–20%, which is not so large because DL transmission power is not so limited even in scenario 3.

For UL VoIP capacity for LTE TDD, configuration 0 is regarded as seriously control channel limited, so the performance of dynamic scheduling is quite poor due to the limited number of control channels available. Performance loss reaches 40–80% compared to semi-persistent scheduling for different AMR codecs. In contrast, configuration 1 is a loose control channel case, so the performance of dynamic scheduling is slightly better than semi-persistent scheduling due to flexible retransmission and frequency domain packet scheduling (FDPS) gain. The performance difference between macro case 1 and macro case 3 at over 50% is a huge loss. The loss is very serious because macro case 3 is a very coverage-limited scenario. The UE transmit power is not large enough to even provide a proper link budget. For macro case 3, TTI bundling and longer delay bound can provide great gains on performance. For example, by using TTI bundling in a 70 ms delay budget for TDD configuration 0, up to 3 dB of energy accumulation

gain can be achieved. However, without TTI bundling, it is hard to achieve good VoIP capacity (i.e. < 40 users/sector) because of the deficient energy accumulation.

To carry out a fair comparison between FDD and TDD, the FDD capacity results should be scaled for different TDD configurations and scheduling options. In DL, the scaling factor used for the dynamic scheduler should depend on the control symbol assumptions because it has control overhead limited performance for dynamic scheduling. With two control symbols, special subframe only has six CCEs available. Hence FDD results for the dynamic scheduler should be scaled with the scaling factor 26/50 = 0.52 for configuration 1 and 16/50 = 0.32 for configuration 0 in DL. However, in DL the scaling factor used for SPS should depend on the number of data symbol assumptions because it has data limited performance for SPS. With two control symbols, the special subframe only has seven data symbols available. Hence, FDD results for SPS should be scaled with the scaling factor 27/50 = 0.54 for configuration 1 and 17/50 = 0.34 for configuration 0.

In UL, there are no data symbols available in the special subframe, only UL control signaling, so we just use the scaling factors 0.6 and 0.4 for configuration 0 and configuration 1 simply in terms of the UL data symbol ratio.

Compared with the scaled FDD results in Chapter 10, we make the following observations. In downlink macro case 1, TDD results for both configurations are at most 5% lower than the scaled FDD results. For dynamic scheduling in the coverage limited macro case 3, the average amount of transmissions per packet is increased. As scheduling flexibility in the time domain is decreased for TDD and HARQ RTT is increased, the scaled TDD performances are about 10% lower than the scaled FDD results – losses are naturally higher for configuration 0 due to the small number of DL subframes.

In UL TDD configuration 0 of macro case 1, FDD has about a 7–10% gain over TDD for semi-persistent scheduling due to more diversity gain in FDD. Moreover, FDD has up to 40% gain over TDD for dynamic scheduling due to the limited control channel resources in TDD. For uplink TDD configuration 1 of macro case 1, TDD has 5–20% gain for AMR 12.2 and AMR 7.95 respectively over FDD for dynamic scheduling due to loose control channel limitation. FDD has ~10% gain for AMR 12.2 and AMR 7.95 respectively over TDD for semi-persistent scheduling due to more diversity.

For macro case 3 with TTI bundling and with a slightly longer packet delay budget (i.e. 70 ms), the difference between FDD and TDD reaches 15–25% for different TDD configurations because there is less energy accumulation in the TDD mode. However, with the shorter delay budget (i.e. 50 ms), TTI bundling is not able to provide visible capacity improvement although 2 dB energy accumulation gain is achieved compared to no bundling. Table 12.7 shows the energy accumulation with a 50 ms packet delay budget. With TTI bundling, a 1.76 dB energy accumulation gain can be achieved in FDD compared to the value in TDD. So the improvement for VoIP capacity is obvious.

Table 12.7 Impact from energy accumulation with TTI bundling

	Configuration 0 (2DL/3UL)	Configuration 1 (3DL/2UL)	FDD
Without bundling	5	5	7
With bundling	8	8 (for most of packets)	12

12.6.4 Evolution of LTE TDD

As mentioned earlier, TDD provides a very efficient way of enabling large peak data rates in unpaired spectrum conditions or when paired spectra with significant bandwidth or duplexing distance can be found. As LTE progresses towards LTE Advanced, introduced in Chapter 2, local area deployment and the outlook for reaching 1 Gbit/s data rates will become a key driver in further developing the TDD mode. In the path towards unleashing the transmission over a very wide bandwidth (e.g. multiple component carriers), TDD poses some interesting challenges for scalability while ensuring backwards compatibility for a smooth evolution.

Another important aspect related to TDD is that of coexistence. Based on Release 8 it seems that synchronization is the key to control interference and provide a reliable system operation in a network with full coverage for UE. However, for indoor or low-cost deployments, the existing techniques for synchronization may become too expensive and hence techniques for more lean synchronization, e.g. Over The Air (OTA) synchronization, could be of interest.

An aspect related to TDD is that of providing dynamic switching of the uplink/downlink configuration depending on the required capacity in the network. It is not yet known if the 3GPP LTE TDD standard will in the future provide a direct means for multiple operators to coexist within the same geographical area or if such coexistence will be guaranteed mainly by means of regulation. A further step in this direction is to look at protocols for slowly or semi-statically modifying the TDD configuration according to the varying load over time either within a coverage area covered by a single operator or in an area with multiple operators, in which case semi-static TDD adaptation would need to be considered together with schemes for flexible spectrum usage and sharing. Finally, it has yet to be seen if fast dynamic TDD for capturing the instantaneous load in a single cell will become possible in practice.

Finally, given the experience with the first deployments it will be interesting to follow to what extent reciprocity can be assumed to be available taking the practical antenna aspects into account. There are several ways in which the LTE TDD standard could be further optimized towards more efficient signaling (and thus performance, given fixed signaling overhead budget) in the case of reciprocity including topics coming as part of the LTE-Advanced work in upcoming 3GPP releases.

12.7 Summary

The LTE TDD operation is very similar to LTE FDD, including exactly the same multiple access solutions, channel encoding, etc. From the 3GPP specifications perspective the close similarity between FDD and TDD is also demonstrated via a single set of specifications for the physical layer. The resulting differences are due to the uplink/downlink split inherent for the TDD operation and cannot be really avoided as in FDD one does not need to change transmission direction but frames are continuous. Still as the resource allocation was 1 ms resolution both in FDD and TDD, the slots, excluding the ones when transmission direction changes, have the same parameters and timing. The sharing of the same spectrum for uplink and downlink gives flexibility for data asymmetry and makes finding new spectrum allocations easier for IMT-Advanced which needs to reach the 1 Gbit/s target. On the other hand the nature of the TDD operation sets the requirements for using a synchronized network to avoid uplink/downlink interference as well as resulting in a shorter range than with a FDD configuration. The LTE TDD mode (or TD-LTE) deployment has also been designed to accommodate coexistence with TD-SCDMA by ensuring compatible parameterization for uplink/downlink split when

sharing the same sites and antennas or for other close proximity installations. The spectrum availability for the TDD operation varies on a country basis, but in China the spectrum has been allocated and also in Europe an additional TDD spectrum (on top of the older allocations like 1900 to 1920 MHz) is being allocated from 2.5 GHz in several countries (with first licensing activities already completed). Also in the 3.5 GHz spectrum there have been recent allocations in certain countries with some unpaired parts, as in some countries there are existing deployments on that band for fixed mobile broadband use. Further development of LTE TDD mode proceeds together with the LTE FDD mode towards the LTE-Advanced to meet the ITU-R IMT-Advanced requirements.

References

[1] 3GPP Technical Specification, TS 36.211 V8.4.0, September 2008.
[2] 3GPP Technical Report, TR 25.814, 'Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA)', 3GPP TSG RAN, September 2006.
[3] 3GPP Technical Specification, TS 36.213 V8.4.0 September 2008.
[4] C. Oestges, B. Clerckx, 'MIMO Wireless Communications', 1st edition, Academic Press, 2007.
[5] J.G. Proakis, 'Digital Communications', 3rd edition, McGraw-Hill Book Co., 1995.
[6] CATT, 'Enhanced Beamforming Technique for LTE-A', 3GPP T-doc R1-082972.
[7] 'Next Generation Mobile Networks (NGMN) Radio Access Performance Evaluation Methodology', A White paper by the NGMN Alliance, January 2008.

13

HSPA Evolution

Harri Holma, Karri Ranta-aho and Antti Toskala

13.1 Introduction

High Speed Packet Access (HSPA) was included in Third Generation Partnership Project (3GPP) Releases 5 and 6 for downlink and for uplink. The 3GPP Releases 7 and 8 have brought a number of HSPA enhancements providing major improvements to the end user performance and to the network efficiency. The work continues further in Release 9.

The HSPA evolution work has progressed in parallel to LTE work in 3GPP. HSPA evolution deployments in practice take place in parallel to LTE deployments. Many of the technical solutions in HSPA evolution and LTE are also similar. The overview of the HSPA evolution and LTE roles is illustrated in Figure 13.1. HSPA evolution is optimized for coexistence with WCDMA/HSPA supporting legacy Release 99 UEs on the same carrier and designed for a simple

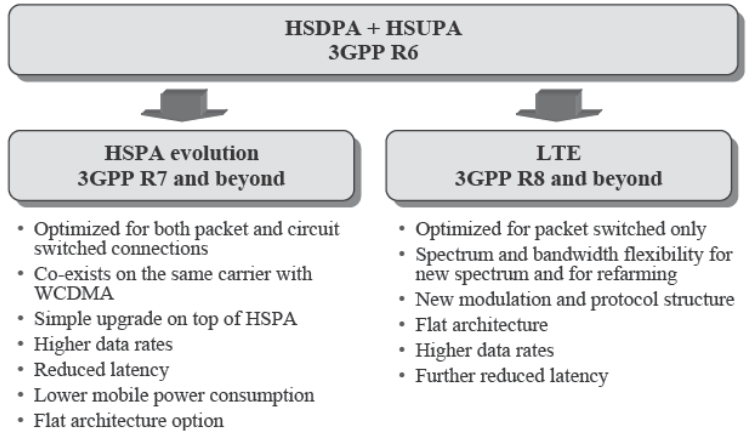


Figure 13.1 Overview of HSPA evolution and LTE roles

LTE for UMTS: OFDMA and SC-FDMA Based Radio Access Edited by Harri Holma and Antti Toskala
 ISBN: 978-0-470-99401-6

upgrade on top of HSPA. The HSPA evolution target is to improve end user performance by lower latency, lower power consumption and higher data rates. The HSPA evolution features are introduced in this chapter. The HSPA evolution is also known as HSPA+.

HSPA evolution also includes inter-working with LTE, which enables both packet handovers and voice handovers from LTE Voice over IP (VoIP) to HSPA circuit switched (CS) voice. The handovers are covered in Chapter 7 and the voice call continuity in Chapter 10.

13.2 Discontinuous Transmission and Reception (DTX/DRX)

Technology evolution in general helps to decrease the mobile terminal power consumption. Also, fast and accurate power control in WCDMA helps to minimize the transmitted power levels. The challenge in 3GPP from Release 99 to Release 6 is still the continuous reception and transmission when the mobile terminal is using HSDPA/HSUPA in Cell_DCH (Dedicated Channel) state. HSPA evolution introduces a few improvements to HSDPA/HSUPA that help to reduce the power consumption for CS voice calls and for all packet services.

With 3GPP Release 6 the UE keeps transmitting the physical control channel even if there is no data channel transmission. The control channel transmission and reception continues until the network commands the UE to Cell_FACH (Forward Access Channel) or Cell_PCH (Paging Channel) state. The Release 7 UE can cut off the control channel transmission as soon as there is no data channel transmission, allowing it to shut down the transmitter completely. This solution is called discontinuous uplink transmission and it brings clear savings in transmitter power consumption [1].

A similar concept is also introduced in the downlink where the UE needs to wake up only occasionally to check if the downlink data transmission is starting again. The UE can use power saving mode during other parts of the frame if there are no data to be received. This solution is called downlink discontinuous reception. The discontinuous transmission concept is illustrated in Figure 13.2 for web browsing. As soon as the web page is downloaded, the connection enters discontinuous transmission and reception.

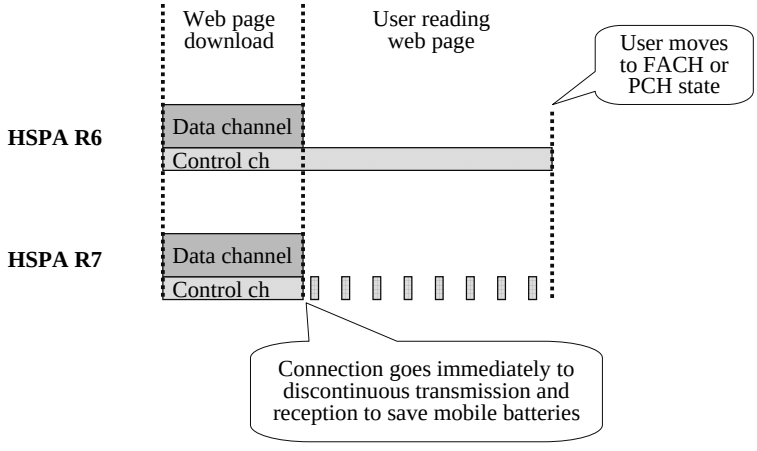


Figure 13.2 Discontinuous transmission and reception with continuous packet connectivity

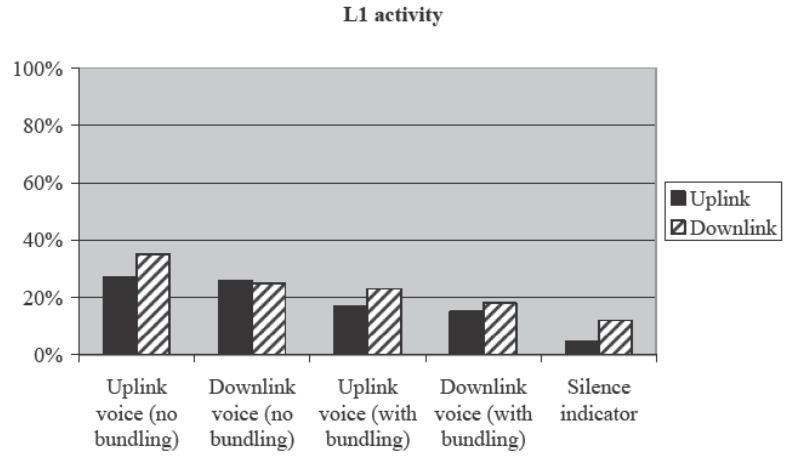


Figure 13.3 Layer 1 activity for voice call

Layer 1 (L1) activity with DTX/DRX for voice service is estimated in Figure 13.3. The packet bundling refers to the solution where two voice packets are transmitted together in the air interface halving the uplink active time as the UE needs to actually send data in the uplink only once every two voice packets arriving from the voice encoder. The activity is reduced from 100% in Release 6 to 20–40% in HSPA evolution. This activity reduction can translate into >50% longer voice talk time. The usage time improvement can be even more for bursty packet data connections.

The Release 99 FACH solution requires continuous reception by the UE, which is challenging from the power consumption point of view especially for always-on applications transmitting frequent keep alive messages. Each keep alive message forces the UE to move to the Cell_FACH state and stay there until the network inactivity timer expires. The discontinuous reception is also introduced for Cell_FACH state in HSPA evolution helping stand-by times when these types of applications are used.

Discontinuous transmission and reception is included in LTE from the beginning in Release 8 specifications. The power saving potential in LTE is even higher than in HSPA because the Transmission Time Interval (TTI) size is shorter than in HSPA (1 ms vs 2 ms) and because there is no need for fast power control related signaling.

13.3 Circuit Switched Voice on HSPA

Voice has remained an important service for the mobile operators. WCDMA Release 99 supports CS voice on Dedicated Channel (DCH) with quite high spectral efficiency. Efficient Voice over IP (VoIP) capability on top of HSPA was defined in Release 7, but VoIP mass market has not yet started. Therefore, CS voice over HSPA was also defined in HSPA evolution. CS voice over HSPA is part of Release 8 but because of the capability indication for the UE, support of the feature was introduced to Release 7, hence it is possible to implement this feature before other Release 8 features. There are two main benefits when running voice on HSPA: UE power

consumption is reduced because of DTX/DRX and the spectral efficiency is improved with HSPA features. Now these benefits can also be used for CS voice calls.

The different 3G voice options are illustrated in Figure 13.4. CS voice over DCH is used currently in commercial networks. Dedicated Release 99 channel is used in L1 and Transparent mode RLC in Layer 2 (L2). From the radio point of view, CS voice over HSPA and VoIP over HSPA use exactly the same L1 VoIP including unacknowledged mode RLC on L2. IP header compression is not needed for CS voice. From the core network point of view, there is again no difference between CS voice over DCH and CS voice over HSPA. In fact, CS core network is not aware if the radio maps CS voice on DCH or on HSPA. CS voice over HSPA could be described as CS voice from the core point of view and VoIP from the radio point of view.

CS voice over HSPA brought the following changes to 3GPP specifications:

- Iu interface = no changes
- Physical layer = no changes
- MAC layer = no changes
- RLC layer = forwarding RLC-UM sequence numbers to upper layers
- PDCP layer = modification of header to identify and timestamp the CS AMR frames + interfacing/inclusion of Jitter Buffer Management
- New de-jitter buffer = UE/RNC implementation dependent entity that absorbs the radio jitter created by HSPA operation so that the CS AMR frames can be delivered in a timely constant fashion to upper layers. The algorithm for the de-jitter buffer is not standardized in 3GPP.

The CS voice over HSPA concept is presented in Figure 13.5. The CS voice connection can be mapped on DCH or on HSPA depending on the UE capability and RNC algorithms. The AMR data rate adaptation can be controlled by RNC depending on the system loading. When CS voice over HSPA is used, there is a clear need for QoS differentiation in HSPA scheduling to guarantee low delays for voice packets also during the high packet data traffic load. Since the packet scheduling and the prioritization are similar for VoIP and for CS voice over HSPA, it will be simple from the radio perspective to add VoIP support later on top of CS voice over HSPA. Therefore, CS voice over HSPA is paving the way for future VoIP introduction. The similar radio solutions also make the handover simpler between VoIP and CS domains.

CS voice on HSPA can take advantage of IP protocol and packet transmissions in all interfaces: air interface carried by HSPA, Iub over IP, Iu-CS over IP and the backbone between

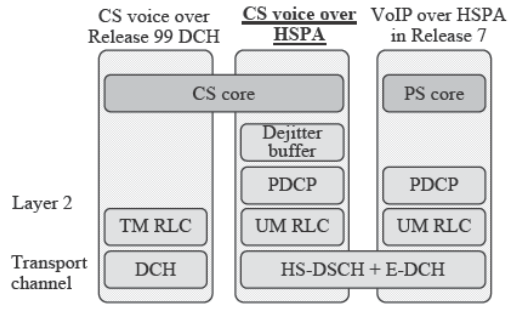


Figure 13.4 Voice options in WCDMA/HSPA

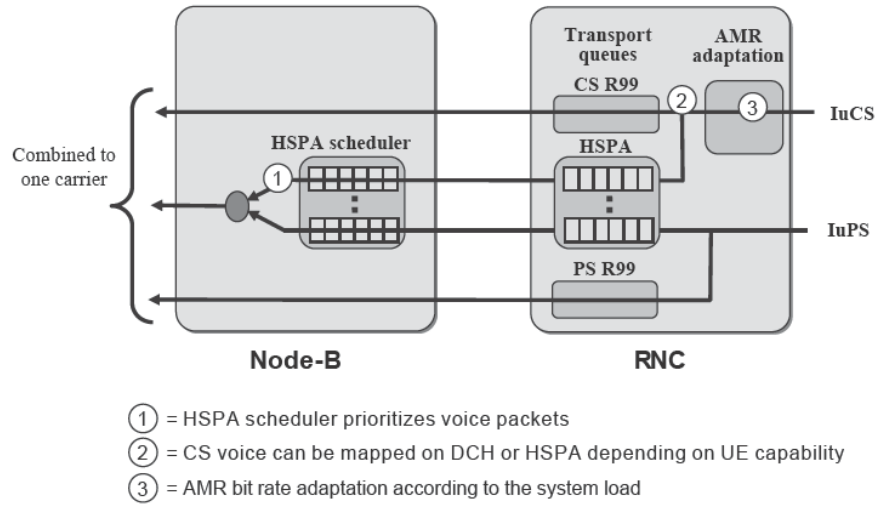


Figure 13.5 CS voice over HSPA overview

Media Gateways (MGW) using IP. The call control signaling is still based on CS protocols [2]. The use of the interfaces is shown in Figure 13.6. CS voice over HSPA can take advantage of the packet network performance and cost while maintaining the existing end-to-end protocols and ecosystem. No changes are required to charging, emergency calls or to roaming.

CS voice over HSPA also increases the spectral efficiency compared to voice over DCH because voice can also take advantage of the HSPA physical layer enhancements:

- UE equalizer increases downlink capacity. Equalizer is included in practice in all HSPA terminals [3].
- Optimized L1 control channel in HSPA reduces control channel overhead. The downlink solution is Fractional DPCH with discontinuous reception and the uplink solution is discontinuous transmission. Also transmission without HS-SCCH can be used on downlink.
- L1 retransmissions can also be used for voice on HSPA since the retransmission delay is only 14 ms.
- HSDPA optimized scheduling allows improvement of the capacity even if the tough delay requirements for voice limit the scheduling freedom compared to best effort data.

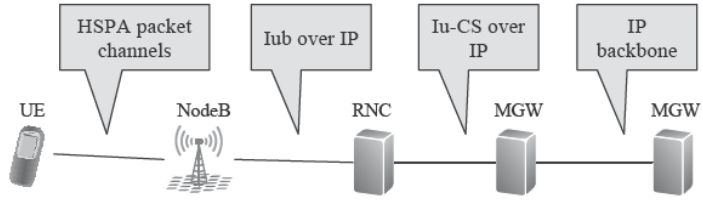


Figure 13.6 CS voice in different interfaces

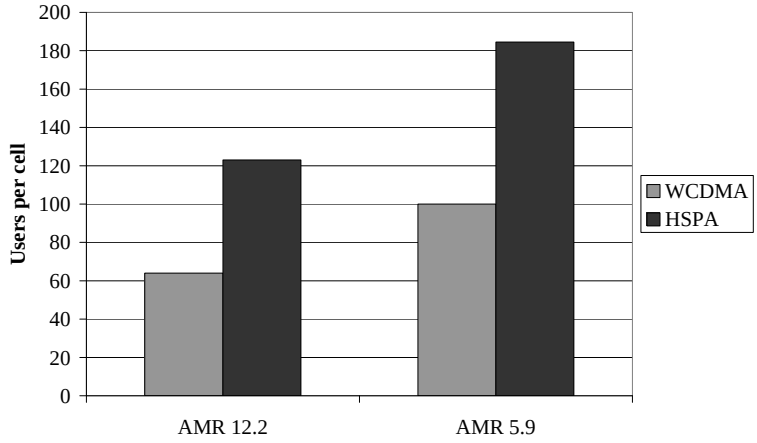


Figure 13.7 Circuit switched voice spectral efficiency with WCDMA and HSPA

The gain in spectral efficiency with CS voice over HSPA is estimated at 50–100% compared to CS voice over DCH. The voice capacity is illustrated in Figure 13.7. The voice capacity evolution including LTE is covered in more detail in Chapter 10.

The voice codec with CS over HSPA can be Narrowband Adaptive Multirate Codec (AMR) or Wideband AMR.

13.4 Enhanced FACH and RACH

WCDMA network data rate and latency are improved with the introduction of Release 5 HSDPA and Release 6 HSUPA. The end user performance can be further improved by minimizing the packet call setup time and the channel allocation time. The expected packet call setup time with Release 7 will be below 1 s. Once the packet call has been established, user data can flow on HSDPA/HSUPA in Cell_DCH state. When the data transmission is inactive for a few seconds, the UE is moved to the Cell_PCH state to minimize the mobile terminal power consumption. When there are more data to be sent or received, the mobile terminal is moved from Cell_PCH to Cell_FACH and to the Cell_DCH state. Release 99 Random Access Channel (RACH) and FACH can be used for signaling and for small amounts of user data. The RACH data rate is very low, typically below 10 kbps, limiting the use of the common channels. Release 5 or Release 6 do not provide any improvements in RACH or FACH performance. The idea in Release 7 Enhanced FACH and Release 8 Enhanced RACH is to use the Release 5 and Release 6 HSPA transport and physical channels also in the Cell_FACH state for improving the end user performance and system efficiency. The concept is illustrated in Figure 13.8. Enhanced FACH and RACH bring a few performance benefits:

- RACH and FACH data rates can be increased beyond 1 Mbps. The end user could get immediate access to relatively high data rates without the latency of channel allocation.
- The state transition from Cell_FACH to Cell_DCH would be practically seamless. Once the network resources for the channel allocation are available, a seamless transition can take place to Cell_DCH since the physical channel is not changed.

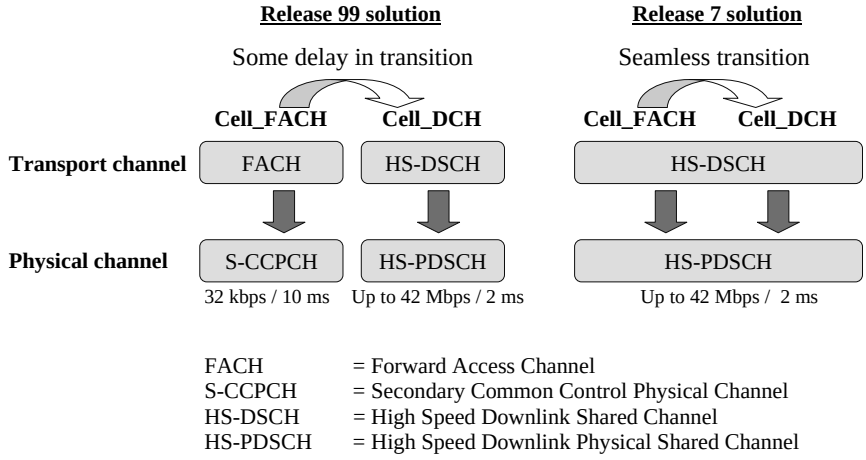


Figure 13.8 Enhanced FACH concept [4]

- Unnecessary state transitions to Cell_DCH can be avoided when more data can be transmitted in Cell_FACH state. Many applications create some background traffic that is today carried on Cell_DCH. Therefore, Enhanced RACH and FACH can reduce the channel element consumption in NodeB.
- Discontinuous reception could be used in Cell_FACH to reduce the power consumption. The discontinuous reception can be implemented since Enhanced FACH uses short 2 ms TTI instead of 10 ms as in Release 99. The discontinuous reception in Cell_FACH state is introduced in 3GPP Release 8.

Since the existing physical channels are used in Enhanced FACH, there are only minor changes in L1 specifications which allow fast implementation of the feature. Enhanced FACH can coexist with Release 99 and with HSDPA/HSUPA on the same carrier. No new power allocation is required for Enhanced FACH since the same HSDPA power allocation is used as for the existing HSDPA.

13.5 Downlink MIMO and 64QAM

The downlink peak data rate with Release 6 HSDPA is 10.8 Mbps with ¾ coding and 14.4 Mbps without any channel coding. There are a number of ways in theory to push the peak data rate higher: larger bandwidth, higher order modulation or multi-antenna transmission with Multiple Input Multiple Output (MIMO). All these solutions are part of HSPA evolution. MIMO and higher order modulation are included in HSPA evolution in Release 7 and Dual carrier (dual cell) HSDPA in Release 8. The 3GPP MIMO concept uses two transmit antennas in the base station and two receive antennas in the terminal and uses a closed loop feedback from the terminal for adjusting the transmit antenna weighting. The diagram of the MIMO transmission is shown in Figure 13.9.

Higher order modulation allows higher peak bit rate without increasing the transmission bandwidth. Release 6 supported Quadrature Phase Shift Keying (QPSK) and 16QAM (Quadrature Amplitude Modulation) transmission in the downlink and dual Binary Phase Shift

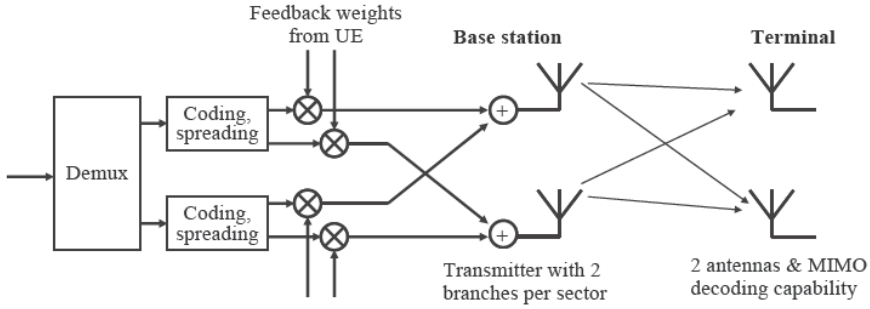


Figure 13.9 2×2 MIMO transmission concept

Keying (BPSK) in the uplink. Dual-channel BPSK modulation is similar to QPSK. Release 7 introduces 64QAM transmission for the downlink and 16QAM for the uplink. 16QAM can double the bit rate compared to QPSK by transmitting 4 bits instead of 2 bits per symbol. 64QAM can increase the peak bit rate by 50% compared to 16QAM since 64QAM transmits 6 bits with single symbol. On the other hand, the constellation points are closer to each other for the higher order modulation and the required signal-to-noise ratio for correct reception is higher. The difference in the required signal-to-noise ratio is approximately 6 dB between 16QAM and QPSK and also between 64QAM and 16QAM. Therefore, downlink 64QAM and uplink 16QAM can be used only when the channel conditions are favorable.

The system simulation results with 64QAM in macro cells are illustrated in Figure 13.10. The 64QAM modulation improves the user data rate with 10–25% probability depending on the scheduling (RR = round robin, PF = proportional fair). The rest of the time the channel conditions are not good enough to enable the reception of 64QAM modulation. The typical capacity gain from 64QAM is less than 10%.

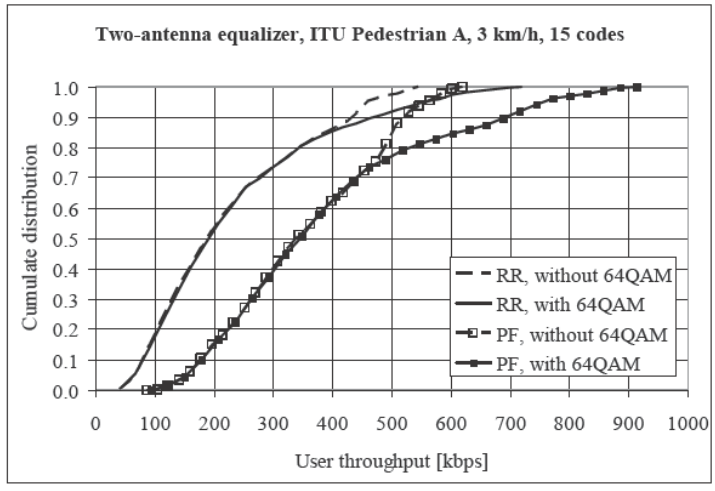


Figure 13.10 Macro cell data rates per user with 64QAM with 20 active users in a cell [5]

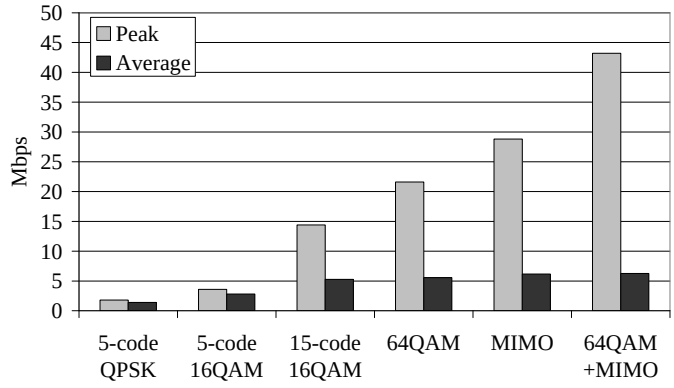


Figure 13.11 Downlink peak rate and average cell rate evolution with HSDPA features

64QAM and MIMO together improve the peak rate by 200% from 14 Mbps to 42 Mbps, but the average cell capacity is only improved by 20% since those high data rate features are useful only for part of the cell area. 64QAM activation can improve the cell capacity by approximately 5% and MIMO by 10%. The evolution of the peak and average rates is illustrated in Figure 13.11. The average rate corresponds to fully loaded macro cells. The practical networks are not always fully loaded, and therefore the gains of 64QAM and MIMO features can be higher in real networks than in fully loaded simulations.

13.6 Dual Carrier HSDPA

The LTE radio improves the data rates compared to HSPA because LTE can use a transmission bandwidth up to 20 MHz compared to 5 MHz in HSPA. The dual carrier (dual cell) HSDPA was specified as part of Release 8 enabling HSDPA to benefit from two adjacent HSDPA carriers in the transmission to a single terminal using a total 10 MHz downlink bandwidth. The uplink solution in Release 8 is still using single 5 MHz carrier. The concept is illustrated in Figure 13.12.

The benefit of the dual carrier for the user data rate is illustrated in Figure 13.13. The dual carrier can double the user data rate at low loading since the user can access the capacity of two carriers instead of just one. The relative benefit decreases when the loading increases. There is still some capacity benefit at high load due to frequency domain scheduling and due to dynamic balancing of the load if both carriers are not 100% loaded all the time. Node B scheduling can optimize the transmission between the two carriers based on the CQI reporting – obtaining partly similar gains as in LTE with frequency domain scheduling. The frequency domain scheduling gain in Dual carrier HSDPA is smaller than with LTE since the scheduling in Dual carrier HSDPA is done in 5 MHz blocks while LTE scheduling is done with 180 kHz resolution.

Both Dual carrier HSDPA and MIMO can boost HSDPA data rates. Those two solutions are compared in Table 13.1. Both solutions can provide the same peak rate of 42 Mbps with 64QAM modulation. MIMO can improve spectral efficiency due to two antenna transmission, while the dual carrier HSDPA brings some improvement to the high loaded case with frequency domain scheduling and a larger trunking gain. The dual carrier solution looks attractive because the data rate improvement is available over the whole cell area equally while MIMO improves the

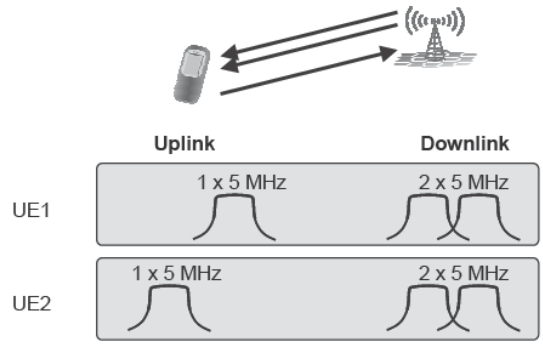


Figure 13.12 Downlink Dual carrier HSDPA concept

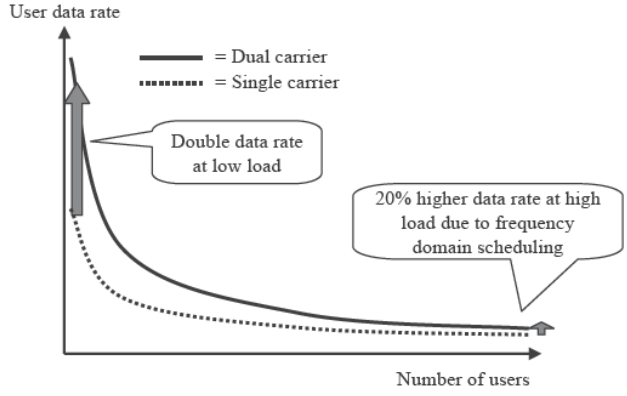


Figure 13.13 Data rate benefit of the dual carrier HSDPA

Table 13.1 Benchmarking of Dual carrier HSDPA and MIMO

	Dual carrier	MIMO
Peak bit rate	42 Mbps	42 Mbps
Spectral efficiency improvement	20% due to frequency domain scheduling and larger trunking gain	10% due to two antenna transmissions
Data rate gain	Similar gain over the whole cell area	Largest gain close to NodeB where dual stream transmission is feasible
NodeB RF requirements	Single power amplifier per sector	Two power amplifiers per sector
UE RF requirements	Possible with 1 antenna terminal	2 antennas required

data rates mostly close to the NodeB. Also, dual carrier HSDPA tends to be easier to upgrade the network since it can be implemented with single 10 MHz power amplifier per sector while MIMO requires two separate power amplifiers.

Table 13.2 HSDPA terminal categories

Category	Codes	Modulation	MIMO	DC-HSPA	Coding	Peak	3GPP
12	5	QPSK	–	–	3/4	1.8 Mbps	Release 5
6	5	16QAM	–	–	3/4	3.6 Mbps	Release 5
8	10	16QAM	–	–	3/4	7.2 Mbps	Release 5
9	15	16QAM	–	–	3/4	10.1 Mbps	Release 5
10	15	16QAM	–	–	1/1	14.0 Mbps	Release 5
13	15	64QAM	–	–	5/6	17.6 Mbps	Release 7
14	15	64QAM	–	–	1/1	21.1 Mbps	Release 7
15	15	16QAM	2 × 2	–	5/6	23.4 Mbps	Release 7
16	15	16QAM	2 × 2	–	1/1	28.0 Mbps	Release 7
17	15	64QAM or MIMO		–	5/6	23.4 Mbps	Release 7
18	15	64QAM or MIMO		–	1/1	28.0 Mbps	Release 7
19	15	64QAM	2 × 2	–	5/6	35.3 Mbps	Release 8
20	15	64QAM	2 × 2	–	1/1	42.2 Mbps	Release 8
21	15	16QAM	–	Yes	5/6	23.4 Mbps	Release 8
22	15	16QAM	–	Yes	1/1	28.0 Mbps	Release 8
23	15	64QAM	–	Yes	5/6	35.3 Mbps	Release 8
24	15	64QAM	–	Yes	1/1	42.2 Mbps	Release 8

3GPP Release 8 does not define the use of MIMO and dual carrier HSDPA at the same time. Therefore, the peak data rate in Release 8 is still 42 Mbps even with dual carrier solution. It would be possible to double the data rate to 84 Mbps by combining MIMO and dual carrier transmission in later 3GPP releases.

New HSDPA terminal categories with 64QAM, MIMO and DC-HSDPA were added in Releases 7 and 8. The HSDPA categories 13 and 14 include 64QAM and categories 15–18 MIMO offering peak rates of 21.1 Mbps and 28.0 Mbps. The combination of 2 × 2 MIMO and 64QAM is part of categories 19 and 20 pushing the peak rate to 42.2 Mbps. The DC-HSDPA categories are 21–24. Categories 21–22 support 16QAM and categories 23–24 support 64QAM. HSDPA terminal categories are listed in Table 13.2.

13.7 Uplink 16QAM

HSUPA Release 6 uplink uses QPSK modulation providing 5.76 Mbps. The uplink data rate within 5 MHz can be increased by using higher order modulation or MIMO transmission. The challenge with uplink MIMO is that the UE needs to have two power amplifiers. Therefore, uplink single user MIMO is not part of HSPA evolution nor LTE in Release 8. The higher order 16QAM modulation was adopted as part of Release 7 for HSUPA, doubling the peak rate to 11.5 Mbps. HSUPA terminal categories are listed in Table 13.3.

The higher order modulation improves the downlink spectral efficiency because the downlink has a limited number of orthogonal resources. The same is not true for uplink since the HSUPA uplink is not orthogonal and there is a practically unlimited number of codes available in uplink. The highest uplink spectral efficiency can be achieved by using QPSK modulation only. In other words, HSUPA 16QAM is a peak data rate feature, not a capacity feature.

The multi-path propagation affects high data rate performance. Therefore, the UE equalizer is used on HSDPA terminals. Also the NodeB receiver can improve the uplink high bit rate

Table 13.3 HSUPA terminal categories

Category	TTI	Modulation	MIMO	Coding	Peak	3GPP
3	10 ms	QPSK	–	3/4	1.4 Mbps	Release
5	10 ms	QPSK	–	3/4	2.0 Mbps	Release
6	2 ms	QPSK	–	1/1	5.7 Mbps	Release
7	2 ms	16QAM	–	1/1	11.5 Mbps	Release

HSUPA performance in multi-path channels by using an equalizer. Another solution is to use four antenna reception in uplink.

High uplink data rates require also high Ec/N0. Fixed reference channel 8 with 8 Mbps and 70% throughput requires Ec/N0 = 12–16 dB [6]. The corresponding uplink noise rise will also be similar to 12–16 dB impacting the coverage of other simultaneous users. It is therefore beneficial to use the uplink interference cancellation to subtract the high bit rate interference from other simultaneous users. The LTE radio uses an orthogonal uplink solution avoiding the intra-cell interference.

The uplink bit rate could be increased from 11.5 Mbps to 23 Mbps by using dual carrier transmission. The uplink dual carrier is not defined in Release 8 but could be considered for later 3GPP releases.

13.8 Layer 2 Optimization

WCDMA Release 99 specification was based on the packet retransmissions running from the Radio Network Controller (RNC) to the UE on L2. The L2 Radio Link Control (RLC) packets had to be relatively small to avoid the retransmission of very large packets in case of transmission errors. Another reason for the relatively small RLC packet size was the need to provide sufficiently small step sizes for adjusting the data rates for Release 99 channels. The RLC packet size in Release 99 is not only small, but it is also fixed for Acknowledged Mode Data and there are just a limited number of block sizes in Unacknowledged Mode Data. This limitation is due to transport channel data rate limitations in Release 99.

The RLC payload size is fixed to 40 bytes in Release 99 for Acknowledged Mode Data. The same RLC solution is applied to HSDPA Release 5 and HSUPA Release 6 as well: the 40-byte packets are transmitted from RNC to the base station for HSDPA. An additional configuration option to use an 80-byte RLC packet size was introduced in Release 5 to avoid extensive RLC protocol overhead, L2 processing and RLC transmission window stalling. With the 2 ms TTI used with HSDPA this leads to possible data rates being multiples of 160 kbps and 320 kbps respectively.

As the data rates are further increased in Release 7, increasing the RLC packet size even further would significantly impact on the granularity of the data rates available for HSDPA scheduling and the possible minimum data rates.

3GPP HSDPA and HSUPA allow the optimization of the L2 operation since L1 retransmissions are used and the probability of L2 retransmissions is very low. Also, the Release 99 transport channel limitation does not apply to HSDPA/HSUPA since the L2 block sizes are independent of the transport formats. Therefore, it is possible to use flexible and considerably larger RLC sizes and introduce segmentation to the Medium Access Control (MAC) layer in the base station.

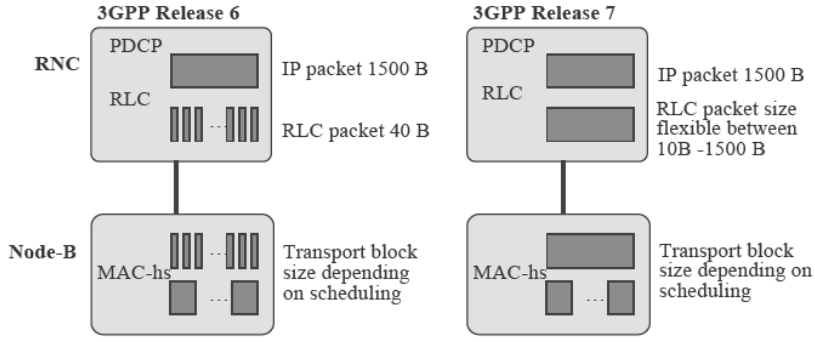


Figure 13.14 Flexible RLC concept

This optimization is included for downlink in Release 7 and for uplink in Release 8 and it is called flexible RLC and MAC segmentation solution. The RLC block size in flexible RLC solution can be as large as an Internet Protocol (IP) packet, which is typically 1500 bytes for download. There is no need for packet segmentation in RNC. By introducing the segmentation to the MAC, the MAC can perform the segmentation of the large RLC PDU based on physical layer requirements when needed. The flexible RLC concept in downlink is illustrated in Figure 13.14.

The flexible RLC and MAC segmentation brings a number of benefits for L2 efficiency and peak bit rates.

- The relative L2 overhead is reduced. With the RLC header of 2 bytes the RLC overhead is 5% for the 40-byte RLC packet. When the RLC packet size increases to 1500 bytes, the RLC header overhead is reduced to below 0.2%. This reduction of the overhead can improve the effective application data throughput.
- The RLC block size can be selected flexibly according to the packet size of each application. This flexibility helps to avoid unnecessary padding, which is no longer needed in the flexible RLC solution. This is relevant especially for small IP packet sizes, which are typical in VoIP or streaming applications.
- Less packet processing is required in the RNC and in UE with octet aligned protocol headers. The number of packets to be processed is reduced since the RLC packet size is increased and octet aligned protocol headers avoid bit shifting in high data rate connections. Both reduce L2 processing load and make the high bit rate implementation easier.
- Full flexibility and resolution of available data rates for the HSDPA scheduler.

13.9 Single Frequency Network (SFN) MBMS

Multimedia broadcast multicast service (MBMS) was added to 3GPP as part of Release 6. 3GPP Release 6 can use soft combining of the MBMS transmission from the adjacent cells. The soft combining considerably improves MBMS performance at the cell edge compared to receiving the signal from a single cell only. Release 6 MBMS therefore provides a very good starting point for broadcast services from the performance point of view.

Even if the soft combining can be used in Release 6, the other cell signals still cause interference to the MBMS reception since the adjacent cells are not orthogonal due to different scrambling codes. If the same scrambling code was used in all cells together with a terminal equalizer, the other cells transmitting the same signal in a synchronized network would be seen just as a single signal with time dispersion. This solution essentially provides a single frequency network with practically no neighboring cell interference. The MBMS over a Single Frequency Network (MBSFN) can enhance MBMS data rates and capacity. The single frequency can be realized with network synchronization and by using the same scrambling code for MBMS transmissions from multiple cells. The MBSFN is included in Release 7 and extended to unpaired bands in Release 8. The unpaired band solution is called Integrated Mobile Broadcast (IMB). The MBSFN solution requires a dedicated carrier for MBMS only transmission, which makes MBSFN a less flexible solution from the spectrum point of view compared to Release 6 MBMS. Release 6 MBMS can coexist with point-to-point traffic on the same carrier.

13.10 Architecture Evolution

3GPP networks will be increasingly used for IP based packet services. 3GPP Release 6 has four network elements in the user and control plane: base station (NodeB), RNC, Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN). The architecture in Release 8 LTE will have only two network elements: base station in the radio network and Access Gateway (a-GW) in the core network. The a-GW consists of control plane Mobility Management Entity (MME) and user plane System Architecture Evolution Gateway (SAE GW). The flat network architecture reduces the network latency and thus improves the overall performance of IP based services. The flat model also improves both user and control plane efficiency. The flat architecture is considered beneficial also for HSPA and it is specified in Release 7. The HSPA flat architecture in Release 7 and LTE flat architecture in Release 8 are exactly the same: NodeB is responsible for the mobility management, ciphering, all retransmissions and header compression both in HSPA and in LTE. The architecture evolution in HSPA is designed to be backwards compatible: existing terminals can operate with the new architecture and the radio and core network functional split is not changed. The architecture evolution is illustrated in Figure 13.15.

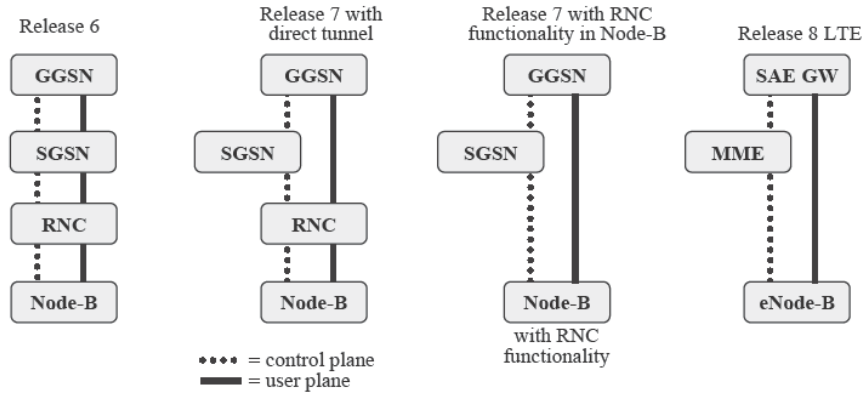


Figure 13.15 Evolution towards flat architecture

Also the packet core network has flat architecture in Release 7. It is called direct tunnel solution and allows the user plane to bypass SGSN. When having the flat architecture with all RNC functionality in the base station and using a direct tunnel solution, only two nodes are needed for user data operation. This achieves flexible scalability and allows the introduction of the higher data rates with HSPA evolution with minimum impact to the other nodes in the network. This is important for achieving a low cost per bit and enabling competitive flat rate data charging. As the gateway in LTE has similar functionality as the GGSN, it is foreseen to enable LTE and HSPA deployment where both connect directly to the same core network element for user plane data handling directly from the base station.

3GPP Release 8 includes a new architecture to support small home NodeBs as well. The home NodeBs are also called femto access points. The home NodeBs are installed by people at home in the same way as WLAN access points are installed now. These home NodeBs use the operator's licensed frequency and are connected to the operator's core network. The output power level is low, typically 100 mW or below. The transport connection uses fixed DSL connections in homes. The handovers and idle mode selections between home NodeBs and the macro network are controlled by the network algorithms and parameters. The flat architecture from Figure 13.15 is not optimized for when there are a very large number of small base stations because the number of core network connections would be very large. Therefore, a new network element – called Home NodeB gateway – is introduced to hide the large number of home NodeBs from the core network. The gateway is located in the operator's premises. The new interface between home NodeBs and the gateway is called Iuh. The interface between the gateway and the core network is the normal Iu interface. The home NodeB architecture is illustrated in Figure 13.16 [7].

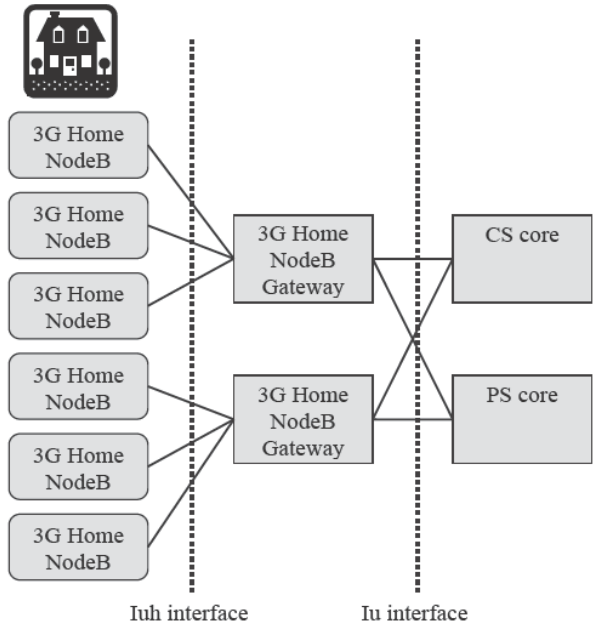


Figure 13.16 Home NodeB architecture [7]

13.11 Summary

While 3GPP is working on a new radio system – LTE (Long Term Evolution) – in Release 8, there are a number of important enhancements also brought to existing High Speed Packet Access (HSPA) standards. HSPA enhancements provide major improvements to the end user performance in increasing peak bit rates, reducing mobile terminal power consumption and reducing channel allocation latency. The peak bit rates can be tripled with 64QAM and 2 × 2 MIMO from 14 Mbps to 42 Mbps. These peak rate features also bring some improvements to the spectral efficiency and cell capacity. The downlink dual carrier HSDPA solution also enhances the data rates by using two adjacent HSDPA carriers for a single user transmission. The combination of Dual carrier and MIMO could push the peak rate to 84 Mbps in further 3GPP releases.

The UE power consumption is considerably reduced in HSPA evolution due to discontinuous transmission and reception. The expected voice call talk time will improve by 50% and the usage time of bursty data applications increases even more. The practical talk times will improve further when the digital and RF technologies evolve and general power consumption is reduced.

HSPA evolution enhances basic voice service by mapping Circuit Switched (CS) voice on top of HSPA channels – a concept called CS voice over HSPA. It is essentially a combination of Voice over IP (VoIP) in the radio and CS voice in the core network. CS voice over HSPA

		Peak rate	Average rate (capacity)	Cell edge rate	Latency gain	Talk time
Downlink	HSDPA 64QAM[1]	+50%	<10%	-	-	-
	HSDPA 2x2MIMO	+100%	<30%	<20%	-	-
	DC-HSDPA	+100%	+20-100%	+20-100%	-	-
Uplink	HSUPA 10 ms (2.0 Mbps)[2]	+600%	+20-100%	<100%	Gain 20 ms	-
	HSUPA 2 ms (5.8 Mbps)	+200%	<30%	-	Gain 15 ms	-
	HSUPA 16QAM	+100%	-	-	-	-
	Advanced NodeB receiver	-	>30%	-	-	-
	DTX/DRX	-	-	-	-	>+50%
	HS-FACH / HS-RACH	-	-	-	Setup time <0.1 s	-
	CS voice over HSPA	-	+80% (voice)	-	-	>+50%

[1]Baseline WCDMA Release 5 downlink 14.4 Mbps

[2]Baseline WCDMA Release 99 uplink 384 kbps

= clear gain >30%

= moderate gain <30%

Figure 13.17 Summary of HSPA evolution features and their benefits

reduces UE power consumption and increases spectral efficiency compared to WCDMA voice while having no impact on the core network and end-to-end ecosystem. It will be easy to upgrade CS voice over HSPA later to VoIP over HSPA since the radio solutions are very similar.

The setup times are reduced by also mapping the RACH/FACH common channels on top of HSDPA/HSUPA. Actually, all the services including signaling can be mapped on top of HSPA in HSPA evolution. There are only a few physical layer channels left from the Release 99 specifications in Release 8 – otherwise, everything is running on top of HSPA. This explains also why the end user performance and the network efficiency are considerably improved compared to Release 99.

The performance benefits of the HSPA evolution features are summarized in Figure 13.17. The exact gains depend on the implementation choices and on the deployment environment.

3GPP Release 7 allows simplification of the network architecture. The number of network elements for the user plane can be reduced from four in Release 6 to two in Release 7. Release 7 architecture is exactly the same as used in LTE in Release 8, which makes the network evolution from HSPA to LTE straightforward. Most of the 3GPP Release 7 and 8 enhancements are expected to be relatively simple upgrades to the HSPA networks as was the case for HSDPA and HSUPA in earlier releases.

References

[1] 3GPP, 'Continuous connectivity for packet data users', 3GPP TR25.903 V7.0.0, March 2007.
[2] 3GPP, Technical Specification 24.008 'Mobile radio interface Layer 3 specification; Core network protocols', V.8.3.0.
[3] Kurjenniemi, J., Nihtilä, T., Lampinen, M., Ristaniemi, T. 'Performance of WCDMA HSDPA Network with Different Advanced Receiver Penetrations', Wireless Personal Multimedia Communications (WPMC), Aalborg, Denmark, 17–22 September 2005.
[4] 3GPP, 'Further discussion on delay enhancements in Rel7', 3GPP R2-061189, August 2006.
[5] 3GPP, '64QAM for HSDPA', 3GPP R1-063335, November 2006.
[6] 3GPP, Technical Specifications 25.104 'Base Station (BS) radio transmission and reception (FDD)', V.8.3.0.
[7] 3GPP, Technical Report 25.820 '3G Home NodeB Study Item Technical Report', V.8.1.1.

Index

LTE for UMTS: OFDMA and SC-FDMA Based Radio Access Edited by Harri Holma and Antti Toskala

ISBN: 978-0-470-99401-6

Attachment B-2

LIBRARY | CATALOG

Search Navigation ▾ < 1 of 3 >

BOOK

LTE for UMTS : Evolution to LTE-Advanced

Full Record MARC Tags

Main title

LTE for UMTS : Evolution to LTE-Advanced / edited by Harri Holma and Antti Toskala.

Edition

Second Edition.

Published/Created

Chichester, West Sussex, United Kingdom : Wiley, 2011, ©2011.

Request this Item | LC Find It

Links

Links	Cover image http://catalogimages.wiley.com/images/db/jimages/9780470660003.jpg

More Information

LCCN Permalink	https://lccn.loc.gov/2010050375
Description	xxxi, 543 pages : illustrations ; 25 cm
ISBN	9780470660003 (hardback) 9781119992943 (oBook)

LC classification	TK5103.4883 .L78 2011
Variant title	Long Term Evolution for Universal Mobile Telecommunications Systems
Related names	Holma, Harri, 1970- Toskala, Antti.
Summary	"Still providing the thorough breakdown of the key areas of LTE of the first edition, this new revised edition of LTE for UMTS includes some important new sections on updates to the standards, as well as a look at possible future developments"--Provided by publisher.
LC Subjects	Universal Mobile Telecommunications System. Wireless communication systems--Standards. Mobile communication systems--Standards. Global system for mobile communications. Long-Term Evolution (Telecommunications)
Browse by shelf order	TK5103.4883
Notes	"Written by experts actively involved in the 3GPP standards and product development, LTE for UMTS, Second Edition gives a complete and up-to-date overview of Long term Evolution (LTE) in a systematic and clear manner. Building upon on the success of the first edition, LTE for UMTS, Second Edition has been revised to now contain improved coverage of the Release 8 LTE details, including field performance results, transport network, self optimized networks and also covering the enhancements done in 3GPP Release 9. This new edition also provides an outlook to Release 10, including the overview of Release 10 LTE-Advanced technology components which enable reaching data rates beyond 1 Gbps. Key updates for the second edition of LTE for UMTS are focused on the new topics from Release 9 & 10, and include: LTE-Advanced; Self optimized networks (SON); Transport network dimensioning; Measurement results"--Provided by publisher. Includes bibliographical references and index.
LCCN	2010050375
Dewey class no.	621.3845/6
Other class no.	TEC061000
Other system no.	(OCoLC)ocn694616471
Invalid ISBN	9781119992950 (ePDF) 9781119992936 (ePub)

Type of material	Book
Content type	text
Media type	unmediated
Carrier type	volume

Item Availability

CALL NUMBER	TK5103.4883 .L78 2011 Copy 1
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

Attachment C-2

LIBRARY | CATALOG

Search Navigation ▾ < 1 of 3 >

BOOK

LTE for UMTS : Evolution to LTE-Advanced

Full Record | MARC Tags

000		03054cam a2200457 i 4500
001		16556413
005		20151026090113.0
008		101129t20112011xxka b 001 0 eng
906	__	|a 7 |b cbc |c orignew |d 1 |e ecip |f 20 |g y-gencatlg
925	0_	|a acquire |b 2 shelf copies |x policy default
955	__	|b rg11 2010-11-29 (telework) |c rg11 2010-11-29 ONIX (telework) to Gen Sci/Tech (STM) |d xh12 2010-12-29 |w rd11 2010-12-29 |a xe07 2011-06-02 1 copy rec'd., to CIP ver. |f rf08 2011-06-21 to BCCD
010	__	|a 2010050375
020	__	|a 9780470660003 (hardback)
020	__	|z 9781119992950 (ePDF)
020	__	|a 9781119992943 (oBook)
020	__	|z 9781119992936 (ePub)
035	__	|a (OCoLC)ocn694616471
040	__	|a DLC |b eng |c DLC |e rda |d YDX |d YDXCP |d DLC
042	__	|a pcc
050	00	|a TK5103.4883 |b .L78 2011
082	00	|a 621.3845/6 |2 22
084	__	|a TEC061000 |2 bisacsh
245	00	|a LTE for UMTS : |b Evolution to LTE-Advanced / |c edited by Harri Holma and Antti Toskala.
246	3_	|a Long Term Evolution for Universal Mobile Telecommunications Systems
250	__	|a Second Edition.
260	__	|a Chichester, West Sussex, United Kingdom : |b Wiley, |c 2011, ©2011.

300	__	|a xxxi, 543 pages : |b illustrations ; |c 25 cm
336	__	|a text |2 rdacontent
337	__	|a unmediated |2 rdamedia
338	__	|a volume |2 rdacarrier
500	__	|a "Written by experts actively involved in the 3GPP standards and product development, LTE for UMTS, Second Edition gives a complete and up-to-date overview of Long term Evolution (LTE) in a systematic and clear manner. Building upon on the success of the first edition, LTE for UMTS, Second Edition has been revised to now contain improved coverage of the Release 8 LTE details, including field performance results, transport network, self optimized networks and also covering the enhancements done in 3GPP Release 9. This new edition also provides an outlook to Release 10, including the overview of Release 10 LTE-Advanced technology components which enable reaching data rates beyond 1 Gbps. Key updates for the second edition of LTE for UMTS are focused on the new topics from Release 9 & 10, and include: LTE-Advanced; Self optimized networks (SON); Transport network dimensioning; Measurement results"--Provided by publisher.
520	__	|a "Still providing the thorough breakdown of the key areas of LTE of the first edition, this new revised edition of LTE for UMTS includes some important new sections on updates to the standards, as well as a look at possible future developments"--Provided by publisher.
504	__	|a Includes bibliographical references and index.
650	_0	|a Universal Mobile Telecommunications System.
650	_0	|a Wireless communication systems |x Standards.
650	_0	|a Mobile communication systems |x Standards.
650	_0	|a Global system for mobile communications.
650	_0	|a Long-Term Evolution (Telecommunications)
700	1_	|a Holma, Harri, |d 1970-
700	1_	|a Toskala, Antti.
856	42	|3 Cover image |u http://catalogimages.wiley.com/images/db/jimages/9780470660003.jpg

Request this Item | LC Find It

Item Availability

CALL NUMBER	TK5103.4883 .L78 2011 Copy 1
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged

Attachment D-2

WorldCat

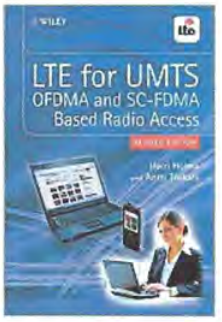
Search WorldCat

Search

Advanced Search Find a Library

Cite/Export Print E-mail Share Permalink

Add to list Add tags Write a review Rate this item: 1 2 3 4 5



Preview this item

LTE for UMTS : Evolution to LTE-Advanced

Author: Harri Holma; Antti Toskala

Publisher: Chichester, West Sussex, United Kingdom : Wiley, 2011, ©2011.

Edition/Format: Print book : English : Second edition View all editions and formats

Summary: "Still providing the thorough breakdown of the key areas of LTE of the first edition, this new revised edition of LTE for UMTS includes some important new sections on updates to the standards, as well as a look at possible future developments"--Provided by publisher.

Rating: (not yet rated) 0 with reviews - Be the first.

Subjects Universal Mobile Telecommunications System.
Wireless communication systems -- Standards.
Mobile communication systems -- Standards.
View all subjects

More like this User lists Similar Items

Get a Copy

Find a copy in the library

AbeBooks	$26.70
Amazon	$89.99
iTunes	$109.99

Find a copy in the library

Enter your location: District of Columbia Find libraries
Submit a complete postal address for best results.

Displaying libraries 1-6 out of 690 for all 28 editions (Washington, DC, USA) Show libraries holding **just this edition**

« First ‹ Prev 1 2 3 Next › Last »

	Library	Held formats	Distance	
1.	**Consumer Financial Protection Bureau** Washington, DC 20552 United States	Book	1 mile MAP IT	Library info Add to favorites
2.	**Howard University** Washington, DC 20059 United States	Book	1 mile MAP IT	Library info Ask a librarian Add to favorites
3.	**Research Center, National Academies of Sciences, Engineering, and Medicine** Washington, DC 20001 United States	Book	1 mile MAP IT	Library info Add to favorites
4.	**US Department of Justice Library** Washington, DC 20530 United States	Book	1 mile MAP IT	Library info Add to favorites
5.	**Federal Communications Commission** Washington, DC 20554 United States	Book	2 miles MAP IT	Library info Add to favorites
6.	**Library of Congress**	Book	2 miles	Library info

Washington, DC 20540 United States MAP IT Ask a librarian Add to favorites

« First ‹ Prev 1 2 3 Next › Last »

Details

Document Type: Book

All Authors / Contributors: Harri Holma; Antti Toskala

Find more information about: Harri Holma Go

ISBN: 9780470660003 0470660007 9781119992943 111999294X 1119992958 9781119992950 1119992931 9781119992936

OCLC Number: 694616471

Notes: "Written by experts actively involved in the 3GPP standards and product development, LTE for UMTS, Second Edition gives a complete and up-to-date overview of Long term Evolution (LTE) in a systematic and clear manner. Building upon on the success of the first edition, LTE for UMTS, Second Edition has been revised to now contain improved coverage of the Release 8 LTE details, including field performance results, transport network, self optimized networks and also covering the enhancements done in 3GPP Release 9. This new edition also provides an outlook to Release 10, including the overview of Release 10 LTE-Advanced technology components which enable reaching data rates beyond 1 Gbps. Key updates for the second edition of LTE for UMTS are focused on the new topics from Release 9 & 10, and include: LTE-Advanced; Self optimized networks (SON); Transport network dimensioning; Measurement results"--Provided by publisher.

Description: xxxi, 543 pages : illustrations ; 25 cm

Contents:
1. Introduction --
2. LTE standardization --
3. System architecture based on 3GPP SAE --
4. Introduction to OFDMA and SC-FDMA and to MIMO in LTE --
5. Physical layer --
6. LTE radio protocols --
7. Mobility --
8. Radio resource management --
9. Self organizing networks (SON) --
10. Performance --
11. LTE measurements --
12. Transport --
13. Voice over IP (VoIP) --
14. Performance requirements --
15 LTE TTD mode --
16. LTE-advanced --
17. HSPA evolution --
Index.

Other Titles: Long Term Evolution for Universal Mobile Telecommunications Systems

Responsibility: edited by Harri Holma and Antti Toskala.

More information: Cover image

Abstract:

Written by experts involved in the 3GPP standards and product development, this title gives a complete overview of Long term Evolution (LTE) in a systematic and clear manner. It provides an outlook Read more...

Reviews

Editorial reviews

Publisher Synopsis
"Written by experts actively involved in the 3GPP standards and product development, LTE for UMTS, Second Edition gives a complete and up-to-date overview of Long Term Evolution (LTE) in a systematic Read more...

User-contributed reviews

Add a review and share your thoughts with other readers. Be the first.

Tags

Add tags for "LTE for UMTS : Evolution to LTE-Advanced". Be the first.

Similar Items

Related Subjects: (9)

- Universal Mobile Telecommunications System.
- Wireless communication systems -- Standards.
- Mobile communication systems -- Standards.
- Global system for mobile communications.
- Long-Term Evolution (Telecommunications)
- TECHNOLOGY & ENGINEERING -- Mobile & Wireless Communications.
- UMTS.
- Trådlös kommunikation.
- Wireless communication systems.

User lists with this item (1)

List 1 ENG (500 items)
by lschitwood updated 2017-11-25

Linked Data

Tags

Add tags for "LTE for UMTS : evolution to LTE-advanced". Be the first.

Similar Items

Related Subjects: (12)

- Universal Mobile Telecommunications System.
- Wireless communication systems -- Standards.
- Mobile communication systems -- Standards.
- Global system for mobile communications.
- Long Term Evolution.
- FDMA.
- OFDMA.
- UMTS.
- Global system for mobile communications
- Mobile communication systems -- Standards
- Universal Mobile Telecommunications System
- Wireless communication systems -- Standards

Linked Data

Attachment A-3

LTE – The UMTS Long Term Evolution

LTE – The UMTS Long Term Evolution

From Theory to Practice

Second Edition

Stefania Sesia
ST-Ericsson, France

Issam Toufik
ETSI, France

Matthew Baker
Alcatel-Lucent, UK

WILEY
A John Wiley & Sons, Ltd., Publication

This edition first published 2011

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ,
United Kingdom

Photograph on cover courtesy of Alcatel-Lucent, from the *ngConnect* LTE-equipped car.
3GPP website reproduced by permission of © 3GPP™.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging-in-Publication Data

Sesia, Stefania.
LTE–the UMTS long term evolution : from theory to practice / Stefania Sesia, Issam Toufik, Matthew Baker.– 2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-66025-6 (hardback)
1. Universal Mobile Telecommunications System. 2. Long-Term Evolution (Telecommunications) I. Toufik, Issam. II. Baker, Matthew (Matthew P.J.) III. Title.
TK5103.4883.S47 2011
621.3845′6–dc22
2010039466

A catalogue record for this book is available from the British Library.

Print ISBN: 9780470660256 (H/B)
ePDF ISBN: 9780470978511
oBook ISBN: 9780470978504
epub ISBN: 9780470978641

Printed in Great Britain by CPI Antony Rowe, Chippenham, Wiltshire.

Dedication

To my family.
Stefania Sesia

To my parents for their sacrifices and unconditional love. To my brother and sisters for their love and continual support. To my friends for being what they are.
Issam Toufik

To the glory of God, who 'so loved the world that He gave His only Son, that whoever believes in Him shall not perish but have eternal life'. — The Bible.
Matthew Baker

Contents

Editors' Biographies

Matthew Baker holds degrees in Engineering and Electrical and Information Sciences from the University of Cambridge. From 1996 to 2009 he worked at Philips Research where he conducted leading-edge research into a variety of wireless communication systems and techniques, including propagation modelling, DECT, Hiperlan and UMTS, as well as leading the Philips RAN standardization team. He has been actively participating in the standardization of both UMTS WCDMA and LTE in 3GPP since 1999, where he has been active in 3GPP TSG RAN Working Groups 1, 2, 4 and 5, contributing several hundred proposals. He now works for Alcatel-Lucent, which he joined in 2009, and he has been Chairman of 3GPP TSG RAN Working Group 1 since being elected to the post in August of that year. He is the author of several international conference papers and inventor of numerous patents. He is a Chartered Engineer, a Member of the Institution of Engineering and Technology and a Visiting Lecturer at the University of Reading, UK.

Stefania Sesia received her Ph.D. degree in Communication Systems and Coding Theory from both Eurecom (Sophia Antipolis, France) and ENST-Paris (Paris, France) in 2005. From 2002 to 2005 she worked at Motorola Research Labs, Paris, towards her Ph.D. thesis. In June 2005 she joined Philips/NXP Semiconductors (now ST-Ericsson) Research and Development Centre in Sophia Antipolis, France where she was technical leader and responsible for the High Speed Downlink Packet Access algorithm development. She has been participating in 3GPP TSG RAN Working Groups 1 and 4 standardization meetings. From 2007 to 2009 she was on secondment from NXP Semiconductors to the European Telecommunications Standard Institute (ETSI) acting as 3GPP TSG RAN and 3GPP TSG RAN Working Group 4 Technical Officer. She is currently back in ST-Ericsson as senior research and development engineer, actively participating in 3GPP TSG RAN Working Group 4 as a delegate. She is the author of several international IEEE conference and journal papers and many contributions to 3GPP, and inventor of numerous US and European patents.

Issam Toufik graduated in Telecommunications Engineering (majoring in Mobile Communication Systems) in 2002 from both ENST-Bretagne (Brest, France) and Eurecom (Sophia Antipolis, France). In 2006, he received his Ph.D. degree in Communication Systems from Eurecom/ENST-Paris, France. From June to August 2005 he worked for Samsung Advanced Institute of Technology (SAIT), South Korea, as a Research Engineer on LTE. In January 2007, he joined NXP Semiconductors/ST-Ericsson, Sophia Antipolis, France, as a Research and Development Engineer for UMTS and LTE algorithm development. In November 2009, he joined the European Telecommunications Standard Institute (ETSI) acting as 3GPP TSG

RAN and 3GPP TSG RAN Working Group 4 Technical Officer. He is the author of several international IEEE conference and journal papers and contributions to 3GPP, and inventor of numerous patents.

List of Contributors

Abe, Tetsushi, NTT DOCOMO
e-mail: abetet@nttdocomo.com

Ancora, Andrea, ST-Ericsson
e-mail: andrea.ancora@stericsson.com

Anderson, Nicholas, Research In Motion
e-mail: nianderson@rim.com

Baker, Matthew, Alcatel-Lucent
e-mail: matthew.baker@alcatel-lucent.com, m.p.j.baker.92@cantab.net

Bertrand, Pierre, Texas Instruments
e-mail: p-bertrand@ti.com

Bhattad, Kapil, Qualcomm
e-mail: kbhattad@qualcomm.com

Bury, Andreas, Blue Wonder Communications
e-mail: andreas.bury@bluwo.com

Chun, SungDuck,LG Electronics
e-mail: duckychun@lge.com

Classon, Brian, Huawei
e-mail: brian.classon@huawei.com

Damnjanovic, Jelena, Qualcomm
e-mail: jelenad@qualcomm.com

Farajidana, Amir, Qualcomm
e-mail: amirf@qualcomm.com

Fischer, Patrick, Bouygues Telecom
e-mail: pfischer@bouyguestelecom.fr

Geirhofer, Stefan, Qualcomm
e-mail: sgeirhofer@qualcomm.com

Gerstenberger, Dirk, Ericsson
e-mail: dirk.gerstenberger@ericsson.com

Gesbert, David, Eurecom
e-mail: david.gesbert@eurecom.fr

Godin, Philippe, Alcatel-Lucent
e-mail: philippe.godin@alcatel-lucent.com

Golitschek, Alexander, Panasonic
e-mail: alexander.golitschek@eu.panasonic.com

Gonsa, Osvaldo, Panasonic
e-mail: osvaldo.gonsa@eu.panasonic.com

Gorokhov, Alex, Qualcomm
e-mail: gorokhov@qualcomm.com

Hardouin, Eric, Orange Labs
e-mail: eric.hardouin@orange-ftgroup.com

Hu, Teck, Alcatel-Lucent
e-mail: teck.hu@alcatel-lucent.com

Hus, Olivier,
e-mail: olivierjhus@gmail.com

Jämsä, Tommi, Elektrobit
e-mail: tommi.jamsa@elektrobit.com

Jiang, Jing, Texas Instruments
e-mail: jing.jiang@ti.com

Kaltenberger, Florian, Eurecom
e-mail: florian.kaltenberger@eurecom.fr

Kazmi, Muhammad, Ericsson
e-mail: muhammad.kazmi@ericsson.com

Knopp, Raymond, Eurecom
e-mail: raymond.knopp@eurecom.fr

Laneman, J. Nicholas, University of Notre Dame
e-mail: jnl@nd.edu

Lee, YoungDae, LG Electronics
e-mail: leego@lge.com

Love, Robert, Motorola Mobility
e-mail: robert.love@motorola.com

Luo, Xiliang, Qualcomm
e-mail: xluo@qualcomm.com

Montojo, Juan, Qualcomm
e-mail: juanm@qualcomm.com

Moulsley, Tim, Fujitsu
e-mail: t.moulsley@btopenworld.com

Nakamura, Takaharu, Fujitsu
e-mail: n.takaharu@jp.fujitsu.com

Nakamura, Takehiro, NTT DOCOMO
e-mail: nakamurata@nttdocomo.co.jp

Nangia, Vijay, Motorola Mobility
e-mail: vijay.nangia@motorola.com

Nimbalker, Ajit, Motorola Mobility
e-mail: aijt.nimbalker@motorola.com

Palat, K. Sudeep, Alcatel-Lucent
e-mail: spalat@alcatel-lucent.com

Payne, Adrian, ERA Technology
e-mail: adrian.w.payne@gmail.com

Ranta-aho, Karri, Nokia Siemens Networks
e-mail: karri.ranta-aho@nsn.com

Rumney, Moray, Agilent
e-mail: moray_rumney@agilent.com

Sälzer, Thomas, Huawei
e-mail: thomas.salzer@huawei.com, thomas.salzer@gmx.de

Sayers, Tony, Ultra Electronics
e-mail: tony.sayers@talktalk.net

Sesia, Stefania, ST-Ericsson
e-mail: stefania.sesia@stericsson.com

Shen, Zukang, CATT
e-mail: shenzukang@catt.cn

Slock, Dirk, Eurecom
e-mail: dirk.slock@eurecom.fr

Suzuki, Hidetoshi, Panasonic
e-mail: Suzuki.Hidetoshi@jp.panasonic.com

Tomatis, Fabrizio, ST-Ericsson
e-mail: fabrizio.tomatis@stericsson.com

Tosato, Filippo, Toshiba
e-mail: filippo.tosato@toshiba-trel.com

Toufik, Issam, ETSI
e-mail: issam.toufik@etsi.org, issam.toufik@eurecom.fr

van der Velde, Himke, Samsung
e-mail: himke.vandervelde@samsung.com

van Rensburg, Cornelius, Huawei
e-mail: cdvanren@ieee.org

Whinnett, Nick, Picochip
e-mail: nickw@picochip.com

Yi, SeungJune, LG Electronics
e-mail: seungjune@lge.com

Ylitalo, Juha, Elektrobit
e-mail: juha.ylitalo@elektrobit.com

Foreword

GSM, and its evolution through GPRS, EDGE, WCDMA and HSPA, is the technology stream of choice for the vast majority of the world's mobile operators. Users have experienced increasing data rates, together with a dramatic reduction in telecommunications charges; they now expect to pay less but receive more. Therefore, in deciding the next steps, there must be a dual approach: seeking considerable performance improvement but at reduced cost. Improved performance must be delivered through systems which are cheaper to install and maintain. LTE and LTE-Advanced represent these next steps and will be the basis on which future mobile telecommunications systems will be built.

Many articles have already been published on the subject of LTE, varying from doctoral theses to network operator analyses and manufacturers' product literature. By their very nature, those publications have viewed the subject from one particular perspective, be it academic, operational or promotional. A very different approach is taken with this book. The authors come from a number of different spheres within the mobile telecommunications ecosystem and collectively bring a refreshing variety of perspectives. What binds the authors together is a thorough knowledge of the subject material which they have derived from their long experience within the standards-setting environment, the 3rd Generation Partnership Project (3GPP). LTE discussions started within 3GPP in 2004, so it is not a particularly new subject. In order to fully appreciate the thinking that conceived this technology, however, it is necessary to have followed the subject from the very beginning and to have witnessed the discussions that took place from the outset. Moreover, it is important to understand the thread that links academia, through research to standardization since it is widely acknowledged that by this route impossible dreams become market realities. Considerable research work has taken place to prove the viability of the technical basis on which LTE is founded and it is essential to draw on that research if any attempt is made to explain LTE to a wider audience. The authors of this book have not only followed the LTE story from the beginning but many have also been active players in WCDMA and its predecessors, in which LTE has its roots.

This book provides a thorough, authoritative and complete tutorial of the LTE system, now fully updated and extended to include LTE-Advanced. It gives a detailed explanation of the advances made in our theoretical understanding and the practical techniques that will ensure the success of this ground-breaking new radio access technology. Where this book is exceptional is that the reader will learn not just how LTE works but why it works.

I am confident that this book will earn its rightful place on the desk of anyone who needs a thorough understanding of the LTE and LTE-Advanced technology, the basis of the world's mobile telecommunications systems for the next decade.

Adrian Scrase, ETSI Vice-President,
International Partnership Projects

Preface to the Second Edition

> Research workers and engineers toil unceasingly on the development of wireless telegraphy. Where this development can lead, we know not. However, with the results already achieved, telegraphy over wires has been extended by this invention in the most fortunate way. Independent of fixed conductor routes and independent of space, we can produce connections between far-distant places, over far-reaching waters and deserts. This is the magnificent practical invention which has flowered upon one of the most brilliant scientific discoveries of our time!

These words accompanied the presentation of the Nobel Prize for Physics to Guglielmo Marconi in December 1909.

Marconi's success was the practical and commercial realization of wireless telegraphy – the art of sending messages without wires – thus exploiting for the first time the amazing capability for wireless communication built into our universe. While others worked on wireless telephony – the transmission of audio signals for voice communication – Marconi interestingly saw no need for this. He believed that the transmission of short text messages was entirely sufficient for keeping in touch.

One could be forgiven for thinking that the explosion of wireless voice communication in the intervening years has proved Marconi wrong; but the resurgence of wireless data transmission at the close of the twentieth century, beginning with the mobile text messaging phenomenon, or 'SMS', reveals in part the depth of insight Marconi possessed.

Nearly 100 years after Marconi received his Nobel prize, the involvement of thousands of engineers around the world in major standardization initiatives such as the 3rd Generation Partnership Project (3GPP) is evidence that the same unceasing toil of research workers and engineers continues apace.

While the first mobile communications standards focused primarily on voice communication, the emphasis now has returned to the provision of systems optimized for data. This trend began with the 3rd Generation Wideband Code Division Multiple Access (WCDMA) system designed in the 3GPP, and is now reaching fulfilment in its successor, the Long-Term Evolution (LTE). LTE was the first cellular communication system optimized from the outset to support packet-switched data services, within which packetized voice communications are just one part. Thus LTE can truly be said to be the heir to Marconi's heritage – the system, unknown indeed to the luminaries of his day, to which his developments have led.

LTE is an enabler. It is not technology for technology's sake, but technology with a purpose, connecting people and information to enable greater things to be achieved. It is already providing higher data rates than ever previously achieved in mobile communications,

combined with wide-area coverage and seamless support for mobility without regard for the type of data being transmitted. To provide this level of functionality and flexibility, it is inevitable that the complexities of the LTE system have far surpassed anything Marconi could have imagined.

One aim of this book, therefore, is to chart an explanatory course through the LTE specifications, to support those who design LTE equipment.

The LTE specification documents themselves do not tell the whole story. Essentially they are a record of decisions taken – decisions which are often compromises between performance and cost, theoretical possibility and practical constraints. We aim therefore to give the reader a detailed insight into the evaluations and trade-offs which lie behind the technology choices inherent in LTE. The specifications also continue to develop, as new releases are produced, and this Second Edition is therefore fully updated to cover Release 9 and the first release of LTE-Advanced, Release 10.

Since the first version of LTE was developed, the theoretical understanding which gave rise to LTE has continued to advance, as the 'unceasing toil' of thousands of engineers continues with the aim of keeping pace with the explosive growth of mobile data traffic. Where the first version of LTE exploited Multiple-Input Multiple-Output (MIMO) antenna techniques to deliver high data rates, the evolution of LTE towards LTE-Advanced extends such techniques further for both downlink and uplink communication, together with support for yet wider bandwidths; meanwhile, heterogeneous (or hierarchical) networks, relaying and Coordinated MultiPoint (CoMP) transmission and reception start to become relevant in LTE-Advanced.

It is particularly these advances in underlying scientific understanding which this book seeks to highlight.

In selecting the technologies to include in LTE and LTE-Advanced, an important consideration is the trade-off between practical benefit and cost of implementation. Fundamental to this assessment is ongoing enhancement in understanding of the radio propagation environment and scenarios of relevance to deployments of LTE and LTE-Advanced. This has been built on significant advances in radio-channel modelling.

The advances in techniques and theoretical understanding continue to be supported by developments in integrated circuit technology and signal processing power which render them feasible where they would have been unthinkable only a few years ago.

Changes in spectrum availability and regulation also influence the development path of LTE towards LTE-Advanced, reinforcing the need for the new technology to be adaptable, capable of being scaled and enhanced to meet new global requirements and deployed in a wide range of different configurations.

With this breadth and depth in mind, the authorship of the chapters of the second edition of this book is even wider than that of the first edition, and again is drawn from all fields of the ecosystem of research and development that has underpinned the design of LTE. They work in the 3GPP standardization itself, in the R&D departments of companies active in LTE, for network operators as well as equipment manufacturers, in universities and in other collaborative research projects. They are uniquely placed to share their insights from the full range of perspectives.

To borrow Marconi's words, where LTE and LTE-Advanced will lead, we know not; but we can be sure that these will not be the last developments in wireless telegraphy.

Matthew Baker, *Stefania Sesia* and *Issam Toufik*

Acknowledgements

Like the first edition, the fully updated and expanded second edition of this book is first and foremost the fruit of a significant team effort, which would not have been successful without the expertise and professionalism displayed by all the contributors, as well as the support of their companies. The dedication of all the co-authors to their task, their patience and flexibility in allowing us to modify and move certain parts of their material for harmonization purposes, are hereby gratefully acknowledged. Particular thanks are due to ST-Ericsson, Alcatel-Lucent and ETSI for giving us the encouragement and working environment to facilitate such a time-consuming project. The help provided by ETSI, 3GPP and others in authorizing us to reproduce certain copyrighted material is also gratefully acknowledged. We would like to express our gratitude to the many experts who kindly provided advice, feedback, reviews and other valuable assistance. We believe their input in all its forms has made this book a more accurate, valuable and even enjoyable resource. These experts include Jacques Achard, Kevin Baum, Martin Beale, Keith Blankenship, Yufei Blankenship, Federico Boccardi, Kevin Boyle, Sarah Boumendil, Alec Brusilovsky, Paul Bucknell, Richard Burbidge, Aaron Byman, Emilio Calvanese Strinati, Choo Chiap Chiau, Anand Dabak, Peter Darwood, Merouane Debbah, Vip Desai, Marko Falck, Antonella Faniuolo, Jeremy Gosteau, Lajos Hanzo, Lassi Hentilä, Shin Horng Wong, Paul Howard, Howard Huang, Alan Jones, Yoshihisa Kishiyama, Achilles Kogiantis, Pekka Kyösti, Daniel Larsson, Jung-Ah Lee, Thierry Lestable, Gert-Jan van Lieshout, Andrew Lillie, Matti Limingoja, Huiheng Mai, Caroline Mathieson, Darren McNamara, Juha Meinilä, Tarik Muharemovic, Gunnar Nitsche, Jukka-Pekka Nuutinen, SungJun Park, Roope Parviainen, Paul Piggin, Claudio Rey, Safouane Sfar, Ken Stewart, Miloš Tesanovic, Paolo Toccacelli, Ludo Tolhuizen, Li Wang, Tim Wilkinson and Steve Zhang.

We would also like to acknowledge the efforts of all participants in 3GPP who, through innumerable contributions and intense discussions often late into the night, facilitated the completion of the LTE specifications for Releases 8, 9 and 10 in such a short space of time.

We would especially like to thank the publishing team at John Wiley & Sons, especially Tiina Ruonamaa, Susan Barclay, Jasmine Chang, Mariam Cheok, Sheena Deuchars, Caitlin Flint, Sarah Hinton, Anna Smart and Sarah Tilley for their professionalism and extensive support and encouragement throughout the preparation of both the first and second editions of this book.

Finally, it should be noted that this book is intended only as a guide to LTE and LTE-Advanced, and the reader should refer to the specifications published by 3GPP for definitive information. Any views expressed in this book are those of the authors and do not necessarily reflect the views of their companies. The editors welcome any suggestions to improve future editions of this book.

The Editors

List of Acronyms

* An asterisk indicates that the acronym can have different meanings depending on the context. The meaning is clearly indicated in the text when used.

3GPP 3rd Generation Partnership Project

3GPP2 3rd Generation Partnership Project 2

ABS Almost Blank Subframe

AC Access Class

ACI Adjacent Channel Interference

ACIR Adjacent Channel Interference Ratio

ACK Acknowledgement

ACLR Adjacent Channel Leakage Ratio

ACS Adjacent Channel Selectivity

ADC Analogue to Digital Converter

ADSL Asymmetric Digital Subscriber Line

AGI Antenna Gain Imbalance

A-GNSS Assisted Global Navigation Satellite System

AM Acknowledged Mode

AMC Adaptive Modulation and Coding

AMPS Analogue Mobile Phone System

AMR Adaptive MultiRate

ANR Automatic Neighbour Relation

ANRF Automatic Neighbour Relation Function

AoA Angle-of-Arrival

AoD Angle-of-Departure

APN Access Point Name

APP A-Posteriori Probability

ARFCN Absolute Radio Frequency Channel Number

ARIB Association of Radio Industries and Businesses

ARP Almost Regular Permutation*

ARP Allocation and Retention Priority*

ARQ Automatic Repeat reQuest

AS Access Stratum*

AS Angular Spread*

A-SEM Additional SEM

ATDMA Advanced TDMA

ATIS Alliance for Telecommunications Industry Solutions

AuC Authentication Centre

AWGN Additive White Gaussian Noise

BCC Base station Colour Code

BCH Broadcast CHannel

BCCH Broadcast Control CHannel

BCJR Algorithm named after its inventors, Bahl, Cocke, Jelinek and Raviv

BER Bit Error Rate

BLER BLock Error Rate

BM-SC Broadcast-Multicast Service Centre

BP Belief Propagation

BPRE Bits Per Resource Element

bps bits per second

BPSK Binary Phase Shift Keying

BSIC Base Station Identification Code

BSR Buffer Status Reports

CAPEX CAPital EXpenditure

CAZAC Constant Amplitude Zero AutoCorrelation

CB Circular Buffer

CBF Coordinated Beamforming

CC Component Carrier

CCCH Common Control CHannel

CCE Control Channel Element

CCI Co-Channel Interference

CCO Cell Change Order

CCSA China Communications Standards Association

CDD Cyclic Delay Diversity

CDF Cumulative Distribution Function

CDL Clustered Delay Line

CDM Code Division Multiplex(ed/ing)

CDMA Code Division Multiple Access

C/I Carrier-to-Interference ratio

CID Cell ID

CIF Carrier Indicator Field

CF Contention-Free

CFI Control Format Indicator

CFO Carrier Frequency Offset

CINR Carrier-to-Interference-and-Noise Ratio

CIR Channel Impulse Response

CM Cubic Metric

CMAS Commercial Mobile Alert Service

CMHH Constant Modulus HouseHolder

CN Core Network

CoMP Coordinated MultiPoint

CODIT UMTS Code DIvision Testbed

COFDM Coded OFDM

CP Cyclic Prefix

CPICH Common PIlot CHannel

CPR Common Phase Rotation

CPT Control PDU Type

CQI Channel Quality Indicator

CRC Cyclic Redundancy Check

CRE Cell Range Expansion

C-RNTI Cell Radio Network Temporary Identifier

CRS Common Reference Signal

CS Circuit-Switched*

CS Cyclic Shift*

CSA Common Subframe Allocation

CSG Closed Subscriber Group

CSI Channel State Information

CSI-RS Channel State Information RS

CSIT Channel State Information at the Transmitter

CTF Channel Transfer Function

CVA Circular Viterbi Algorithm

CVQ Channel Vector Quantization

CW Continuous-Wave

DAB Digital Audio Broadcasting

DAC Digital to Analogue Converter

DAI Downlink Assignment Index

dB deci-Bel

d.c. direct current

DCCH Dedicated Control CHannel

DCFB Direct Channel FeedBack

DCI Downlink Control Information

DFT Discrete Fourier Transform

DFT-S-OFDM DFT-Spread OFDM

Diffserv Differentiated Services

DL DownLink

DL-SCH DownLink Shared CHannel

DMB Digital Mobile Broadcasting

DM-RS DeModulation-RS

DOA Direction Of Arrival

DPC Dirty-Paper Coding

DRB Data Radio Bearer

DRX Discontinuous Reception

DS-CDMA Direct-Sequence Code Division Multiple Access

DSP Digital Signal Processor

DTCH Dedicated Traffic CHannel

DTX Discontinuous Transmission

DVB-H Digital Video Broadcasting – Handheld

DVB-T Digital Video Broadcasting – Terrestrial

DwPTS Downlink Pilot TimeSlot

ECGI E-UTRAN Cell Global Identifier

ECM EPS Connection Management

EDGE Enhanced Data rates for GSM Evolution

EESM Exponential Effective SINR Mapping

eICIC enhanced Inter-Cell Interference Coordination

EMEA Europe, Middle East and Africa

EMM EPS Mobility Management

eNodeB evolved NodeB

EPA Extended Pedestrian A

EPC Evolved Packet Core

EPG Electronic Programme Guide

ePHR extended Power Headroom Report

EPS Evolved Packet System

E-RAB E-UTRAN Radio Access Bearer

E-SMLC Evolved Serving Mobile Location Centre

ESP Encapsulating Security Payload

ETSI European Telecommunications Standards Institute

ETU Extended Typical Urban

ETWS Earthquake and Tsunami Warning System

E-UTRA Evolved-UTRA

E-UTRAN Evolved-UTRAN

EVA Extended Vehicular A

EVM Error Vector Magnitude

FACH Forward Access CHannel

FB Frequency Burst

FCC Federal Communications Commission

FCCH Frequency Control CHannel

FDD Frequency Division Duplex

FDE Frequency-Domain Equalizer

FDM Frequency Division Multiplexing

FDMA Frequency Division Multiple Access

FDSS Frequency-Domain Spectral Shaping

FFT Fast Fourier Transform

FI Framing Info

FIR Finite Impulse Response

FMS First Missing SDU

FSTD Frequency Switched Transmit Diversity

FTP File Transfer Protocol

FTTH Fibre-To-The-Home

GBR Guaranteed Bit Rate

GCL Generalized Chirp-Like

GERAN GSM EDGE Radio Access Network

GGSN Gateway GPRS Support Node

GMSK Gaussian Minimum-Shift Keying

GNSS Global Navigation Satellite System

GPRS General Packet Radio Service

GPS Global Positioning System

GSM Global System for Mobile communications

GT Guard Time

GTP GPRS Tunnelling Protocol

GTP-U GTP-User plane

HARQ Hybrid Automatic Repeat reQuest

HD-FDD Half-Duplex FDD

HeNB Home eNodeB

HFN Hyper Frame Number

HII High Interference Indicator

HLR Home Location Register

HRPD High Rate Packet Data

HSDPA High Speed Downlink Packet Access

HSPA High Speed Packet Access

HSPA+ High Speed Packet Access Evolution

HSS Home Subscriber Server

HSUPA High Speed Uplink Packet Access

HTTP HyperText Transfer Protocol

ICI Inter-Carrier Interference

ICIC Inter-Cell Interference Coordination

IDFT Inverse Discrete Fourier Transform

IETF Internet Engineering Task Force

IFDMA Interleaved Frequency Division Multiple Access

IFFT Inverse Fast Fourier Transform

i.i.d. Independent identically distributed

IM Implementation Margin

IMD Inter-Modulation Distortion

IMS IP Multimedia Subsystem

IMSI International Mobile Subscriber Identity

IMT International Mobile Telecommunications

InH Indoor Hotspot

IP Internet Protocol

IR Incremental Redundancy

IRC Interference Rejection Combining

ISD Inter-Site Distance

ISI Inter-Symbol Interference

IST-WINNER Information Society Technologies - Wireless world INitiative NEw Radio

ITU International Telecommunication Union

ITU-R ITU Radiocommunication sector

J-TACS Japanese Total Access Communication System

JT Joint Transmission

LA Local Area

LAC Local Area Code

LB Long Block

LBP Layered Belief Propagation

LBRM Limited Buffer Rate Matching

LCID Logical Channel ID

LDPC Low-Density Parity Check

L-GW LIPA GateWay

LI Length Indicator

LIPA Local IP Access

LLR Log-Likelihood Ratio

LMMSE Linear MMSE

LNA Low Noise Amplifier

LO Local Oscillator

LOS Line-Of-Sight

LPP LTE Positioning Protocol

LS Least Squares

LSF Last Segment Flag

LTE Long-Term Evolution

MA Metropolitan Area

MAC Medium Access Control

MAC-I Message Authentication Code for Integrity

MAN Metropolitan Area Network

MAP Maximum A posteriori Probability

MBL Mobility Load Balancing

MBMS Multimedia Broadcast/Multicast Service

MBMS GW MBMS GateWay

MBR Maximum Bit Rate

MBSFN Multimedia Broadcast Single Frequency Network

MCCH Multicast Control CHannel

MCE Multicell Coordination Entity

MCH Multicast CHannel

MCL Minimum Coupling Loss

MCS Modulation and Coding Scheme

Mcps Megachips per second

MDS Minimum Discernible Signal

MDT Minimization of Drive Tests

MeNB Macro eNodeB

MIB Master Information Block

MIMO Multiple-Input Multiple-Output

MIP Mobile Internet Protocol

MISO Multiple-Input Single-Output

ML Maximum Likelihood

MLD Maximum Likelihood Detector

MME Mobility Management Entity

MMSE Minimum MSE

MO Mobile Originated

MOP Maximum Output Power

MPS Multimedia Priority Service

M-PSK M-ary Phase-Shift Keying

MQE Minimum Quantization Error

MRB Multicast Radio Bearer

MRC Maximum Ratio Combining

M-RNTI MBMS Radio Network Temporary Identifier

MRO Mobility Robustness Optimization

MSA MCH Subframe Allocation

MSAP MCH Subframe Allocation Pattern

MSB Most Significant Bit

MSD Maximum Sensitivity Degradation

MSE Mean Squared Error

MSI MCH Scheduling Information

MSISDN Mobile Station International Subscriber Directory Number

MSP MCH Scheduling Period

MSR Maximum Sensitivity Reduction

MTC Machine-Type Communications

MTCH Multicast Traffic CHannel

MU-MIMO Multi-User MIMO

MUE Macro User Equipment

NACC Network Assisted Cell Change

NACK Negative ACKnowledgement

NACS NonAdjacent Channel Selectivity

NAS Non Access Stratum

NCC Network Colour Code

NCL Neighbour Cell List

NDI New Data Indicator

NF Noise Figure

NGMN Next Generation Mobile Networks

NLM Network Listen Mode

NLMS Normalized Least-Mean-Square

NLOS Non-Line-Of-Sight

NMT Nordic Mobile Telephone

NNSF NAS Node Selection Function

NodeB The base station in WCDMA systems

NR Neighbour cell Relation

NRT Neighbour Relation Table

O&M Operation and Maintenance

OBPD Occupied Bandwidth Power De-rating

OBW Occupied BandWidth

OCC Orthogonal Cover Code

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

OPEX OPerational Expenditure

OSG Open Subscriber Group

OTDOA Observed Time Difference Of Arrival

OI Overload Indicator

OMA Open Mobile Alliance

OOB Out-Of-Band

P/S Parallel-to-Serial

PA Power Amplifier

PAN Personal Area Network

PAPR Peak-to-Average Power Ratio

PBCH Physical Broadcast CHannel

PBR Prioritized Bit Rate

PCC Policy Control and Charging*

PCC Primary Component Carrier*

PCCH Paging Control CHannel

P-CCPCH Primary Common Control Physical CHannel

PCEF Policy Control Enforcement Function

PCell Primary serving Cell

PCFICH Physical Control Format Indicator CHannel

PCG Project Coordination Group

PCH Paging CHannel

PCI Physical Cell Identity

P-CPICH Primary Common PIlot CHannel

PCRF Policy Control and charging Rules Function

PDCCH Physical Downlink Control CHannel

PDCP Packet Data Convergence Protocol

PDN Packet Data Network

PDP Power Delay Profile

PDSCH Physical Downlink Shared CHannel

PDU Protocol Data Unit

PF Paging Frame

PFS Proportional Fair Scheduling

P-GW PDN GateWay

PHICH Physical Hybrid ARQ Indicator CHannel

PHR Power Headroom Report

PLL Phase-Locked Loop

PLMN Public Land Mobile Network

P-MCCH Primary MCCH

PMCH Physical Multicast CHannel

PMI Precoding Matrix Indicators

PMIP Proxy MIP

PN Pseudo-Noise
PO Paging Occasion
PRACH Physical Random Access CHannel
PRB Physical Resource Block
P-RNTI Paging RNTI
PRG Precoder Resource block Group
PRS Positioning Reference Signal
PS Packet-Switched
P-SCH Primary Synchronization CHannel
PSD Power Spectral Density
PSS Primary Synchronization Signal
PTI Precoder Type Indication
PUCCH Physical Uplink Control CHannel
PUSCH Physical Uplink Shared CHannel
PVI Precoding Vector Indicator
PWS Public Warning System
QAM Quadrature Amplitude Modulation
QCI QoS Class Identifier
QoS Quality-of-Service
QPP Quadratic Permutation Polynomial
QPSK Quadrature Phase Shift Keying
RA Random Access
RAC Routing Area Code
RACH Random Access CHannel
RAN Radio Access Network
RAR Random Access Response
RA-RNTI Random Access Radio Network Temporary Identifier
RAT Radio Access Technology
RB Resource Block
RE Resource Element
REG Resource Element Group
RF Radio Frequency*
RF Resegmentation Flag*
RFC Request For Comments
RI Rank Indicator
RIM RAN Information Management
RIT Radio Interface Technology
RLC Radio Link Control
RLF Radio Link Failure
RLS Recursive Least Squares
RM Rate Matching*
RM Reed-Muller*
RMa Rural Macrocell
RN Relay Node
RNC Radio Network Controller
RNTI Radio Network Temporary Identifier
RNTP Relative Narrowband Transmit Power
ROHC RObust Header Compression
RoT Rise over Thermal
R-PDCCH Relay Physical Downlink Control Channel
RPRE Received Power per Resource Element
RPF RePetition Factor
R-PLMN Registered PLMN
RRC Radio Resource Control*
RRC Root-Raised-Cosine*
RRH Remote Radio Head
RRM Radio Resource Management
RS Reference Signal
RSCP Received Signal Code Power
RSRP Reference Signal Received Power
RSRQ Reference Signal Received Quality
RSSI Received Signal Strength Indicator
RSTD Reference Signal Time Difference
RTCP Real-time Transport Control Protocol
RTD Round-Trip Delay
RTP Real-time Transport Protocol
RTT Round-Trip Time
RV Redundancy Version
S/P Serial-to-Parallel
S1AP S1 Application Protocol
SAE System Architecture Evolution
SAP Service Access Point
SAW Stop-And-Wait
SB Short Block*
SB Synchronization Burst*
SBP Systematic Bit Puncturing

SCC Secondary Component Carrier

SC-FDMA Single-Carrier Frequency Division Multiple Access

SCH Synchronization CHannel

SCM Spatial Channel Model

SCME Spatial Channel Model Extension

SCTP Stream Control Transmission Protocol

SDMA Spatial Division Multiple Access

SDO Standards Development Organization

SDU Service Data Unit

SeGW Security GateWay

SEM Spectrum Emission Mask

SFBC Space-Frequency Block Code

SFDR Spurious-Free Dynamic Range

SFN System Frame Number

SGSN Serving GPRS Support Node

S-GW Serving GateWay

SI System Information

SIB System Information Block

SIC Successive Interference Cancellation

SIMO Single-Input Multiple-Output

SINR Signal-to-Interference plus Noise Ratio

SIP Session Initiation Protocol

SIR Signal-to-Interference Ratio

SI-RNTI System Information Radio Network Temporary Identifier

SISO Single-Input Single-Output*

SISO Soft-Input Soft-Output*

SLP SUPL Location Platform

S-MCCH Secondary MCCH

SMS Short Message Service

SN Sequence Number

SNR Signal-to-Noise Ratio

SO Segmentation Offset

SON Self-Optimizing Networks

SORTD Space Orthogonal-Resource Transmit Diversity

SPA Sum-Product Algorithm

SPS Semi-Persistent Scheduling

SPS-C-RNTI Semi-Persistent Scheduling C-RNTI

SR Scheduling Request

SRB Signalling Radio Bearer

SRIT Set of Radio Interface Technology

SRNS Serving Radio Network Subsystem

SRS Sounding Reference Signal

S-SCH Secondary Syncronization CHannel

SSS Secondary Synchronization Signal

STBC Space-Time Block Code

S-TMSI SAE-Temporary Mobile Subscriber Identity

STTD Space-Time Transmit Diversity

SU-MIMO Single-User MIMO

SUPL Secure User Plane Location

SVD Singular-Value Decomposition

TA Tracking Area

TAC Tracking Area Code

TACS Total Access Communication System

TAI Tracking Area Identity

TB Transport Block

TCP Transmission Control Protocol

TDC Time-Domain Coordination

TDD Time Division Duplex

TDL Tapped Delay Line

TDMA Time Division Multiple Access

TD-SCDMA Time Division Synchronous Code Division Multiple Access

TEID Tunnelling End ID

TF Transport Format

TFT Traffic Flow Template

TM Transparent Mode

TMD Transparent Mode Data

TNL Transport Network Layer

TNMSE Truncated Normalized Mean-Squared Error

TPC Transmitter Power Control

TPD Total Power De-rating

TPMI Transmitted Precoding Matrix Indicator

TR Tone Reservation

TSC Training Sequence Code

TSG Technical Specification Group

TTA Telecommunications Technology Association

TTC Telecommunications Technology Committee

TTFF Time To First Fix

TTI Transmission Time Interval

TU Typical Urban

UCI Uplink Control Information

UDP User Datagram Protocol

UE User Equipment

UL UpLink

ULA Uniform Linear Array

UL-SCH UpLink Shared CHannel

UM Unacknowledged Mode

UMa Urban Macrocell

UMB Ultra-Mobile Broadband

UMi Urban Microcell

UMTS Universal Mobile Telecommunications System

UP Unitary Precoding

UpPTS Uplink Pilot TimeSlot

US Uncorrelated-Scattered

USIM Universal Subscriber Identity Module

UTRA Universal Terrestrial Radio Access

UTRAN Universal Terrestrial Radio Access Network

VA Viterbi Algorithm

VCB Virtual Circular Buffer

VCO Voltage-Controlled Oscillator

VoIP Voice-over-IP

VRB Virtual Resource Block

WA Wide Area

WAN Wide Area Network

WCDMA Wideband Code Division Multiple Access

WFT Winograd Fourier Transform

WG Working Group

WiMAX Worldwide interoperability for Microwave Access

WINNER Wireless world INitiative NEw Radio

WLAN Wireless Local Area Network

WPD Waveform Power De-rating

WRC World Radiocommunication Conference

WSS Wide-Sense Stationary

WSSUS Wide-Sense Stationary Uncorrelated Scattering

ZC Zadoff–Chu

ZCZ Zero Correlation Zone

ZF Zero-Forcing

ZFEP Zero-Forcing Equal Power

1

Introduction and Background

Thomas Sälzer and Matthew Baker

1.1 The Context for the Long Term Evolution of UMTS

1.1.1 Historical Context

The Long Term Evolution of UMTS is one of the latest steps in an advancing series of mobile telecommunications systems. Arguably, at least for land-based systems, the series began in 1947 with the development of the concept of *cells* by Bell Labs, USA. The use of cells enabled the capacity of a mobile communications network to be increased substantially, by dividing the coverage area up into small cells each with its own base station operating on a different frequency.

The early systems were confined within national boundaries. They attracted only a small number of users, as the equipment on which they relied was expensive, cumbersome and power-hungry, and therefore was only really practical in a car.

The first mobile communication systems to see large-scale commercial growth arrived in the 1980s and became known as the 'First Generation' systems. The First Generation used analogue technology and comprised a number of independently developed systems worldwide (e.g. AMPS (Analogue Mobile Phone System, used in America), TACS (Total Access Communication System, used in parts of Europe), NMT (Nordic Mobile Telephone, used in parts of Europe) and J-TACS (Japanese Total Access Communication System, used in Japan and Hong Kong)).

Global roaming first became a possibility with the development of the 'Second Generation' system known as GSM (Global System for Mobile communications), which was based on digital technology. The success of GSM was due in part to the collaborative spirit in which it was developed. By harnessing the creative expertise of a number of companies working

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

together under the auspices of the European Telecommunications Standards Institute (ETSI), GSM became a robust, interoperable and widely accepted standard.

Fuelled by advances in mobile handset technology, which resulted in small, fashionable terminals with a long battery life, the widespread acceptance of the GSM standard exceeded initial expectations and helped to create a vast new market. The resulting near-universal penetration of GSM phones in the developed world provided an ease of communication never previously possible, first by voice and text message, and later also by more advanced data services. Meanwhile in the developing world, GSM technology had begun to connect communities and individuals in remote regions where fixed-line connectivity was non-existent and would be prohibitively expensive to deploy.

This ubiquitous availability of user-friendly mobile communications, together with increasing consumer familiarity with such technology and practical reliance on it, thus provides the context for new systems with more advanced capabilities. In the following section, the series of progressions which have succeeded GSM is outlined, culminating in the development of the system known as LTE – the Long Term Evolution of UMTS (Universal Mobile Telecommunications System).

1.1.2 LTE in the Mobile Radio Landscape

In contrast to transmission technologies using media such as copper lines and optical fibres, the radio spectrum is a medium shared between different, and potentially interfering, technologies.

As a consequence, regulatory bodies – in particular, ITU-R (International Telecommunication Union – Radiocommunication Sector) [1], but also regional and national regulators – play a key role in the evolution of radio technologies since they decide which parts of the spectrum and how much bandwidth may be used by particular types of service and technology. This role is facilitated by the *standardization* of families of radio technologies – a process which not only provides specified interfaces to ensure interoperability between equipment from a multiplicity of vendors, but also aims to ensure that the allocated spectrum is used as efficiently as possible, so as to provide an attractive user experience and innovative services.

The complementary functions of the regulatory authorities and the standardization organizations can be summarized broadly by the following relationship:

$$\text{Aggregated data rate} = \underbrace{\text{bandwidth}}_{\substack{\text{regulation and licences} \\ \text{(ITU-R, regional regulators)}}} \times \underbrace{\text{spectral efficiency}}_{\substack{\text{technology} \\ \text{and standards}}}$$

On a worldwide basis, ITU-R defines technology families and associates specific parts of the spectrum with these families. Facilitated by ITU-R, spectrum for mobile radio technologies is identified for the radio technologies which meet ITU-R's requirements to be designated as members of the *International Mobile Telecommunications* (IMT) family. Effectively, the IMT family comprises systems known as 'Third Generation' (for the first time providing data rates up to 2 Mbps) and beyond.

From the technology and standards angle, three main organizations have recently been developing standards relevant to IMT requirements, and these organisations continue to shape the landscape of mobile radio systems as shown in Figure 1.1.

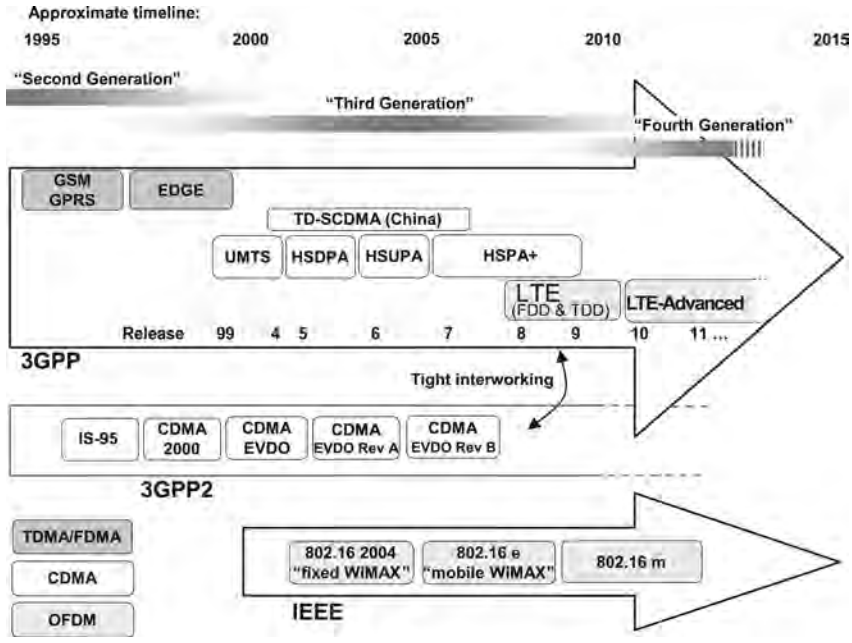


Figure 1.1: Approximate timeline of the mobile communications standards landscape.

The uppermost evolution track shown in Figure 1.1 is that developed in the 3rd Generation Partnership Project (3GPP), which is currently the dominant standards development group for mobile radio systems and is described in more detail below.

Within the 3GPP evolution track, three multiple access technologies are evident: the 'Second Generation' GSM/GPRS/EDGE family[1] was based on Time- and Frequency-Division Multiple Access (TDMA/FDMA); the 'Third Generation' UMTS family marked the entry of Code Division Multiple Access (CDMA) into the 3GPP evolution track, becoming known as *Wideband* CDMA (owing to its 5 MHz carrier bandwidth) or simply WCDMA; finally LTE has adopted Orthogonal Frequency-Division Multiplexing (OFDM), which is the access technology dominating the latest evolutions of all mobile radio standards.

In continuing the technology progression from the GSM and UMTS technology families within 3GPP, the LTE system can be seen as completing the trend of expansion of service provision beyond voice calls towards a multiservice air interface. This was already a key aim of UMTS and GPRS/EDGE, but LTE was designed from the start with the goal of evolving the radio access technology under the assumption that all services would be packet-switched, rather than following the circuit-switched model of earlier systems. Furthermore, LTE is accompanied by an evolution of the non-radio aspects of the complete system, under the term 'System Architecture Evolution' (SAE) which includes the Evolved Packet Core (EPC) network. Together, LTE and SAE comprise the Evolved Packet System (EPS), where both the core network and the radio access are fully packet-switched.

[1]The maintenance and development of specifications for the GSM family was passed to 3GPP from ETSI.

The standardization of LTE and SAE does not mean that further development of the other radio access technologies in 3GPP has ceased. In particular, the enhancement of UMTS with new releases of the specifications continues in 3GPP, to the greatest extent possible while ensuring backward compatibility with earlier releases: the original 'Release 99' specifications of UMTS have been extended with high-speed downlink and uplink enhancements (HSDPA[2] and HSUPA[3] in Releases 5 and 6 respectively), known collectively as 'HSPA' (High-Speed Packet Access). HSPA has been further enhanced in Release 7 (becoming known as HSPA+) with higher-order modulation and, for the first time in a cellular communication system, multistream 'MIMO'[4] operation, while Releases 8, 9 and 10 introduce support for multiple 5 MHz carriers operating together in downlink and uplink. These backward-compatible enhancements enable network operators who have invested heavily in the WCDMA technology of UMTS to generate new revenues from new features while still providing service to their existing subscribers using legacy terminals.

The first version of LTE was made available in Release 8 of the 3GPP specification series. It was able to benefit from the latest understanding and technology developments from HSPA and HSPA+, especially in relation to optimizations of the protocol stack, while also being free to adopt radical new technology without the constraints of backward compatibility or a 5 MHz carrier bandwidth. However, LTE also has to satisfy new demands, for example in relation to spectrum flexibility for deployment. LTE can operate in Frequency-Division Duplex (FDD) and Time-Division Duplex (TDD) modes in a harmonized framework designed also to support the evolution of TD-SCDMA (Time-Division Synchronous Code Division Multiple Access), which was developed in 3GPP as an additional branch of the UMTS technology path, essentially for the Chinese market.

A second version of LTE was developed in Release 9, and Release 10 continues the progression with the beginning of the next significant step known as LTE-Advanced.

A second evolution track shown in Figure 1.1 is led by a partnership organization similar to 3GPP and known as 3GPP2. CDMA2000 was developed based on the American 'IS-95' standard, which was the first mobile cellular communication system to use CDMA technology; it was deployed mainly in the USA, Korea and Japan. Standardization in 3GPP2 has continued with parallel evolution tracks towards data-oriented systems (EV-DO), to a certain extent taking a similar path to the evolutions in 3GPP. It is important to note that LTE will provide tight interworking with systems developed by 3GPP2, which allows a smooth migration to LTE for operators who previously followed the 3GPP2 track.

The third path of evolution has emerged from the IEEE 802 LAN/MAN[5] standards committee, which created the '802.16' family as a broadband wireless access standard. This family is also fully packet-oriented. It is often referred to as *WiMAX*, on the basis of a so-called 'System Profile' assembled from the 802.16 standard and promoted by the WiMAX Forum. The WiMAX Forum also ensures the corresponding product certification. While the first version, known as 802.16-2004, was restricted to fixed access, the following version 802.16e includes basic support of mobility and is therefore often referred to as 'mobile WiMAX'. However, it can be noted that in general the WiMAX family has not been designed with the same emphasis on mobility and compatibility with operators' core networks as the

[2]High-Speed Downlink Packet Access.
[3]High-Speed Uplink Packet Access.
[4]Multiple-Input Multiple-Output antenna system.
[5]Local Area Network / Metropolitan Area Network.

3GPP technology family, which includes core network evolutions in addition to the radio access network evolution. Nevertheless, the latest generation developed by the IEEE, known as 802.16m, has similar targets to LTE-Advanced which are outlined in Chapter 27.

The overall pattern is of an evolution of mobile radio towards flexible, packet-oriented, multiservice systems. The aim of all these systems is towards offering a mobile broadband user experience that can approach that of current fixed access networks such as Asymmetric Digital Subscriber Line (ADSL) and Fibre-To-The-Home (FTTH).

1.1.3 The Standardization Process in 3GPP

The collaborative standardization model which so successfully produced the GSM system became the basis for the development of UMTS. In the interests of producing truly global standards, the collaboration for both GSM and UMTS was expanded beyond ETSI to encompass regional Standards Development Organizations (SDOs) from Japan (ARIB and TTC), Korea (TTA), North America (ATIS) and China (CCSA), as shown in Figure 1.2.

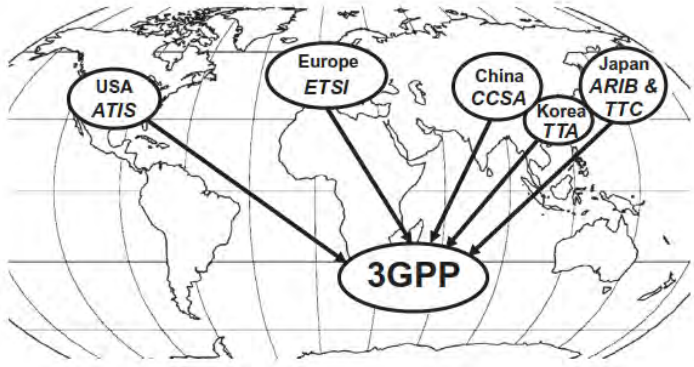


Figure 1.2: 3GPP is a global partnership of six regional SDOs.

So the 3GPP was born and by 2011 boasted 380 individual member companies.

The successful creation of such a large and complex system specification as that for UMTS or LTE requires a well-structured organization with pragmatic working procedures. 3GPP is divided into four Technical Specification Groups (TSGs), each of which is comprised of a number of Working Groups (WGs) with responsibility for a specific aspect of the specifications as shown in Figure 1.3.

A distinctive feature of the working methods of these groups is the consensus-driven approach to decision-making.

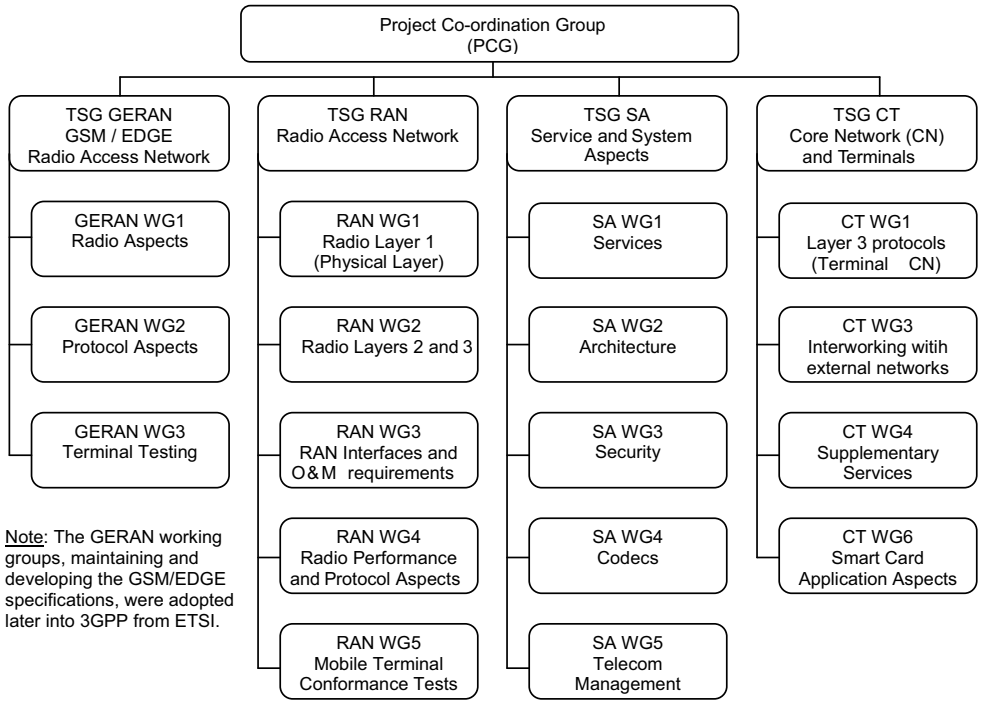


Figure 1.3: The Working Group structure of 3GPP. Reproduced by permission of © 3GPP.

All documents submitted to 3GPP are publicly available on the 3GPP website,[6] including contributions from individual companies, technical reports and technical specifications.

In reaching consensus around a technology, the WGs take into account a variety of considerations, including but not limited to performance, implementation cost, complexity and compatibility with earlier versions or deployments. Simulations are frequently used to compare performance of different techniques, especially in the WGs focusing on the physical layer of the air interface and on performance requirements. This requires consensus first to be reached around the simulation assumptions to be used for the comparison, including, in particular, understanding and defining the scenarios of interest to network operators.

The LTE standardization process was inaugurated at a workshop in Toronto in November 2004, when a broad range of companies involved in the mobile communications business presented their visions for the future evolution of the specifications to be developed in 3GPP. These visions included both initial perceptions of the *requirements* which needed to be satisfied, and proposals for *suitable technologies* to meet those requirements.

The requirements are reviewed in detail in Section 1.2, while the key technologies are introduced in Section 1.3.

[6] www.3gpp.org.

1.2 Requirements and Targets for the Long Term Evolution

Discussion of the key requirements for the new LTE system led to the creation of a formal 'Study Item' in 3GPP with the specific aim of 'evolving' the 3GPP radio access technology to ensure competitiveness over a ten-year time-frame. Under the auspices of this Study Item, the requirements for LTE Release 8 were refined and crystallized, being finalized in June 2005.

They can be summarized as follows:

- reduced delays, in terms of both connection establishment and transmission latency;
- increased user data rates;
- increased cell-edge bit-rate, for uniformity of service provision;
- reduced cost per bit, implying improved spectral efficiency;
- greater flexibility of spectrum usage, in both new and pre-existing bands;
- simplified network architecture;
- seamless mobility, including between different radio-access technologies;
- reasonable power consumption for the mobile terminal.

It can also be noted that network operator requirements for next generation mobile systems were formulated by the Next Generation Mobile Networks (NGMN) alliance of network operators [2], which served as an additional reference for the development and assessment of the LTE design. Such operator-driven requirements have also guided the development of LTE-Advanced (see Chapters 27 to 31).

To address these objectives, the LTE system design covers both the radio interface and the radio network architecture.

1.2.1 System Performance Requirements

Improved system performance compared to existing systems is one of the main requirements from network operators, to ensure the competitiveness of LTE and hence to arouse market interest. In this section, we highlight the main performance metrics used in the definition of the LTE requirements and its performance assessment.

Table 1.1 summarizes the main performance requirements to which the first release of LTE was designed. Many of the figures are given relative to the performance of the most advanced available version of UMTS, which at the time of the definition of the LTE requirements was HSDPA/HSUPA Release 6 – referred to here as the *reference baseline*. It can be seen that the target requirements for LTE represent a significant step from the capacity and user experience offered by the third generation mobile communications systems which were being deployed at the time when the first version of LTE was being developed.

As mentioned above, HSPA technologies are also continuing to be developed to offer higher spectral efficiencies than were assumed for the reference baseline. However, LTE has been able to benefit from avoiding the constraints of backward compatibility, enabling the inclusion of advanced MIMO schemes in the system design from the beginning, and highly flexible spectrum usage built around new multiple access schemes.

Table 1.1: Summary of key performance requirement targets for LTE Release 8.

		Absolute requirement	Release 6 (for comparison)	Comments
Downlink	Peak transmission rate	> 100 Mbps	14.4 Mbps	LTE in 20 MHz FDD, 2 × 2 spatial multiplexing. Reference: HSDPA in 5 MHz FDD, single antenna transmission
	Peak spectral efficiency	> 5 bps/Hz	3 bps/Hz	
	Average cell spectral efficiency	> 1.6 2.1 bps/Hz/cell	0.53 bps/Hz/cell	LTE: 2 × 2 spatial multiplexing, Interference Rejection Combining (IRC) receiver [3]. Reference: HSDPA, Rake receiver [4], 2 receive antennas
	Cell edge spectral efficiency	> 0.04 0.06 bps/Hz/user	0.02 bps/Hz/user	As above, 10 users assumed per cell
	Broadcast spectral efficiency	> 1 bps/Hz	N/A	Dedicated carrier for broadcast mode
Uplink	Peak transmission rate	> 50 Mbps	11 Mbps	LTE in 20 MHz FDD, single antenna transmission. Reference: HSUPA in 5 MHz FDD, single antenna transmission
	Peak spectral efficiency	> 2.5 bps/Hz	2 bps/Hz	
	Average cell spectral efficiency	> 0.66 1.0 bps/Hz/cell	0.33 bps/Hz/cell	LTE: single antenna transmission, IRC receiver [3]. Reference: HSUPA, Rake receiver [4], 2 receive antennas
	Cell edge spectral efficiency	> 0.02 0.03 bps/Hz/user	0.01 bps/Hz/user	As above, 10 users assumed per cell
System	User plane latency (two way radio delay)	< 10 ms		LTE target approximately one fifth of Reference.
	Connection set up latency	< 100 ms		Idle state → active state
	Operating bandwidth	1.4 20 MHz	5 MHz	(initial requirement started at 1.25 MHz)
	VoIP capacity	NGMN preferred target expressed in [2] is > 60 sessions/MHz/cell		

The requirements shown in Table 1.1 are discussed and explained in more detail below. Chapter 26 shows how the overall performance of the LTE system meets these requirements.

1.2.1.1 Peak Rates and Peak Spectral Efficiency

For marketing purposes, the first parameter by which different radio access technologies are usually compared is the peak per-user data rate which can be achieved. This peak data rate generally scales according to the amount of spectrum used, and, for MIMO systems, according to the minimum of the number of transmit and receive antennas (see Section 11.1).

The peak data rate can be defined as the maximum throughput per user assuming the whole bandwidth being allocated to a single user with the highest modulation and coding scheme and the maximum number of antennas supported. Typical radio interface overhead (control channels, pilot signals, guard intervals, etc.) is estimated and taken into account for a given operating point. For TDD systems, the peak data rate is generally calculated for the downlink

and uplink periods separately. This makes it possible to obtain a single value independent of the uplink/downlink ratio and a fair system comparison that is agnostic of the duplex mode. The maximum spectral efficiency is then obtained simply by dividing the peak rate by the used spectrum allocation.

The target peak data rates for downlink and uplink in LTE Release 8 were set at 100 Mbps and 50 Mbps respectively within a 20 MHz bandwidth,[7] corresponding to respective peak spectral efficiencies of 5 and 2.5 bps/Hz. The underlying assumption here is that the terminal has two receive antennas and one transmit antenna. The number of antennas used at the base station is more easily upgradeable by the network operator, and the first version of the LTE specifications was therefore designed to support downlink MIMO operation with up to four transmit and receive antennas. The MIMO techniques enabling high peak data rates are described in detail in Chapter 11.

When comparing the capabilities of different radio communication technologies, great emphasis is often placed on the peak data rate capabilities. While this is one indicator of how technologically advanced a system is and can be obtained by simple calculations, it may not be a key differentiator in the usage scenarios for a mobile communication system in practical deployment. Moreover, it is relatively easy to design a system that can provide very high peak data rates for users close to the base station, where interference from other cells is low and techniques such as MIMO can be used to their greatest extent. It is much more challenging to provide high data rates with good coverage and mobility, but it is exactly these latter aspects which contribute most strongly to user satisfaction.

In typical deployments, individual users are located at varying distances from the base stations, the propagation conditions for radio signals to individual users are rarely ideal, and the available resources must be shared between many users. Consequently, although the claimed peak data rates of a system are genuinely achievable in the right conditions, it is rare for a single user to be able to experience the peak data rates for a sustained period, and the envisaged applications do not usually require this level of performance.

A differentiator of the LTE system design compared to some other systems has been the recognition of these 'typical deployment constraints' from the beginning. During the design process, emphasis was therefore placed not only on providing a competitive peak data rate for use when conditions allow, but also importantly on *system level performance*, which was evaluated during several performance verification steps.

System-level evaluations are based on simulations of multicell configurations where data transmission from/to a population of mobiles is considered in a typical deployment scenario. The sections below describe the main metrics used as requirements for system level performance. In order to make these metrics meaningful, parameters such as the deployment scenario, traffic models, channel models and system configuration need to be defined.

The key definitions used for the system evaluations of LTE Release 8 can be found in an input document from network operators addressing the performance verification milestone in the LTE development process [5]. This document takes into account deployment scenarios and channel models agreed during the LTE Study Item [6], and is based on an evaluation methodology elaborated by NGMN operators in [7]. The reference deployment scenarios which were given special consideration for the LTE performance evaluation covered macrocells with base station separations of between 500 m and 1.7 km, as well as microcells using MIMO with base station separations of 130 m. A range of mobile terminal

[7] Four times the bandwidth of a WCDMA carrier.

speeds were studied, focusing particularly on the range 3–30 km/h, although higher mobile speeds were also considered important.

1.2.1.2 Cell Throughput and Spectral Efficiency

Performance at the cell level is an important criterion, as it relates directly to the number of cell sites that a network operator requires, and hence to the capital cost of deploying the system. For LTE Release 8, it was chosen to assess the cell level performance with full-queue traffic models (i.e. assuming that there is never a shortage of data to transmit if a user is given the opportunity) and a relatively high system load, typically 10 users per cell.

The requirements at the cell level were defined in terms of the following metrics:

- Average cell throughput [bps/cell] and spectral efficiency [bps/Hz/cell];
- Aaverage user throughput [bps/user] and spectral efficiency [bps/Hz/user];
- Cell-edge user throughput [bps/user] and spectral efficiency [bps/Hz/user] (the metric used for this assessment is the 5-percentile user throughput, obtained from the cumulative distribution function of the user throughput).

For the UMTS Release 6 reference baseline, it was assumed that both the terminal and the base station use a single transmit antenna and two receive antennas; for the terminal receiver the assumed performance corresponds to a two-branch Rake receiver [4] with linear combining of the signals from the two antennas.

For the LTE system, the use of two transmit and receive antennas was assumed at the base station. At the terminal, two receive antennas were assumed, but still only a single transmit antenna. The receiver for both downlink and uplink is assumed to be a linear receiver with optimum combining of the signals from the antenna branches [3].

The original requirements for the cell level metrics were only expressed as relative gains compared to the Release 6 reference baseline. The absolute values provided in Table 1.1 are based on evaluations of the reference system performance that can be found in [8] and [9] for downlink and uplink respectively.

1.2.1.3 Voice Capacity

Unlike full queue traffic (such as file download) which is typically delay-tolerant and does not require a guaranteed bit-rate, real-time traffic such as Voice over IP (VoIP) has tight delay constraints. It is important to set system capacity requirements for such services – a particular challenge in fully packet-based systems like LTE which rely on adaptive scheduling.

The system capacity requirement is defined as the number of satisfied VoIP users, given a particular traffic model and delay constraints. The details of the traffic model used for evaluating LTE can be found in [5]. Here, a VoIP user is considered to be in outage (i.e. not satisfied) if more than 2% of the VoIP packets do not arrive successfully at the radio receiver within 50 ms and are therefore discarded. This assumes an overall end-to-end delay (from mobile terminal to mobile terminal) below 200 ms. The system capacity for VoIP can then be defined as the number of users present per cell when more than 95% of the users are satisfied.

The NGMN group of network operators expressed a preference for the ability to support 60 satisfied VoIP sessions per MHz – an increase of two to four times what can typically be achieved in the Release 6 reference case.

1.2.1.4 Mobility and Cell Ranges

LTE is required to support communication with terminals moving at speeds of up to 350 km/h, or even up to 500 km/h depending on the frequency band. The primary scenario for operation at such high speeds is usage on high-speed trains – a scenario which is increasing in importance across the world as the number of high-speed rail lines increases and train operators aim to offer an attractive working environment to their passengers. These requirements mean that handover between cells has to be possible without interruption – in other words, with imperceptible delay and packet loss for voice calls, and with reliable transmission for data services.

These targets are to be achieved by the LTE system in typical cells of radius up to 5 km, while operation should continue to be possible for cell ranges of 100 km and more, to enable wide-area deployments.

1.2.1.5 Broadcast Mode Performance

The requirements for LTE included the integration of an efficient broadcast mode for high rate Multimedia Broadcast/Multicast Services (MBMS) such as mobile TV, based on a Single Frequency Network mode of operation as explained in detail in Chapter 13. The spectral efficiency requirement is given in terms of a carrier dedicated to broadcast transmissions – i.e. not shared with unicast transmissions.

In broadcast systems, the system throughput is limited to what is achievable for the users in the worst conditions. Consequently, the broadcast performance requirement was defined in terms of an achievable system throughput (bps) and spectral efficiency (bps/Hz) assuming a coverage of 98% of the nominal coverage area of the system. This means that only 2% of the locations in the nominal coverage area are in outage – where outage for broadcast services is defined as experiencing a packet error rate higher than 1%. This broadcast spectral efficiency requirement was set to 1 bps/Hz [10].

While the broadcast mode was not available in Release 8 due to higher prioritization of other service modes, Release 9 incorporates a broadcast mode employing Single Frequency Network operation on a mixed unicast-broadcast carrier.

1.2.1.6 User Plane Latency

User plane latency is an important performance metric for real-time and interactive services. On the radio interface, the minimum user plane latency can be calculated based on signalling analysis for the case of an unloaded system. It is defined as the average time between the first transmission of a data packet and the reception of a physical layer acknowledgement. The calculation should include typical HARQ[8] retransmission rates (e.g. 0–30%). This definition therefore considers the capability of the system design, without being distorted by the scheduling delays that would appear in the case of a loaded system. The round-trip latency is obtained simply by multiplying the one-way user plane latency by a factor of two.

LTE is also required to be able to operate with an IP-layer one-way data-packet latency across the radio access network as low as 5 ms in optimal conditions. However, it is recognized that the actual delay experienced in a practical system will be dependent on system loading and radio propagation conditions. For example, HARQ plays a key role in

[8]Hybrid Automatic Repeat reQuest – see Section 10.3.2.5.

maximizing spectral efficiency at the expense of increased delay while retransmissions take place, whereas maximal spectral efficiency may not be essential in situations when minimum latency is required.

1.2.1.7 Control Plane Latency and Capacity

In addition to the user plane latency requirement, call setup delay was required to be significantly reduced compared to previous cellular systems. This not only enables a good user experience but also affects the battery life of terminals, since a system design which allows a fast transition from an idle state to an active state enables terminals to spend more time in the low-power idle state.

Control plane latency is measured as the time required for performing the transitions between different LTE states. LTE is based on only two main states, 'RRC_IDLE' and 'RRC_CONNECTED' (i.e. 'active') (see Section 3.1).

LTE is required to support transition from idle to active in less than 100 ms (excluding paging delay and Non-Access Stratum (NAS) signalling delay).

The LTE system capacity is dependent not only on the supportable throughput but also on the number of users simultaneously located within a cell which can be supported by the control signalling. For the latter aspect, LTE is required to support at least 200 active-state users per cell for spectrum allocations up to 5 MHz, and at least 400 users per cell for wider spectrum allocations; only a small subset of these users would be actively receiving or transmitting data at any given time instant, depending, for example, on the availability of data to transmit and the prevailing radio channel conditions. An even larger number of non-active users may also be present in each cell, and therefore able to be paged or to start transmitting data with low latency.

1.2.2 Deployment Cost and Interoperability

Besides the system performance aspects, a number of other considerations are important for network operators. These include reduced deployment cost, spectrum flexibility and enhanced interoperability with legacy systems – essential requirements to enable deployment of LTE networks in a variety of scenarios and to facilitate migration to LTE.

1.2.2.1 Spectrum Allocations and Duplex Modes

As demand for suitable radio spectrum for mobile communications increases, LTE is required to be able to operate in a wide range of frequency bands and sizes of spectrum allocations in both uplink and downlink. LTE can use spectrum allocations ranging from 1.4 to 20 MHz with a single carrier and addresses all frequency bands currently identified for IMT systems by ITU-R [1] including those below 1 GHz.

This will include deploying LTE in spectrum currently occupied by older radio access technologies – a practice often known as 'spectrum refarming'.

New frequency bands are continually being introduced for LTE in a release-independent way, meaning that any of the LTE Releases can be deployed in a new frequency band once the Radio-Frequency (RF) requirements have been specified [11].

The ability to operate in both paired and unpaired spectrum is required, depending on spectrum availability (see Chapter 23). LTE provides support for FDD, TDD and half-duplex

FDD operation in a unified design, ensuring a high degree of commonality which facilitates implementation of multimode terminals and allows worldwide roaming.

Starting from Release 10, LTE also provides means for flexible spectrum use via aggregation of contiguous and non-contiguous spectrum assets for high data rate services using a total bandwidth of up to 100 MHz (see Chapter 28).

1.2.2.2 Inter-Working with Other Radio Access Technologies

Flexible interoperation with other radio access technologies is essential for service continuity, especially during the migration phase in early deployments of LTE with partial coverage, where handover to legacy systems will often occur.

LTE relies on an evolved packet core network which allows interoperation with various access technologies, in particular earlier 3GPP technologies (GSM/EDGE and UTRAN[9]) as well as non-3GPP technologies (e.g. WiFi, CDMA2000 and WiMAX).

However, service continuity and short interruption times can only be guaranteed if measurements of the signals from other systems and fast handover mechanisms are integrated in the LTE radio access design. LTE therefore supports tight inter-working with all legacy 3GPP technologies and some non-3GPP technologies such as CDMA2000.

1.2.2.3 Terminal Complexity and Cost

A key consideration for competitive deployment of LTE is the availability of low-cost terminals with long battery life, both in stand-by and during activity. Therefore, low terminal complexity has been taken into account where relevant throughout the LTE system, as well as designing the system wherever possible to support low terminal power consumption.

1.2.2.4 Network Architecture Requirements

LTE is required to allow a cost-effective deployment by an improved radio access network architecture design including:

- Flat architecture consisting of just one type of node, the base station, known in LTE as the *eNodeB* (see Chapter 2);
- Effective protocols for the support of packet-switched services (see Chapters 3 to 4);
- Open interfaces and support of multivendor equipment interoperability;
- efficient mechanisms for operation and maintenance, including self-optimization functionalities (see Chapter 25);
- Support of easy deployment and configuration, for example for so-called home base stations (otherwise known as femto-cells) (see Chapter 24).

[9]Universal Terrestrial Radio Access Network.

1.3 Technologies for the Long Term Evolution

The fulfilment of the extensive range of requirements outlined above is only possible thanks to advances in the underlying mobile radio technology. As an overview, we outline here three fundamental technologies that have shaped the LTE radio interface design: *multicarrier* technology, *multiple-antenna* technology, and the application of *packet-switching* to the radio interface. Finally, we summarize the combinations of capabilities that are supported by different categories of LTE mobile terminal in Releases 8 and 9.

1.3.1 Multicarrier Technology

Adopting a multicarrier approach for multiple access in LTE was the first major design choice. After initial consolidation of proposals, the candidate schemes for the downlink were Orthogonal Frequency-Division Multiple Access (OFDMA)[10] and Multiple WCDMA, while the candidate schemes for the uplink were Single-Carrier Frequency-Division Multiple Access (SC-FDMA), OFDMA and Multiple WCDMA. The choice of multiple-access schemes was made in December 2005, with OFDMA being selected for the downlink, and SC-FDMA for the uplink. Both of these schemes open up the frequency domain as a new dimension of flexibility in the system, as illustrated schematically in Figure 1.4.

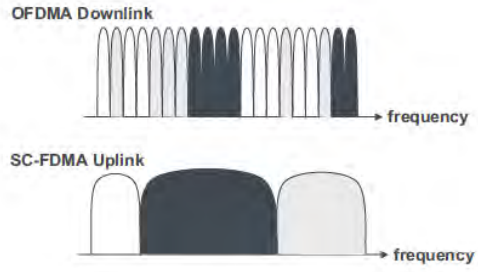


Figure 1.4: Frequency-domain view of the LTE multiple-access technologies.

OFDMA extends the multicarrier technology of OFDM to provide a very flexible multiple-access scheme. OFDM subdivides the bandwidth available for signal transmission into a multitude of narrowband subcarriers, arranged to be mutually orthogonal, which either individually or in groups can carry independent information streams; in OFDMA, this subdivision of the available bandwidth is exploited in sharing the subcarriers among multiple users.[11]

This resulting flexibility can be used in various ways:

[10] OFDM technology was already well understood in 3GPP as a result of an earlier study of the technology in 2003–4.

[11] The use of the frequency domain comes in addition to the well-known time-division multiplexing which continues to play an important role in LTE.

- Different spectrum bandwidths can be utilized without changing the fundamental system parameters or equipment design;
- Transmission resources of variable bandwidth can be allocated to different users and scheduled freely in the frequency domain;
- Fractional frequency re-use and interference coordination between cells are facilitated.

Extensive experience with OFDM has been gained in recent years from deployment of digital audio and video broadcasting systems such as DAB, DVB and DMB.[12] This experience has highlighted some of the key advantages of OFDM, which include:

- Robustness to time-dispersive radio channels, thanks to the subdivision of the wide-band transmitted signal into multiple narrowband subcarriers, enabling inter-symbol interference to be largely constrained within a guard interval at the beginning of each symbol;
- Low-complexity receivers, by exploiting frequency-domain equalization;
- Simple combining of signals from multiple transmitters in broadcast networks.

These advantages, and how they arise from the OFDM signal design, are explained in detail in Chapter 5.

By contrast, the transmitter design for OFDM is more costly, as the Peak-to-Average Power Ratio (PAPR) of an OFDM signal is relatively high, resulting in a need for a highly-linear RF power amplifier. However, this limitation is not inconsistent with the use of OFDM for *downlink* transmissions, as low-cost implementation has a lower priority for the base station than for the mobile terminal.

In the uplink, however, the high PAPR of OFDM is difficult to tolerate for the transmitter of the mobile terminal, since it is necessary to compromise between the output power required for good outdoor coverage, the power consumption, and the cost of the power amplifier. SC-FDMA, which is explained in detail in Chapter 14, provides a multiple-access technology which has much in common with OFDMA – in particular the flexibility in the frequency domain, and the incorporation of a guard interval at the start of each transmitted symbol to facilitate low-complexity frequency-domain equalization at the receiver. At the same time, SC-FDMA has a significantly lower PAPR. It therefore resolves to some extent the dilemma of how the uplink can benefit from the advantages of multicarrier technology while avoiding excessive cost for the mobile terminal transmitter and retaining a reasonable degree of commonality between uplink and downlink technologies.

In Release 10, the uplink multiple access scheme is extended to allow multiple clusters of subcarriers in the frequency domain, as explained in Section 28.3.6.

1.3.2 Multiple Antenna Technology

The use of multiple antenna technology allows the exploitation of the spatial-domain as another new dimension. This becomes essential in the quest for higher spectral efficiencies. As will be detailed in Chapter 11, with the use of multiple antennas the theoretically achievable spectral efficiency scales linearly with the minimum of the number of transmit and receive antennas employed, at least in suitable radio propagation environments.

[12] Digital Audio Broadcasting, Digital Video Broadcasting and Digital Mobile Broadcasting.

Multiple antenna technology opens the door to a large variety of features, but not all of them easily deliver their theoretical promises when it comes to implementation in practical systems. Multiple antennas can be used in a variety of ways, mainly based on three fundamental principles, schematically illustrated in Figure 1.5:

- **Diversity gain.** Use of the spatial diversity provided by the multiple antennas to improve the robustness of the transmission against multipath fading.
- **Array gain.** Concentration of energy in one or more given directions via precoding or beamforming. This also allows multiple users located in different directions to be served simultaneously (so-called multi-user MIMO).
- **Spatial multiplexing gain.** Transmission of multiple signal streams to a single user on multiple spatial layers created by combinations of the available antennas.

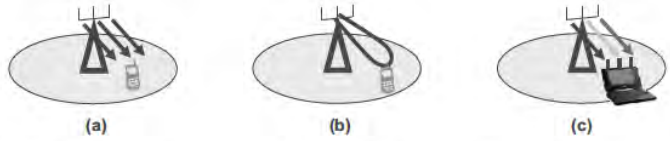


Figure 1.5: Three fundamental benefits of multiple antennas:
(a) diversity gain; (b) array gain; (c) spatial multiplexing gain.

A large part of the LTE Study Item phase was therefore dedicated to the selection and design of the various multiple antenna features to be included in the first release of LTE. The final system includes several complementary options which allow for adaptability according to the network deployment and the propagation conditions of the different users.

1.3.3 Packet-Switched Radio Interface

As has already been noted, LTE has been designed as a completely packet-oriented multi-service system, without the reliance on circuit-switched connection-oriented protocols prevalent in its predecessors. In LTE, this philosophy is applied across all the layers of the protocol stack.

The route towards fast packet scheduling over the radio interface was already opened by HSDPA, which allowed the transmission of short packets having a duration of the same order of magnitude as the coherence time of the fast fading channel, as shown in Figure 1.6. This calls for a joint optimization of the physical layer configuration and the resource management carried out by the link layer protocols according to the prevailing propagation conditions. This aspect of HSDPA involves tight coupling between the lower two layers of the protocol stack – the MAC (Medium Access Control layer – see Chapter 4) and the physical layer.

In HSDPA, this coupling already included features such as fast channel state feedback, dynamic link adaptation, scheduling exploiting multi-user diversity, and fast retransmission protocols. In LTE, in order to improve the system latency, the packet duration was further reduced from the 2 ms used in HSDPA down to just 1 ms. This short transmission interval,

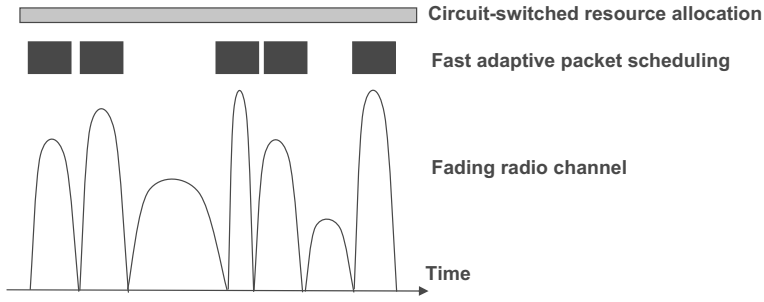


Figure 1.6: Fast scheduling and link adaptation.

together with the new dimensions of frequency and space, has further extended the field of cross-layer techniques between the MAC and physical layers to include the following techniques in LTE:

- Adaptive scheduling in both the frequency and spatial dimensions;
- Adaptation of the MIMO configuration including the selection of the number of spatial layers transmitted simultaneously;
- Link adaptation of modulation and code-rate, including the number of transmitted codewords;
- Several modes of fast channel state reporting.

These different levels of optimization are combined with very sophisticated control signalling.

1.3.4 User Equipment Categories

In practice it is important to recognize that the market for UEs is large and diverse, and there is therefore a need for LTE to support a range of categories of UE with different capabilities to satisfy different market segments. In general, each market segment attaches different priorities to aspects such as peak data rate, UE size, cost and battery life. Some typical trade-offs include the following:

- Support for the highest data rates is key to the success of some applications, but generally requires large amounts of memory for data processing, which increases the cost of the UE.
- UEs which may be embedded in large devices such as laptop computers are often not significantly constrained in terms of acceptable power consumption or the number of antennas which may be used; on the other hand, other market segments require ultra-slim hand-held terminals which have little space for multiple antennas or large batteries.

The wider the range of UE categories supported, the closer the match which may be made between a UE's supported functionality and the requirements of a particular market segment.

However, support for a large number of UE categories also has drawbacks in terms of the signalling overhead required for each UE to inform the network about its supported functionality, as well as increased costs due to loss of economies of scale and increased complexity for testing the interoperability of many different configurations.

The first release of LTE was therefore designed to support a compact set of five categories of UE, ranging from relatively low-cost terminals with similar capabilities to UMTS HSPA, up to very high-capability terminals which exploit the LTE technology to the maximum extent possible.

The five Release 8 UE categories are summarized in Table 1.2. It can be seen that the highest category of Release 8 LTE UE possesses peak data rate capabilities far exceeding the LTE Release 8 targets. Full details are specified in [12].

Table 1.2: Categories of LTE user equipment in Releases 8 and 9.

	UE category				
	1	2	3	4	5
Supported downlink data rate (Mbps)	10	50	100	150	300
Supported uplink data rate (Mbps)	5	25	50	50	75
Number of receive antennas required	2	2	2	2	4
Number of downlink MIMO layers supported	1	2	2	2	4
Support for 64QAM modulation in downlink	✔	✔	✔	✔	✔
Support for 64QAM modulation in uplink	✖	✖	✖	✖	✔
Relative memory requirement for physical layer processing (normalized to category 1 level)	1	4.9	4.9	7.3	14.6

Additional UE categories are introduced in Release 10, and these are explained in Section 27.5.

The LTE specifications deliberately avoid large numbers of optional features for the UEs, preferring to take the approach that if a feature is sufficiently useful to be worth including in the specifications then support of it should be mandatory. Nevertheless, a very small number of optional Release 8 features, whose support is indicated by each UE by specific signalling, are listed in [12]; such features are known as 'UE capabilities'. Some additional UE capabilities are added in later releases.

In addition, it is recognized that it is not always possible to complete conformance testing and Inter-Operability Testing (IOT) of every mandatory feature simultaneously for early deployments of LTE. Therefore, the development of conformance test cases for LTE was prioritized according to the likelihood of early deployment of each feature. Correspondingly, *Feature Group Indicators* (FGIs) are used for certain groups of lower priority mandatory features, to enable a UE to indicate whether IOT has been successfully completed for those features; the grouping of features corresponding to each FGI can be found in Annex B.1 of [13]. For UEs of Release 9 and later, it becomes mandatory for certain of these FGIs to be set to indicate that the corresponding feature(s) have been implemented and successfully tested.

1.3.5 From the First LTE Release to LTE-Advanced

As a result of intense activity by a larger number of contributing companies than ever before in 3GPP, the specifications for the first LTE release (Release 8) had reached a sufficient level of completeness by December 2007 to enable LTE to be submitted to ITU-R as a member of the IMT family of radio access technologies. It is therefore able to be deployed in IMT-designated spectrum, and the first commercial deployments were launched towards the end of 2009 in northern Europe.

Meanwhile, 3GPP has continued to improve the LTE system and to develop it to address new markets. In this section, we outline the new features introduced in the second LTE release, Release 9, and those provided by LTE Release 10, which begins the next significant step known as LTE-Advanced.

Increasing LTE's suitability for different markets and deployments was the first goal of Release 9. One important market with specific regulatory requirements is North America. LTE Release 9 therefore provides improved support for Public Warning Systems (PWS) and some accurate positioning methods (see Chapter 19). One positioning method uses the Observed Time Difference of Arrival (OTDOA) principle, supported by specially designed new reference signals inserted in the LTE downlink transmissions. Measurements of these positioning reference signals received from different base stations allow a UE to calculate its position very accurately, even in locations where other positioning means such as GPS fail (e.g. indoors). Enhanced Cell-ID-based techniques are also supported.

Release 9 also introduces support for a broadcast mode based on Single Frequency Network type transmissions (see Chapter 13).

The MIMO transmission modes are further developed in Release 9, with an extension of the Release 8 beamforming mode to support two orthogonal spatial layers that can be transmitted to a single user or multiple users, as described in Section 11.2.2.3. The design of this mode is forward-compatible for extension to more than two spatial layers in Release 10.

Release 9 also addresses specific deployments and, in particular, low power nodes (see Chapter 24). It defines new requirements for pico base stations and home base stations, in addition to improving support for Closed Subscriber Groups (CSG). Support for self-optimization of the networks is also enhanced in Release 9, as described in Chapter 25.

1.3.5.1 LTE-Advanced

The next version of LTE, Release 10, develops LTE to LTE-Advanced. While LTE Releases 8 and 9 already satisfy to a large extent the requirements set by ITU-R for the IMT-Advanced designation [14] (see Section 27.1), Release 10 will fully satisfy them and even exceed them in several aspects where 3GPP has set more demanding performance targets than those of ITU-R. The requirements for LTE-Advanced are discussed in detail in Chapter 27.

The main Release 10 features that are directly related to fulfilment of the IMT-Advanced requirements are:

- Carrier aggregation, allowing the total transmission bandwidth to be increased up to 100 MHz (see Chapter 28);
- Uplink MIMO transmission for peak spectral efficiencies greater than 7.5 bps/Hz and targeting up to 15 bps/Hz (see Chapter 29);

- Downlink MIMO enhancements, targeting peak spectral efficiencies up to 30 bps/Hz (see Chapter 29).

Besides addressing the IMT-Advanced requirements, Release 10 also provides some new features to enhance LTE deployment, such as support for relaying (see Chapter 30), enhanced inter-cell interference coordination (see Chapter 31) and mechanisms to minimize the need for drive tests by supporting extended measurement reports from the terminals (see Chapters 25 and 31).

1.4 From Theory to Practice

With commercial deployment of LTE now a reality, the advances in theoretical understanding and technology which underpin the LTE specifications are being exploited practically. This book is written with the primary aim of illuminating the transition from the underlying academic progress to the realization of useful advances in the provision of mobile communication services. Particular focus is given to the physical layer of the Radio Access Network (RAN), as it is here that many of the most dramatic technical advances are manifested. This should enable the reader to develop an understanding of the background to the technology choices in the LTE system, and hence to understand better the LTE specifications and how they may be implemented.

Parts I to IV of the book describe the features of LTE Releases 8 and 9, including indications of the aspects that are further enhanced in Release 10, while the details of the major new features of Release 10 are explained in Part V.

Part I sets the radio interface in the context of the network architecture and protocols, including radio resource management aspects, as well as explaining the new developments in these areas which distinguish LTE from previous systems.

In Part II, the physical layer of the RAN downlink is covered in detail, beginning with an explanation of the theory of the new downlink multiple access technology, OFDMA, in Chapter 5. This sets the context for the details of the LTE downlink design in Chapters 6 to 9. As coding, link adaptation and multiple antenna operation are of fundamental importance in fulfilling the LTE requirements, two chapters are then devoted to these topics, covering both the theory and the practical implementation in LTE.

Chapter 12 shows how these techniques can be applied to the system-level operation of the LTE system, focusing on applying the new degrees of freedom to multi-user scheduling and interference coordination.

Finally for the downlink, Chapter 13 covers broadcast operation – a mode which has its own unique challenges in a cellular system but which is nonetheless important in enabling a range of services to be provided to the end user.

Part III addresses the physical layer of the RAN uplink, beginning in Chapter 14 with an introduction to the theory behind the new uplink multiple access technology, SC-FDMA. This is followed in Chapters 15 to 18 with an analysis of the detailed uplink structure and operation, including the design of the associated procedures for random access, timing control and power control which are essential to the efficient operation of the uplink.

This leads on to Part IV, which examines a number of aspects of LTE related to its deployment as a mobile cellular system. Chapter 19 explains the UE positioning techniques introduced in Release 9. Chapter 20 provides a thorough analysis of the characteristics

of the radio propagation environments in which LTE systems will be deployed, since an understanding of the propagation environment underpins much of the technology adopted for the LTE specifications. The new technologies and bandwidths adopted in LTE also have implications for the radio-frequency implementation of the mobile terminals in particular, and some of these are analysed in Chapter 21. The LTE system is designed to operate not just in wide bandwidths but also in a diverse range of spectrum allocation scenarios, and Chapter 23 therefore addresses the different duplex modes applicable to LTE and the effects that these may have on system design and operation. Chapter 24 addresses aspects of special relevance to deployment of low-power base stations such as Home eNodeBs and picocells, while Chapter 25 explains the advanced techniques for self-optimization of the network. Part IV concludes with a dedicated chapter examining a wide range of aspects of the overall system performance achievable with the first release of LTE.

Finally, Part V explains in detail the major new features included in Release 10 for LTE-Advanced, as 3GPP continues to respond to the ever-higher expectations of end-users. Chapters 28 to 30 address the technologies of carrier aggregation, enhanced MIMO and relaying respectively, and Chapter 31 covers enhanced Inter-Cell Interference Coordination, Minimization of Drive Tests and Machine-Type Communications. Chapter 32 provides an evaluation of the system performance achievable with LTE-Advanced Release 10, and concludes with a further look into the future.

References[13]

[1] ITU, International Telecommunications Union, www.itu.int/itu-r.

[2] NGMN, 'Next Generation Mobile Networks Beyond HSPA & EVDO – A white paper', www.ngmn.org, December 2006.

[3] J. H. Winters, 'Optimum Combining in Digital Mobile Radio with Cochannel Interference'. *IEEE Journal on Selected Areas in Communications*, Vol. 2, July 1984.

[4] R. Price and P. E. Green, 'A Communication Technique for Multipath Channels' in *Proceedings of the IRE*, Vol. 46, March 1958.

[5] Orange, China Mobile, KPN, NTT DoCoMo, Sprint, T-Mobile, Vodafone, and Telecom Italia, 'R1-070674: LTE Physical Layer Framework for Performance Verification', www.3gpp.org, 3GPP TSG RAN WG1, meeting 48, St Louis, USA, February 2007.

[6] 3GPP Technical Report 25.814, 'Physical Layer Aspects for Evolved UTRA', www.3gpp.org.

[7] NGMN, 'Next Generation Mobile Networks Radio Access Performance Evaluation Methodology', www.ngmn.org, June 2007.

[8] Ericsson, 'R1-072578: Summary of Downlink Performance Evaluation', www.3gpp.org, 3GPP TSG RAN WG1, meeting 49, Kobe, Japan, May 2007.

[9] Nokia, 'R1-072261: LTE Performance Evaluation – Uplink Summary', www.3gpp.org, 3GPP TSG RAN WG1, meeting 49, Kobe, Japan, May 2007.

[10] 3GPP Technical Report 25.913, 'Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)', www.3gpp.org.

[11] 3GPP Technical Specification 36.307, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements on User Equipments (UEs) supporting a release-independent frequency band', www.3gpp.org.

[13] All web sites confirmed 1st March 2011.

[12] 3GPP Technical Specification 36.306, 'Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio access capabilities', www.3gpp.org.

[13] 3GPP Technical Specification 36.331, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification', www.3gpp.org.

[14] ITU-R Report M.2134, 'Requirements related to technical performance for IMT-Advanced radio interface(s)', www.itu.int/itu-r.

Part I

Network Architecture and Protocols

2

Network Architecture

Sudeep Palat and Philippe Godin

2.1 Introduction

As mentioned in the preceding chapter, LTE has been designed to support only Packet-Switched (PS) services, in contrast to the Circuit-Switched (CS) model of previous cellular systems. It aims to provide seamless Internet Protocol (IP) connectivity between User Equipment (UE) and the Packet Data Network (PDN), without any disruption to the end users' applications during mobility. While the term 'LTE' encompasses the evolution of the radio access through the Evolved-UTRAN[1] (E-UTRAN), it is accompanied by an evolution of the non-radio aspects under the term 'System Architecture Evolution' (SAE) which includes the Evolved Packet Core (EPC) network. Together LTE and SAE comprise the Evolved Packet System (EPS).

EPS uses the concept of *EPS bearers* to route IP traffic from a gateway in the PDN to the UE. A bearer is an IP packet flow with a defined Quality of Service (QoS). The E-UTRAN and EPC together set up and release bearers as required by applications. EPS natively supports voice services over the IP Multimedia Subsystem (IMS) using Voice over IP (VoIP), but LTE also supports interworking with legacy systems for traditional CS voice support.

This chapter presents the overall EPS network architecture, giving an overview of the functions provided by the Core Network (CN) and E-UTRAN. The protocol stack across the different interfaces is then explained, along with an overview of the functions provided by the different protocol layers. Section 2.4 outlines the end-to-end bearer path including QoS aspects, provides details of a typical procedure for establishing a bearer and discusses the inter-working with legacy systems for CS voice services. The remainder of the chapter presents the network interfaces in detail, with particular focus on the E-UTRAN interfaces

[1] Universal Terrestrial Radio Access Network.

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

and associated procedures, including those for the support of user mobility. The network elements and interfaces used solely to support broadcast services are covered in Chapter 13, and the aspects related to UE positioning in Chapter 19.

2.2 Overall Architectural Overview

EPS provides the user with IP connectivity to a PDN for accessing the Internet, as well as for running services such as VoIP. An EPS bearer is typically associated with a QoS. Multiple bearers can be established for a user in order to provide different QoS streams or connectivity to different PDNs. For example, a user might be engaged in a voice (VoIP) call while at the same time performing web browsing or File Transfer Protocol (FTP) download. A VoIP bearer would provide the necessary QoS for the voice call, while a best-effort bearer would be suitable for the web browsing or FTP session. The network must also provide sufficient security and privacy for the user and protection for the network against fraudulent use.

Release 9 of LTE introduced several additional features. To meet regulatory requirements for commercial voice, services such as support of IMS, emergency calls and UE positioning (see Chapter 19) were introduced. Enhancements to Home cells (HeNBs) were also introduced in Release 9 (see Chapter 24).

All these features are supported by means of several EPS network elements with different roles. Figure 2.1 shows the overall network architecture including the network elements and the standardized interfaces. At a high level, the network is comprised of the CN (i.e. EPC) and the access network (i.e. E-UTRAN). While the CN consists of many logical nodes, the access network is made up of essentially just one node, the evolved NodeB (eNodeB), which connects to the UEs. Each of these network elements is inter-connected by means of interfaces which are standardized in order to allow multivendor interoperability.

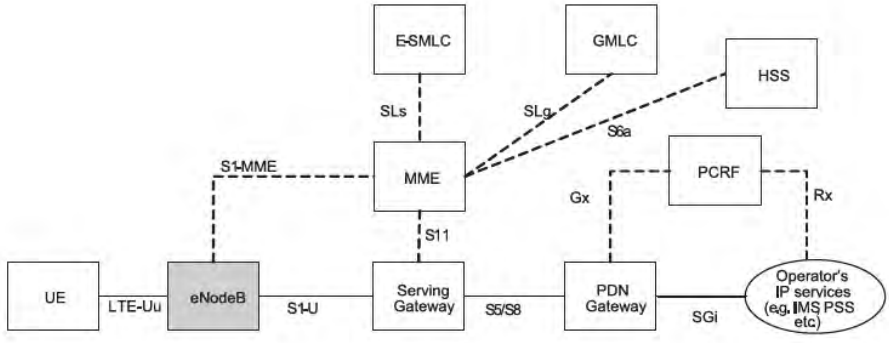


Figure 2.1: The EPS network elements.

The functional split between the EPC and E-UTRAN is shown in Figure 2.2. The EPC and E-UTRAN network elements are described in more detail below.

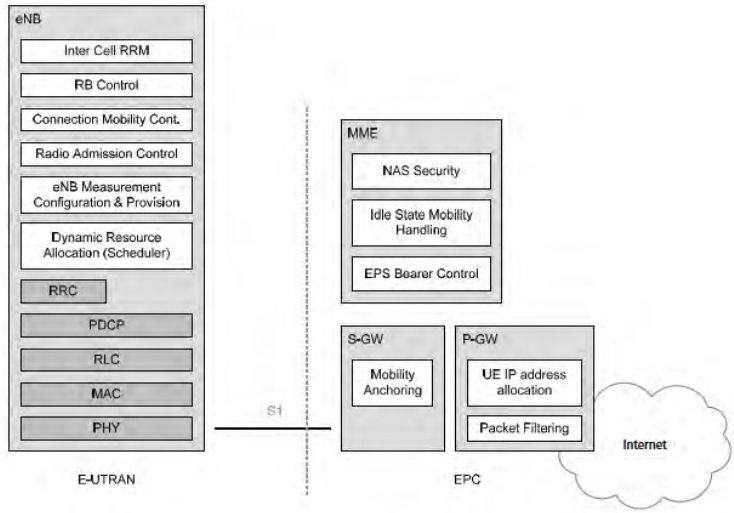


Figure 2.2: Functional split between E-UTRAN and EPC.
Reproduced by permission of © 3GPP.

2.2.1 The Core Network

The CN (called the EPC in SAE) is responsible for the overall control of the UE and the establishment of the bearers. The main logical nodes of the EPC are:

- PDN Gateway (P-GW);
- Serving GateWay (S-GW);
- Mobility Management Entity (MME) ;
- Evolved Serving Mobile Location Centre (E-SMLC).

In addition to these nodes, the EPC also includes other logical nodes and functions such as the Gateway Mobile Location Centre (GMLC), the Home Subscriber Server (HSS) and the Policy Control and Charging Rules Function (PCRF). Since the EPS only provides a bearer path of a certain QoS, control of multimedia applications such as VoIP is provided by the IMS which is considered to be outside the EPS itself. When a user is roaming outside his home country network, the user's P-GW, GMLC and IMS domain may be located in either the home network or the visited network. The logical CN nodes (specified in [1]) are shown in Figure 2.1 and discussed in more detail below.

- **PCRF.** The PCRF is responsible for policy control decision-making, as well as for controlling the flow-based charging functionalities in the Policy Control Enforcement Function (PCEF) which resides in the P-GW. The PCRF provides the QoS authorization (QoS class identifier and bit rates) that decides how a certain data flow will be treated in the PCEF and ensures that this is in accordance with the user's subscription profile.

- **GMLC.** The GMLC contains functionalities required to support LoCation Services (LCS). After performing authorization, it sends positioning requests to the MME and receives the final location estimates.

- **Home Subscriber Server (HSS).** The HSS contains users' SAE subscription data such as the EPS-subscribed QoS profile and any access restrictions for roaming (see Section 2.2.3). It also holds information about the PDNs to which the user can connect. This could be in the form of an Access Point Name (APN) (which is a label according to DNS[2] naming conventions describing the access point to the PDN), or a PDN Address (indicating subscribed IP address(es)). In addition, the HSS holds dynamic information such as the identity of the MME to which the user is currently attached or registered. The HSS may also integrate the Authentication Centre (AuC) which generates the vectors for authentication and security keys (see Section 3.2.3.1).

- **P-GW.** The P-GW is responsible for IP address allocation for the UE, as well as QoS enforcement and flow-based charging according to rules from the PCRF. The P-GW is responsible for the filtering of downlink user IP packets into the different QoS-based bearers. This is performed based on Traffic Flow Templates (TFTs) (see Section 2.4). The P-GW performs QoS enforcement for Guaranteed Bit Rate (GBR) bearers. It also serves as the mobility anchor for inter-working with non-3GPP technologies such as CDMA2000 and WiMAX networks (see Section 2.4.2 and Chapter 22 for more information about mobility).

- **S-GW.** All user IP packets are transferred through the S-GW, which serves as the local mobility anchor for the data bearers when the UE moves between eNodeBs. It also retains the information about the bearers when the UE is in idle state (known as EPS Connection Management IDLE (ECM-IDLE), see Section 2.2.1.1) and temporarily buffers downlink data while the MME initiates paging of the UE to re-establish the bearers. In addition, the S-GW performs some administrative functions in the visited network, such as collecting information for charging (e.g. the volume of data sent to or received from the user) and legal interception. It also serves as the mobility anchor for inter-working with other 3GPP technologies such as GPRS[3] and UMTS[4] (see Section 2.4.2 and Chapter 22 for more information about mobility).

- **MME.** The MME is the control node which processes the signalling between the UE and the CN. The protocols running between the UE and the CN are known as the *Non-Access Stratum* (NAS) protocols.

  The main functions supported by the MME are classified as:

  Functions related to bearer management. This includes the establishment, maintenance and release of the bearers, and is handled by the session management layer in the NAS protocol.

  Functions related to connection management. This includes the establishment of the connection and security between the network and UE, and is handled by the connection or mobility management layer in the NAS protocol layer.

[2]Domain Name System.
[3]General Packet Radio Service.
[4]Universal Mobile Telecommunications System.

NAS control procedures are specified in [1] and are discussed in more detail in the following section.

Functions related to inter-working with other networks. This includes handing over of voice calls to legacy networks and is explained in more detail in Section 2.4.2.

- **E-SMLC.** The E-SMLC manages the overall coordination and scheduling of resources required to find the location of a UE that is attached to E-UTRAN. It also calculates the final location based on the estimates it receives, and it estimates the UE speed and the achieved accuracy. The positioning functions and protocols are explained in detail in Chapter 19.

2.2.1.1 Non-Access Stratum (NAS) Procedures

The NAS procedures, especially the connection management procedures, are fundamentally similar to UMTS. The main change from UMTS is that EPS allows concatenation of some procedures so as to enable faster establishment of the connection and the bearers.

The MME creates a *UE context*when a UE is turned on and attaches to the network. It assigns to the UE a unique short temporary identity termed the SAE-Temporary Mobile Subscriber Identity (S-TMSI) which identifies the UE context in the MME. This UE context holds user subscription information downloaded from the HSS. The local storage of subscription data in the MME allows faster execution of procedures such as bearer establishment since it removes the need to consult the HSS every time. In addition, the UE context also holds dynamic information such as the list of bearers that are established and the terminal capabilities.

To reduce the overhead in the E-UTRAN and the processing in the UE, all UE-related information in the access network can be released during long periods of data inactivity. The UE is then in the ECM-IDLE state. The MME retains the UE context and the information about the established bearers during these idle periods.

To allow the network to contact an ECM-IDLE UE, the UE updates the network as to its new location whenever it moves out of its current Tracking Area (TA); this procedure is called a 'Tracking Area Update'. The MME is responsible for keeping track of the user location while the UE is in ECM-IDLE.

When there is a need to deliver downlink data to an ECM-IDLE UE, the MME sends a paging message to all the eNodeBs in its current TA, and the eNodeBs page the UE over the radio interface. On receipt of a paging message, the UE performs a service request procedure which results in moving the UE to the ECM-CONNECTED state. UE-related information is thereby created in the E-UTRAN, and the bearers are re-established. The MME is responsible for the re-establishment of the radio bearers and updating the UE context in the eNodeB. This transition between the UE states is called an 'idle-to-active transition'. To speed up the idle-to-active transition and bearer establishment, EPS supports concatenation of the NAS and AS[5] procedures for bearer activation (see also Section 2.4.1). Some inter-relationship between the NAS and AS protocols is intentionally used to allow procedures to run simultaneously, rather than sequentially as in UMTS. For example, the bearer establishment procedure can be executed by the network without waiting for the completion of the security procedure.

[5] Access Stratum – the protocols which run between the eNodeBs and the UE.

Security functions are the responsibility of the MME for both signalling and user data. When a UE attaches with the network, a mutual authentication of the UE and the network is performed between the UE and the MME/HSS. This authentication function also establishes the security keys which are used for encryption of the bearers, as explained in Section 3.2.3.1. The security architecture for SAE is specified in [2].

The NAS also handles IMS Emergency calls, whereby UEs without regular access to the network (i.e. terminals without a Universal Subscriber Identity Module (USIM) or UEs in limited service mode) are allowed access to the network using an 'Emergency Attach' procedure; this bypasses the security requirements but only allows access to an emergency P-GW.

2.2.2 The Access Network

The access network of LTE, E-UTRAN, simply consists of a network of eNodeBs, as illustrated in Figure 2.3. For normal user traffic (as opposed to broadcast), there is no centralized controller in E-UTRAN; hence the E-UTRAN architecture is said to be flat.

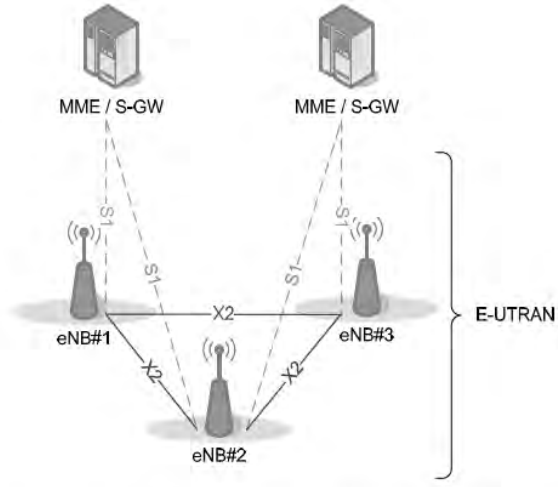


Figure 2.3: Overall E-UTRAN architecture. Reproduced by permission of © 3GPP.

The eNodeBs are normally inter-connected with each other by means of an interface known as *X2*, and to the EPC by means of the *S1* interface – more specifically, to the MME by means of the S1-MME interface and to the S-GW by means of the S1-U interface.

The protocols which run between the eNodeBs and the UE are known as the *Access Stratum* (AS) protocols.

The E-UTRAN is responsible for all radio-related functions, which can be summarized briefly as:

- **Radio Resource Management.** This covers all functions related to the radio bearers, such as radio bearer control, radio admission control, radio mobility control, scheduling and dynamic allocation of resources to UEs in both uplink and downlink.
- **Header Compression.** This helps to ensure efficient use of the radio interface by compressing the IP packet headers which could otherwise represent a significant overhead, especially for small packets such as VoIP (see Section 4.2.2).
- **Security.** All data sent over the radio interface is encrypted (see Sections 3.2.3.1 and 4.2.3).
- **Positioning.** The E-UTRAN provides the necessary measurements and other data to the E-SMLC and assists the E-SMLC in finding the UE position (see Chapter 19).
- **Connectivity to the EPC.** This consists of the signalling towards the MME and the bearer path towards the S-GW.

On the network side, all of these functions reside in the eNodeBs, each of which can be responsible for managing multiple cells. Unlike some of the previous second- and third-generation technologies, LTE integrates the radio controller function into the eNodeB. This allows tight interaction between the different protocol layers of the radio access network, thus reducing latency and improving efficiency. Such distributed control eliminates the need for a high-availability, processing-intensive controller, which in turn has the potential to reduce costs and avoid 'single points of failure'. Furthermore, as LTE does not support soft handover there is no need for a centralized data-combining function in the network.

One consequence of the lack of a centralized controller node is that, as the UE moves, the network must transfer all information related to a UE, i.e. the UE context, together with any buffered data, from one eNodeB to another. As discussed in Section 2.3.1.1, mechanisms are therefore needed to avoid data loss during handover. The operation of the X2 interface for this purpose is explained in more detail in Section 2.6.

An important feature of the S1 interface linking the access network to the CN is known as *S1-flex*. This is a concept whereby multiple CN nodes (MME/S-GWs) can serve a common geographical area, being connected by a mesh network to the set of eNodeBs in that area (see Section 2.5). An eNodeB may thus be served by multiple MME/S-GWs, as is the case for eNodeB#2 in Figure 2.3. The set of MME/S-GW nodes serving a common area is called an *MME/S-GW pool* , and the area covered by such a pool of MME/S-GWs is called a *pool area*. This concept allows UEs in the cell(s) controlled by one eNodeB to be shared between multiple CN nodes, thereby providing a possibility for load sharing and also eliminating single points of failure for the CN nodes. The UE context normally remains with the same MME as long as the UE is located within the pool area.

2.2.3 Roaming Architecture

A network run by one operator in one country is known as a Public Land Mobile Network (PLMN). Roaming, where users are allowed to connect to PLMNs other than those to which they are directly subscribed, is a powerful feature for mobile networks, and LTE/SAE is no exception. A roaming user is connected to the E-UTRAN, MME and S-GW of the visited LTE network. However, LTE/SAE allows the P-GW of either the visited or the home network to be used, as shown in Figure 2.4. Using the home network's P-GW allows the user to access

the home operator's services even while in a visited network. A P-GW in the visited network allows a 'local breakout' to the Internet in the visited network.

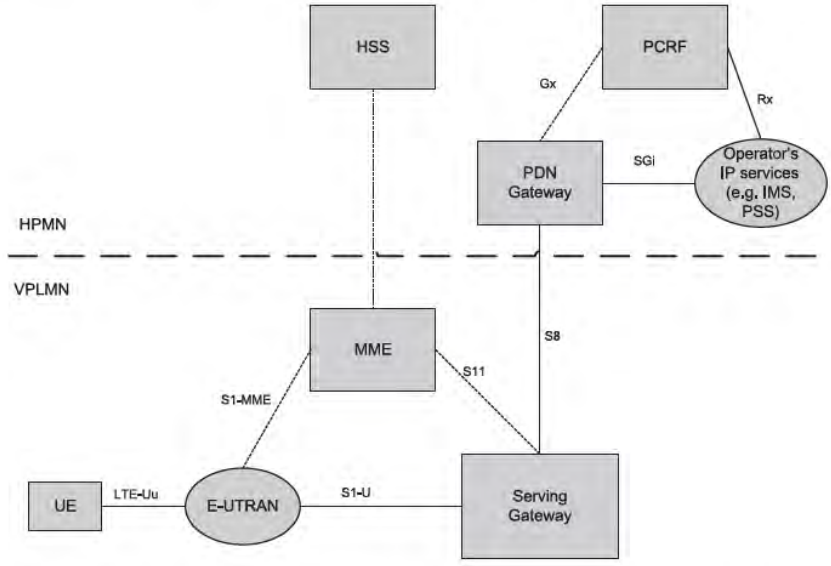


Figure 2.4: Roaming architecture for 3GPP accesses with P-GW in home network.

2.3 Protocol Architecture

We outline here the radio protocol architecture of E-UTRAN.

2.3.1 User Plane

An IP packet for a UE is encapsulated in an EPC-specific protocol and tunnelled between the P-GW and the eNodeB for transmission to the UE. Different tunnelling protocols are used across different interfaces. A 3GPP-specific tunnelling protocol called the GPRS Tunnelling Protocol (GTP) [4] is used over the core network interfaces, S1 and S5/S8.[6]

The E-UTRAN user plane protocol stack, shown greyed in Figure 2.5, consists of the Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC) and Medium Access Control (MAC) sublayers which are terminated in the eNodeB on the network side. The respective roles of each of these layers are explained in detail in Chapter 4.

[6]SAE also provides an option to use Proxy Mobile IP (PMIP) on S5/S8. More details on the MIP-based S5/S8 interface can be found in [3].

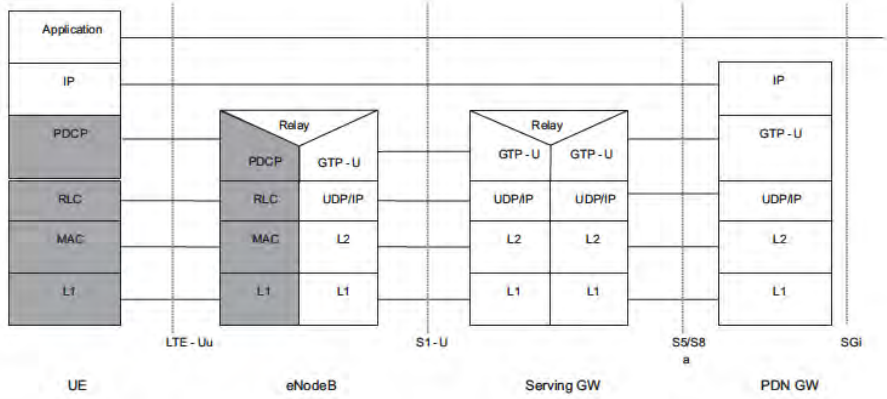


Figure 2.5: The E-UTRAN user plane protocol stack. Reproduced by permission of © 3GPP.

2.3.1.1 Data Handling During Handover

In the absence of any centralized controller node, data buffering during handover due to user mobility in the E-UTRAN must be performed in the eNodeB itself. Data protection during handover is a responsibility of the PDCP layer and is explained in detail in Section 4.2.4.

The RLC and MAC layers both start afresh in a new cell after handover is completed.

2.3.2 Control Plane

The protocol stack for the control plane between the UE and MME is shown in Figure 2.6.

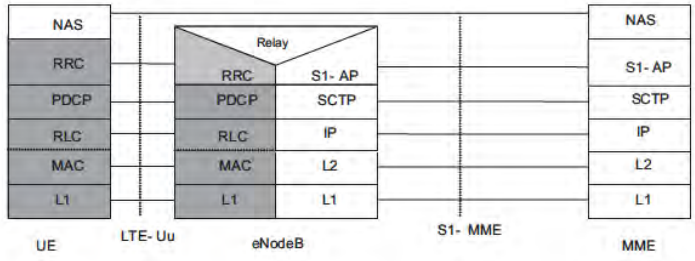


Figure 2.6: Control plane protocol stack. Reproduced by permission of © 3GPP.

The greyed region of the stack indicates the AS protocols. The lower layers perform the same functions as for the user plane with the exception that there is no header compression function for control plane.

The Radio Resource Control (RRC) protocol is known as 'Layer 3' in the AS protocol stack. It is the main controlling function in the AS, being responsible for establishing the radio bearers and configuring all the lower layers using RRC signalling between the eNodeB and the UE. These functions are detailed in Section 3.2.

2.4 Quality of Service and EPS Bearers

In a typical case, multiple applications may be running in a UE at the same time, each one having different QoS requirements. For example, a UE can be engaged in a VoIP call while at the same time browsing a web page or downloading an FTP file. VoIP has more stringent requirements for QoS in terms of delay and delay jitter than web browsing and FTP, while the latter requires a much lower packet loss rate. In order to support multiple QoS requirements, different bearers are set up within EPS, each being associated with a QoS.

Broadly, bearers can be classified into two categories based on the nature of the QoS they provide:

- **Minimum Guaranteed Bit Rate (GBR) bearers** which can be used for applications such as VoIP. These have an associated GBR value for which dedicated transmission resources are permanently allocated (e.g. by an admission control function in the eNodeB) at bearer establishment/modification. Bit rates higher than the GBR may be allowed for a GBR bearer if resources are available. In such cases, a Maximum Bit Rate (MBR) parameter, which can also be associated with a GBR bearer, sets an upper limit on the bit rate which can be expected from a GBR bearer.

- **Non-GBR bearers** which do not guarantee any particular bit rate. These can be used for applications such as web browsing or FTP transfer. For these bearers, no bandwidth resources are allocated permanently to the bearer.

In the access network, it is the eNodeB's responsibility to ensure that the necessary QoS for a bearer over the radio interface is met. Each bearer has an associated Class Identifier (QCI), and an Allocation and Retention Priority (ARP).

Each QCI is characterized by priority, packet delay budget and acceptable packet loss rate. The QCI label for a bearer determines the way it is handled in the eNodeB. Only a dozen such QCIs have been standardized so that vendors can all have the same understanding of the underlying service characteristics and thus provide the corresponding treatment, including queue management, conditioning and policing strategy. This ensures that an LTE operator can expect uniform traffic handling behaviour throughout the network regardless of the manufacturers of the eNodeB equipment. The set of standardized QCIs and their characteristics (from which the PCRF in an EPS can select) is provided in Table 2.1 (from Section 6.1.7 in [5]).

The priority and packet delay budget (and, to some extent, the acceptable packet loss rate) from the QCI label determine the RLC mode configuration (see Section 4.3.1), and how the scheduler in the MAC (see Section 4.4.2.1) handles packets sent over the bearer (e.g. in terms of scheduling policy, queue management policy and rate shaping policy). For example, a packet with a higher priority can be expected to be scheduled before a packet with lower priority. For bearers with a low acceptable loss rate, an Acknowledged Mode (AM) can be

Table 2.1: Standardized QoS Class Identifiers (QCIs) for LTE.

QCI	Resource type	Priority	Packet delay budget (ms)	Packet error loss rate	Example services
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational video (live streaming)
3	GBR	5	300	10^{-6}	Non-conversational video (buffered streaming)
4	GBR	3	50	10^{-3}	Real time gaming
5	Non-GBR	1	100	10^{-6}	IMS signalling
6	Non-GBR	7	100	10^{-3}	Voice, video (live streaming), interactive gaming
7	Non-GBR	6	300	10^{-6}	Video (buffered streaming)
8	Non-GBR	8	300	10^{-6}	TCP-based (e.g. WWW, e-mail) chat, FTP, p2p file sharing, progressive video, etc.
9	Non-GBR	9	300	10^{-6}	

used within the RLC protocol layer to ensure that packets are delivered successfully across the radio interface (see Section 4.3.1.3).

The ARP of a bearer is used for call admission control – i.e. to decide whether or not the requested bearer should be established in case of radio congestion. It also governs the prioritization of the bearer for pre-emption with respect to a new bearer establishment request. Once successfully established, a bearer's ARP does not have any impact on the bearer-level packet forwarding treatment (e.g. for scheduling and rate control). Such packet forwarding treatment should be solely determined by the other bearer-level QoS parameters such as QCI, GBR and MBR.

An EPS bearer has to cross multiple interfaces as shown in Figure 2.7 – the S5/S8 interface from the P-GW to the S-GW, the S1 interface from the S-GW to the eNodeB, and the radio interface (also known as the LTE-Uu interface) from the eNodeB to the UE. Across each interface, the EPS bearer is mapped onto a lower layer bearer, each with its own bearer identity. Each node must keep track of the binding between the bearer IDs across its different interfaces.

An S5/S8 bearer transports the packets of an EPS bearer between a P-GW and an S-GW. The S-GW stores a one-to-one mapping between an S1 bearer and an S5/S8 bearer. The bearer is identified by the GTP tunnel ID across both interfaces.

An S1 bearer transports the packets of an EPS bearer between an S-GW and an eNodeB. A radio bearer [6] transports the packets of an EPS bearer between a UE and an eNodeB. An E-UTRAN Radio Access Bearer (E-RAB) refers to the concatenation of an S1 bearer and the corresponding radio bearer. An eNodeB stores a one-to-one mapping between a radio bearer ID and an S1 bearer to create the mapping between the two. The overall EPS bearer service architecture is shown in Figure 2.8.

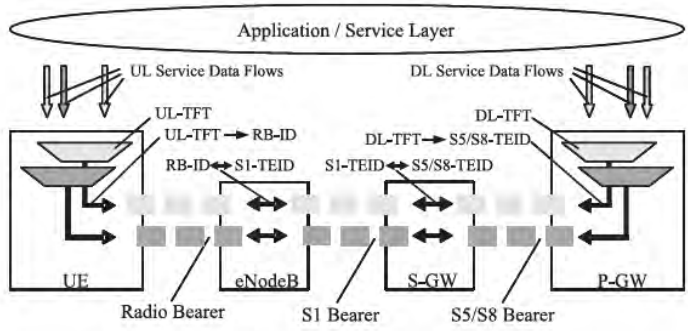


Figure 2.7: LTE/SAE bearers across the different interfaces. Reproduced by permission of © 3GPP.

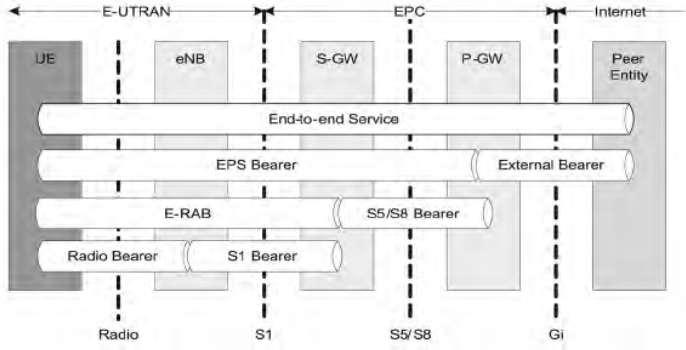


Figure 2.8: The overall EPS bearer service architecture. Reproduced by permission of © 3GPP.

IP packets mapped to the same EPS bearer receive the same bearer-level packet forwarding treatment (e.g. scheduling policy, queue management policy, rate shaping policy, RLC configuration). Providing different bearer-level QoS thus requires that a separate EPS bearer is established for each QoS flow, and user IP packets must be filtered into the different EPS bearers.

Packet filtering into different bearers is based on Traffic Flow Templates (TFTs). The TFTs use IP header information such as source and destination IP addresses and Transmission Control Protocol (TCP) port numbers to filter packets such as VoIP from web browsing traffic so that each can be sent down the respective bearers with appropriate QoS. An UpLink TFT (UL TFT) associated with each bearer in the UE filters IP packets to EPS bearers in the uplink direction. A DownLink TFT (DL TFT) in the P-GW is a similar set of downlink packet filters.

As part of the procedure by which a UE attaches to the network, the UE is assigned an IP address by the P-GW and at least one bearer is established, called the default bearer, and it remains established throughout the lifetime of the PDN connection in order to provide the UE with always-on IP connectivity to that PDN. The initial bearer-level QoS parameter values of the default bearer are assigned by the MME, based on subscription data retrieved from the HSS. The PCEF may change these values in interaction with the PCRF or according to local configuration. Additional bearers called dedicated bearers can also be established at any time during or after completion of the attach procedure. A dedicated bearer can be either GBR or non-GBR (the default bearer always has to be a non-GBR bearer since it is permanently established). The distinction between default and dedicated bearers should be transparent to the access network (e.g. E-UTRAN). Each bearer has an associated QoS, and if more than one bearer is established for a given UE, then each bearer must also be associated with appropriate TFTs. These dedicated bearers could be established by the network, based for example on a trigger from the IMS domain, or they could be requested by the UE. The dedicated bearers for a UE may be provided by one or more P-GWs.

The bearer-level QoS parameter values for dedicated bearers are received by the P-GW from the PCRF and forwarded to the S-GW. The MME only transparently forwards those values received from the S-GW over the S11 reference point to the E-UTRAN.

2.4.1 Bearer Establishment Procedure

This section describes an example of the end-to-end bearer establishment procedure across the network nodes using the functionality described in the previous sections.

A typical bearer establishment flow is shown in Figure 2.9. Each of the messages is described below.

When a bearer is established, the bearers across each of the interfaces discussed above are established.

The PCRF sends a 'PCC[7] Decision Provision' message indicating the required QoS for the bearer to the P-GW. The P-GW uses this QoS policy to assign the bearer-level QoS parameters. The P-GW then sends to the S-GW a 'Create Dedicated Bearer Request' message including the QoS and UL TFT to be used in the UE.

The S-GW forwards the Create Dedicated Bearer Request message (including bearer QoS, UL TFT and S1-bearer ID) to the MME (message 3 in Figure 2.9).

The MME then builds a set of session management configuration information including the UL TFT and the EPS bearer identity, and includes it in the 'Bearer Setup Request' message which it sends to the eNodeB (message 4 in Figure 2.9). The session management configuration is NAS information and is therefore sent transparently by the eNodeB to the UE.

The Bearer Setup Request also provides the QoS of the bearer to the eNodeB; this information is used by the eNodeB for call admission control and also to ensure the necessary QoS by appropriate scheduling of the user's IP packets. The eNodeB maps the EPS bearer QoS to the radio bearer QoS. It then signals a 'RRC Connection Reconfiguration' message (including the radio bearer QoS, session management configuration and EPS radio bearer identity) to the UE to set up the radio bearer (message 5 in Figure 2.9). The RRC Connection Reconfiguration message contains all the configuration parameters for the radio interface.

[7]Policy Control and Charging.

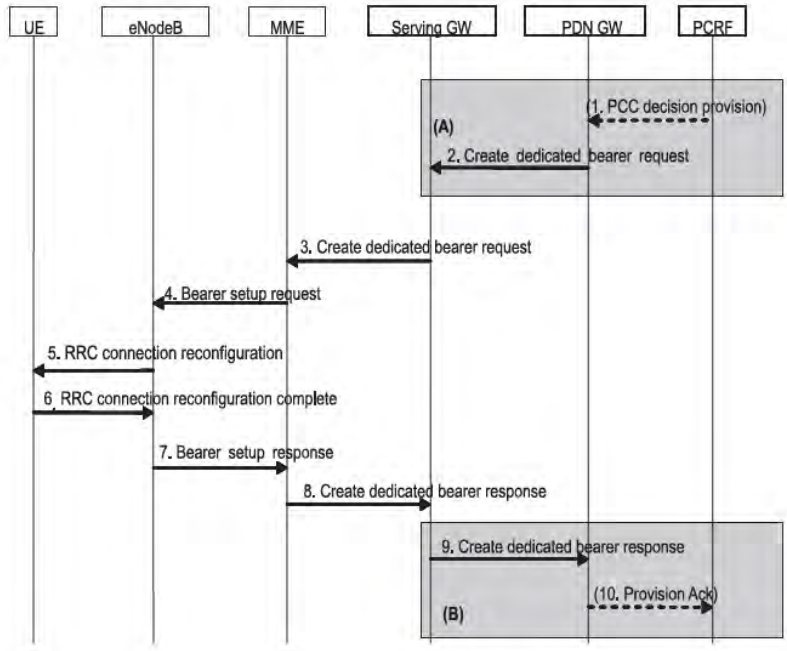


Figure 2.9: An example message flow for an LTE/SAE bearer establishment. Reproduced by permission of © 3GPP.

This is mainly for the configuration of the Layer 2 (PDCP, RLC and MAC) parameters, but also the Layer 1 parameters required for the UE to initialize the protocol stack.

Messages 6 to 10 are the corresponding response messages to confirm that the bearers have been set up correctly.

2.4.2 Inter-Working with other RATs

EPS also supports inter-working and mobility (handover) with networks using other Radio Access Technologies (RATs), notably GSM[8], UMTS, CDMA2000 and WiMAX. The architecture for inter-working with 2G and 3G GPRS/UMTS networks is shown in Figure 2.10. The S-GW acts as the mobility anchor for inter-working with other 3GPP technologies such as GSM and UMTS, while the P-GW serves as an anchor allowing seamless mobility to non-3GPP networks such as CDMA2000 or WiMAX. The P-GW may also support a Proxy Mobile Internet Protocol (PMIP) based interface. While VoIP is the primary mechanism for voice services, LTE also supports inter-working with legacy systems for CS voice services.

[8]Global System for Mobile Communications.

This is controlled by the MME and is based on two procedures outlined in Sections 2.4.2.1 and 2.4.2.2.

More details of the radio interface procedures for inter-working with other RATs are specified in [3] and covered in Sections 2.5.6.2 and 3.2.4.

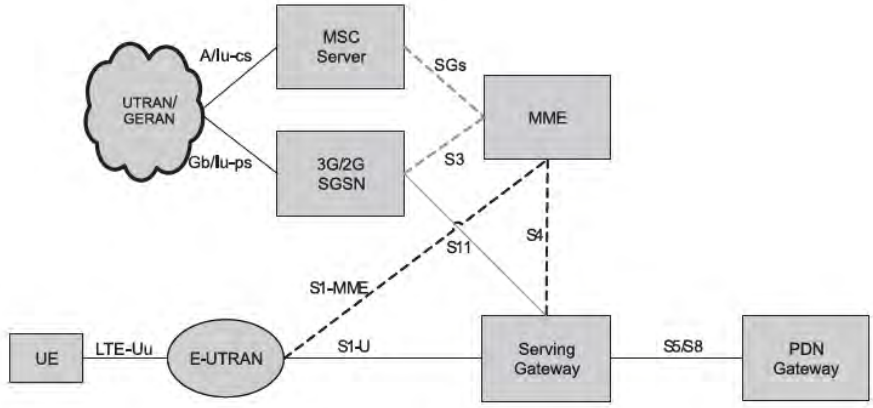


Figure 2.10: Architecture for 3G UMTS interworking.

2.4.2.1 Circuit-Switched Fall Back (CSFB)

LTE natively supports VoIP only using IMS services. However, in case IMS services are not deployed from the start, LTE also supports a Circuit-Switched FallBack (CSFB) mechanism which allows CS voice calls to be handled via legacy RATs for UEs that are camped on LTE.

CSFB allows a UE in LTE to be handed over to a legacy RAT to originate a CS voice call. This is supported by means of an interface, referred to as SGs[9], between the MME and the Mobile Switching Centre (MSC) of the legacy RAT shown in Figure 2.10. This interface allows the UE to attach with the MSC and register for CS services while still in LTE. Moreover it carries paging messages from the MSC for incoming voice calls so that UEs can be paged over LTE. The network may choose a handover, cell change order, or redirection procedure to move the UE to the legacy RAT.

Figure 2.11 shows the message flow for a CSFB call from LTE to UMTS, including paging from the MSC via the SGs interface and MME in the case of UE-terminated calls, and the sending of an Extended Service Request NAS message from the UE to the MME to trigger either a handover or redirection to the target RAT in the case of a UE-originated call. In the latter case, the UE then originates the CS call over the legacy RAT using the procedure defined in the legacy RAT specification. Further details of CSFB can be found in [7].

[9]SGs is an extension of the Gs interface between the Serving GPRS Support Node (SGSN) and the Mobile Switching Centre (MSC)

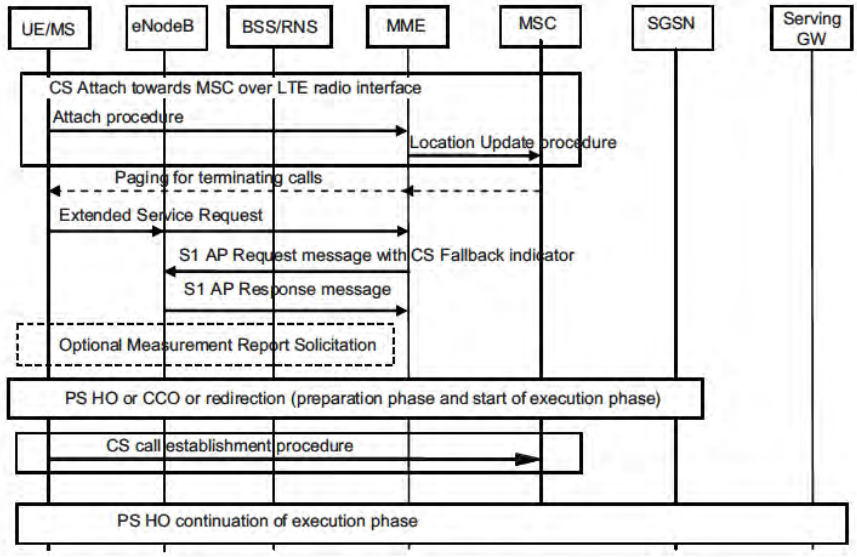


Figure 2.11: Message sequence diagram for CSFB from LTE to UMTS/GERAN.

2.4.2.2 Single Radio Voice Call Continuity (SRVCC)

If ubiquitous coverage of LTE is not available, it is possible that a UE involved in a VoIP call over LTE might then move out of LTE coverage to enter a legacy RAT cell which only offers CS voice services. The Single Radio Voice Call Continuity (SRVCC) procedure is designed for handover of a Packet Switched (PS) VoIP call over LTE to a CS voice call in the legacy RAT, involving the transfer of a PS bearer into a CS bearer.

Figure 2.12 shows an overview of the functions involved in SRVCC. The eNodeB may detect that the UE is moving out of LTE coverage and trigger a handover procedure towards the MME by means of an SRVCC indication. The MME is responsible for the SRVCC procedure and also for the transfer of the PS E-RAB carrying VoIP into a CS bearer. The MSC Server then initiates the session transfer procedure to IMS and coordinates it with the CS handover procedure to the target cell. The handover command provided to the UE to request handover to the legacy RAT also provides the information to set up the CS and PS radio bearers. The UE can continue with the call over the CS domain on completion of the handover. Further details of SRVCC can be found in [8].

2.5 The E-UTRAN Network Interfaces: S1 Interface

The S1 interface connects the eNodeB to the EPC. It is split into two interfaces, one for the control plane and the other for the user plane. The protocol structure for the S1 and the functionality provided over S1 are discussed in more detail below.

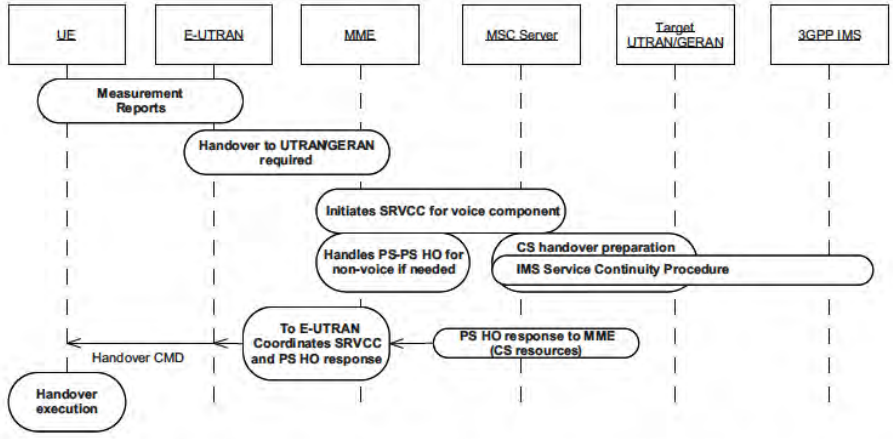


Figure 2.12: The main procedures involved in an SRVCC handover of a PS VoIP call from LTE to CS voice call in UMTS/GERAN.

2.5.1 Protocol Structure over S1

The protocol structure over S1 is based on a full IP transport stack with no dependency on legacy SS7[10] network configuration as used in GSM or UMTS networks. This simplification provides one area of potential savings on operational expenditure with LTE networks.

2.5.1.1 Control Plane

Figure 2.13 shows the protocol structure of the S1 control plane which is based on the Stream Control Transmission Protocol / IP (SCTP/IP) stack.

The SCTP protocol is well known for its advanced features inherited from TCP which ensure the required reliable delivery of the signalling messages. In addition, it makes it possible to benefit from improved features such as the handling of multistreams to implement transport network redundancy easily and avoid head-of-line blocking or multihoming (see 'IETF RFC4960' [9]).

A further simplification in LTE (compared to the UMTS Iu interface, for example) is the direct mapping of the S1-AP (S1 Application Protocol) on top of SCTP which results in a simplified protocol stack with no intermediate connection management protocol. The individual connections are directly handled at the application layer. Multiplexing takes place between S1-AP and SCTP whereby each stream of an SCTP association is multiplexed with the signalling traffic of multiple individual connections.

One further area of flexibility that comes with LTE lies in the lower layer protocols for which full optionality has been left regarding the choice of the IP version and the choice

[10]Signalling System #7 (SS7) is a communications protocol defined by the International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T) with a main purpose of setting up and tearing down telephone calls. Other uses include Short Message Service (SMS), number translation, prepaid billing mechanisms, and many other services.

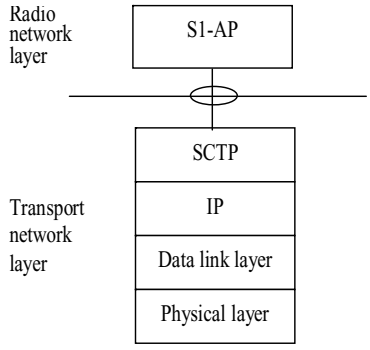


Figure 2.13: S1-MME control plane protocol stack. Reproduced by permission of © 3GPP.

of the data link layer. For example, this enables the operator to start deployment using IP version 4 with the data link tailored to the network deployment scenario.

2.5.1.2 User Plane

Figure 2.14 shows the protocol structure of the S1 user plane, which is based on the GTP/ User Datagram Protocol (UDP) IP stack which is already well known from UMTS networks.

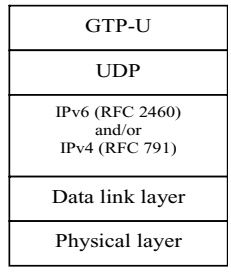


Figure 2.14: S1-U user plane protocol stack. Reproduced by permission of © 3GPP.

One of the advantages of using GTP-User plane (GTP-U) is its inherent facility to identify tunnels and also to facilitate intra-3GPP mobility.

The IP version number and the data link layer have been left fully optional, as for the control plane stack.

A transport bearer is identified by the GTP tunnel endpoints and the IP address (source Tunnelling End ID (TEID), destination TEID, source IP address, destination IP address).

The S-GW sends downlink packets of a given bearer to the eNodeB IP address (received in S1-AP) associated to that particular bearer. Similarly, the eNodeB sends upstream packets of a given bearer to the EPC IP address (received in S1-AP) associated to that particular bearer.

Vendor-specific traffic categories (e.g. real-time traffic) can be mapped onto Differentiated Services (Diffserv) code points (e.g. expedited forwarding) by network O&M (Operation and Maintenance) configuration to manage QoS differentiation between the bearers.

2.5.2 Initiation over S1

The initialization of the S1-MME control plane interface starts with the identification of the MMEs to which the eNodeB must connect, followed by the setting up of the Transport Network Layer (TNL).

With the support of the S1-flex function in LTE, an eNodeB must initiate an S1 interface towards each MME node of the pool area to which it belongs. This list of MME nodes of the pool together with an initial corresponding remote IP address can be directly configured in the eNodeB at deployment (although other means may also be used). The eNodeB then initiates the TNL establishment with that IP address. Only one SCTP association is established between one eNodeB and one MME.

During the establishment of the SCTP association, the two nodes negotiate the maximum number of streams which will be used over that association. However, multiple pairs of streams[11] are typically used in order to avoid the head-of-line blocking issue mentioned above. Among these pairs of streams, one particular pair must be reserved by the two nodes for the signalling of the common procedures (i.e. those which are not specific to one UE). The other streams are used for the sole purpose of the dedicated procedures (i.e. those which are specific to one UE).

Once the TNL has been established, some basic application-level configuration data for the system operation is automatically exchanged between the eNodeB and the MME through an 'S1 SETUP' procedure initiated by the eNodeB. This procedure is one case of a Self-Optimizing Network process and is explained in detail in Section 25.3.1.

Once the S1 SETUP procedure has been completed, the S1 interface is operational.

2.5.3 Context Management over S1

Within each pool area, a UE is associated to one particular MME for all its communications during its stay in this pool area. This creates a context in this MME for the UE. This particular MME is selected by the NAS Node Selection Function (NNSF) in the first eNodeB from which the UE entered the pool.

Whenever the UE becomes active (i.e. makes a transition from idle to active mode) under the coverage of a particular eNodeB in the pool area, the MME provides the UE context information to this eNodeB using the 'INITIAL CONTEXT SETUP REQUEST' message (see Figure 2.15). This enables the eNodeB in turn to create a context and manage the UE while it is in active mode.

Even though the setup of bearers is otherwise relevant to a dedicated 'Bearer Management' procedure described below, the creation of the eNodeB context by the INITIAL CONTEXT SETUP procedure also includes the creation of one or several bearers including the default bearers.

At the next transition back to idle mode following a 'UE CONTEXT RELEASE' message sent from the MME, the eNodeB context is erased and only the MME context remains.

[11] Note that a stream is unidirectional and therefore pairs must be used.

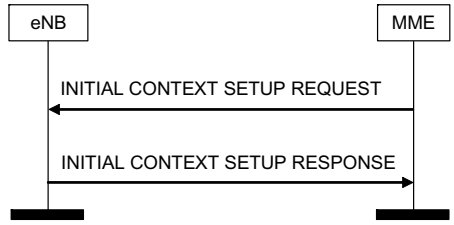


Figure 2.15: Initial context setup procedure. Reproduced by permission of © 3GPP.

2.5.4 Bearer Management over S1

LTE uses independent dedicated procedures respectively covering the setup, modification and release of bearers. For each bearer requested to be set up, the transport layer address and the tunnel endpoint are provided to the eNodeB in the 'BEARER SETUP REQUEST' message to indicate the termination of the bearer in the S-GW where uplink user plane data must be sent. Conversely, the eNodeB indicates in the 'BEARER SETUP RESPONSE' message the termination of the bearer in the eNodeB where the downlink user plane data must be sent.

For each bearer, the QoS parameters (see Section 2.4) requested for the bearer are also indicated. Independently of the standardized QCI values, it is also still possible to use extra proprietary labels for the fast introduction of new services if vendors and operators agree upon them.

2.5.5 Paging over S1

As mentioned in Section 2.5.3, in order to re-establish a connection towards a UE in idle mode, the MME distributes a 'PAGING REQUEST' message to the relevant eNodeBs based on the TAs where the UE is expected to be located. When receiving the paging request, the eNodeB sends a page over the radio interface in the cells which are contained within one of the TAs provided in that message.

The UE is normally paged using its S-TMSI. The 'PAGING REQUEST' message also contains a UE identity index value in order for the eNodeB to calculate the paging occasions at which the UE will switch on its receiver to listen for paging messages (see Section 3.4).

In Release 10, paging differentiation is introduced over the S1 interface to handle Multimedia Priority Service (MPS)[12] users. In case of MME or RAN overload, it is necessary to page a UE with higher priority during the establishment of a mobile-terminated MPS call. In case of MME overload, the MME can itself discriminate between the paging messages and discard the lower priority ones. In case of RAN overload in some cells, the eNodeB can perform this discrimination based on a new Paging Priority Indicator sent by the MME. The MME can signal up to eight such priority values to the eNodeB. In case of an IMS MPS call, the terminating UE will further set up an RRC connection with the same eNodeB that will also get automatically prioritized. In case of a CS fallback call, the eNodeB will instead

[12]MPS allows the delivery of calls or complete sessions of a high priority nature, in case for example of public safety or national security purposes, from mobile to mobile, mobile to fixed, and fixed to mobile networks during network congestion conditions.

signal to the UE that it must set the cause value ‘high priority terminating call’ when trying to establish the UMTS RRC Connection.

2.5.6 Mobility over S1

LTE/SAE supports mobility within LTE/SAE, and also to other systems using both 3GPP and non-3GPP technologies. The mobility procedures over the radio interface are defined in Section 3.2. These mobility procedures also involve the network interfaces. The sections below discuss the procedures over S1 to support mobility. The mobility performance requirements from the UE point of view are outlined in Chapter 22.

2.5.6.1 Intra-LTE Mobility

There are two types of handover procedure in LTE for UEs in active mode: the S1-handover procedure and the X2-handover procedure.

For intra-LTE mobility, the X2-handover procedure is normally used for the inter-eNodeB handover (described in Section 2.6.3). However, when there is no X2 interface between the two eNodeBs, or if the source eNodeB has been configured to initiate handover towards a particular target eNodeB via the S1 interface, then an S1-handover will be triggered.

The S1-handover procedure has been designed in a very similar way to the UMTS Serving Radio Network Subsystem (SRNS) relocation procedure and is shown in Figure 2.16: it consists of a preparation phase involving the core network, where the resources are first prepared at the target side (steps 2 to 8), followed by an execution phase (steps 8 to 12) and a completion phase (after step 13).

Compared to UMTS, the main difference is the introduction of the ‘STATUS TRANSFER’ message sent by the source eNodeB (steps 10 and 11). This message has been added in order to carry some PDCP status information that is needed at the target eNodeB in cases when PDCP status preservation applies for the S1-handover (see Section 4.2.4); this is in alignment with the information which is sent within the X2 ‘STATUS TRANSFER’ message used for the X2-handover (see below). As a result of this alignment, the handling of the handover by the target eNodeB as seen from the UE is exactly the same, regardless of the type of handover (S1 or X2) the network had decided to use; indeed, the UE is unaware of which type of handover is used by the network.

The ‘Status Transfer’ procedure is assumed to be triggered in parallel with the start of data forwarding after the source eNodeB has received the ‘HANDOVER COMMAND’ message from the source MME. This data forwarding can be either direct or indirect, depending on the availability of a direct path for the user plane data between the source eNodeB and the target eNodeB.

The ‘HANDOVER NOTIFY’ message (step 13), which is sent later by the target eNodeB when the arrival of the UE at the target side is confirmed, is forwarded by the MME to trigger the update of the path switch in the S-GW towards the target eNodeB. In contrast to the X2-handover, the message is not acknowledged and the resources at the source side are released later upon reception of a ‘RELEASE RESOURCE’ message directly triggered from the source MME (step 17 in Figure 2.16).

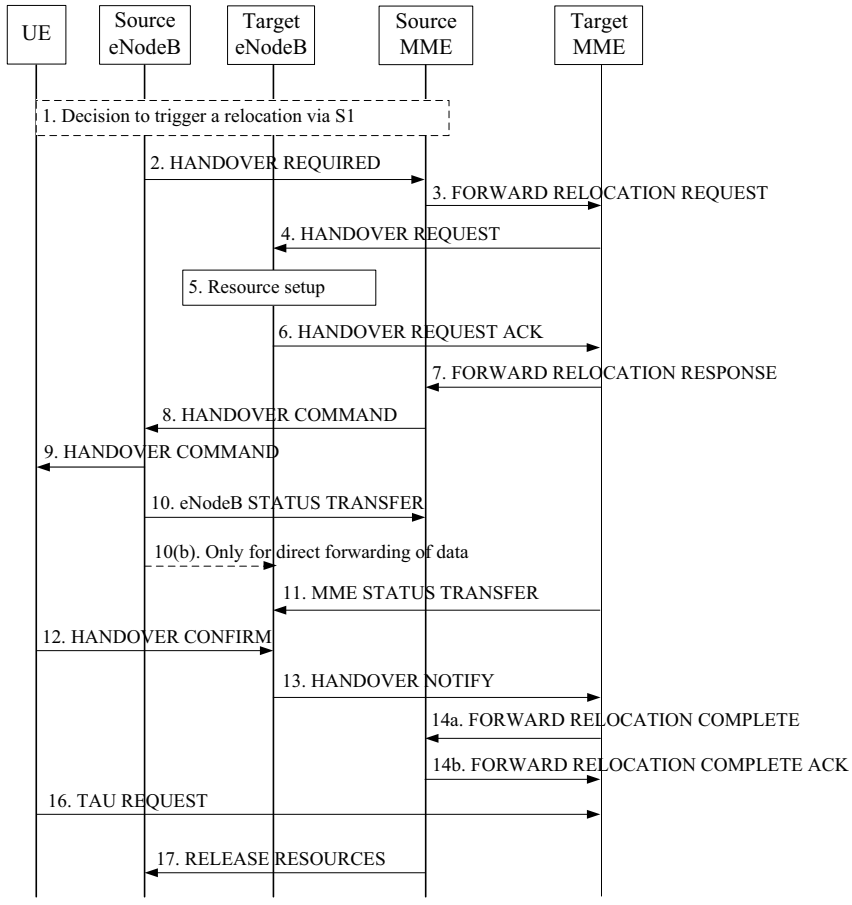


Figure 2.16: S1-based handover procedure. Reproduced by permission of © 3GPP.

2.5.6.2 Inter-RAT Mobility

One key element of the design of LTE is the need to co-exist with other Radio Access Technologies (RATs).

For mobility from LTE towards UMTS, the handover process can reuse the S1-handover procedures described above, with the exception of the 'STATUS TRANSFERŠ message which is not needed at steps 10 and 11 since no PDCP context is continued.

For mobility towards CDMA2000, dedicated uplink and downlink procedures have been introduced in LTE. They essentially aim at tunnelling the CDMA2000 signalling between the UE and the CDMA2000 system over the S1 interface, without being interpreted by the eNodeB on the way. An 'UPLINK S1 CDMA2000 TUNNELLING' message is sent from the eNodeB to the MME; this also includes the RAT type in order to identify which CDMA2000

RAT the tunnelled CDMA2000 message is associated with in order for the message to be routed to the correct node within the CDMA2000 system.

2.5.6.3 Mobility towards Home eNodeBs

Mobility towards HeNBs involves additional functions from the source LTE RAN node and the MME. In addition to the E-UTRAN Cell Global Identifier (ECGI), the source RAN node should include the Closed Subscriber Group Identity (CSG ID) and the access mode of the target HeNB in the 'HANDOVER REQUIRED' message to the MME so that the MME can perform the access control to that HeNB. If the target HeNB operates in closed access mode (see Chapter 24) and the MME fails the access control, the MME will reject the handover by sending back a 'HANDOVER PREPARATION FAILURE' message. Otherwise the MME will accept and continue the handover while indicating to the target HeNB whether the UE is a 'CSG member' if the HeNB is operating in hybrid mode. A detailed description of mobility towards the HeNB and the associated call flow is provided in Chapter 24.

2.5.7 Load Management over S1

Three types of load management procedures apply over S1: a normal 'load balancing' procedure to distribute the traffic, an 'overload' procedure to overcome a sudden peak in the loading and a 'load rebalancing' procedure to partially/fully offload an MME.

The MME load balancing procedure aims to distribute the traffic to the MMEs in the pool evenly according to their respective capacities. To achieve that goal, the procedure relies on the normal NNSF present in each eNodeB as part of the S1-flex function. Provided that suitable weight factors corresponding to the capacity of each MME node are available in the eNodeBs beforehand, a weighted NNSF done by every eNodeB in the network normally achieves a statistically balanced distribution of load among the MME nodes without further action. However, specific actions are still required for some particular scenarios:

- If a new MME node is introduced (or removed), it may be necessary temporarily to increase (or decrease) the weight factor normally corresponding to the capacity of this node in order to make it catch more (or less) traffic at the beginning until it reaches an adequate level of load.

- In case of an unexpected peak in the loading, an 'OVERLOAD' message can be sent over the S1 interface by the overloaded MME. When received by an eNodeB, this message calls for a temporary restriction of a certain type of traffic. An MME can adjust the reduction of traffic it desires by defining the number of eNodeBs to which it sends the 'OVERLOAD' message and by defining the types of traffic subject to restriction.Two new rejection types are introduced in Release 10 to combat CN Overload:

  - 'reject low priority access', which can be used by the MME to reduce access of some low-priority devices or applications such as Machine-Type Communication (MTC) devices (see Section 31.4);
  - 'permit high priority sessions', to allow access only to high-priority users and mobile-terminated services.

- Finally, if the MME wants to force rapidly the offload of part or all of its UEs, it will use the rebalancing function. This function forces the UEs to reattach to another MME by using a specific 'cause value' in the 'UE Release Command S1' message. In a first step it applies to idle mode UEs and in a second step it may also apply to UEs in connected mode (if the full MME offload is desired, e.g. for maintenance reasons).

2.5.8 Trace Function

In order to trace the activity of a UE in connected mode, two types of trace session can be started in the eNodeB:

- Signalling-Based Trace. This is triggered by the MME and is uniquely identified by a trace identity. Only one trace session can be activated at a time for one UE. The MME indicates to the eNodeB the interfaces to trace (e.g. S1, X2, Uu) and the associated trace depth. The trace depth represents the granularity of the signalling to be traced from the high-level messages down to the detailed ASN.1[13] and is comprised of three levels: minimum, medium and maximum. The MME also indicates the IP address of a Trace Collection Entity where the eNodeB must send the resulting trace record file. If an X2 handover preparation has started at the time when the eNodeB receives the order to trace, the eNodeB will signal back a TRACE FAILURE INDICATION message to the MME, and it is then up to the MME to take appropriate action based on the indicated failure reason. Signalling-based traces are propagated at X2 and S1 handover.

- Management-Based Trace. This is triggered in the eNodeB when the conditions required for tracing set by O&M are met. The eNodeB then allocates a trace identity that it sends to the MME in a CELL TRAFFIC TRACE message over S1, together with the Trace Collection Entity identity that shall be used by the MME for the trace record file (in order to assemble the trace correctly in the Trace Collection Entity). Management-based traces are propagated at X2 and S1 handover.

In Release 10, the trace function supports the Minimization of Drive Tests (MDT) feature, which is explained in Section 31.3.

2.5.9 Delivery of Warning Messages

Two types of warning message may need to be delivered with the utmost urgency over a cellular system, namely Earthquake and Tsunami Warning System (ETWS)) messages and Commercial Mobile Alert System (CMAS) messages (see Section 13.7). The delivery of ETWS messages is already supported since Release 8 via the S1 Write-Replace Warning procedure which makes it possible to carry either primary or secondary notifications over S1 for the eNodeB to broadcast over the radio. The Write-Replace Warning procedure also includes a Warning Area List where the warning message needs to be broadcast. It can be a list of cells, tracking areas or emergency area identities. The procedure also contains information on how the broadcast is to be performed (for example, the number of broadcasts requested).

[13] Abstract Syntax Notation One

In contrast to ETWS, the delivery of CMAS messages is only supported from Release 9 onwards. One difference between the two public warning systems is that in ETWS the eNodeB can only broadcast one message at a time, whereas CMAS allows the broadcast of multiple concurrent warning messages over the radio. Therefore an ongoing ETWS broadcast needs to be overwritten if a new ETWS warning has to be delivered immediately in the same cell. With CMAS, a new Kill procedure has also been added to allow easy cancellation of an ongoing broadcast when needed. This Kill procedure includes the identity of the message to be stopped and the Warning Area where it is to be stopped.

2.6 The E-UTRAN Network Interfaces: X2 Interface

The X2 interface is used to inter-connect eNodeBs. The protocol structure for the X2 interface and the functionality provided over X2 are discussed below.

2.6.1 Protocol Structure over X2

The control plane and user plane protocol stacks over the X2 interface are the same as over the S1 interface, as shown in Figures 2.17 and 2.18 respectively (with the exception that in Figure 2.17 the X2-AP (X2 Application Protocol) is substituted for the S1-AP). This also means again that the choice of the IP version and the data link layer are fully optional. The use of the same protocol structure over both interfaces provides advantages such as simplifying the data forwarding operation.

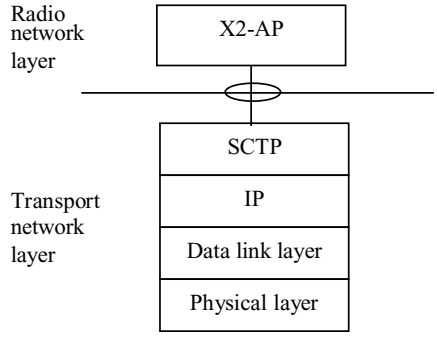


Figure 2.17: X2 signalling bearer protocol stack. Reproduced by permission of © 3GPP.

2.6.2 Initiation over X2

The X2 interface may be established between one eNodeB and some of its neighbour eNodeBs in order to exchange signalling information when needed. However, a full mesh is not mandated in an E-UTRAN network. Two types of information may typically need to be exchanged over X2 to drive the establishment of an X2 interface between two eNodeBs:

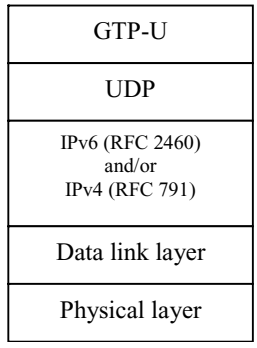


Figure 2.18: Transport network layer for data streams over X2. Reproduced by permission of © 3GPP.

load or interference related information (see Section 2.6.4) and handover related information (see mobility in Section 2.6.3).

Because these two types of information are fully independent of one another, it is possible that an X2 interface may be present between two eNodeBs for the purpose of exchanging load or interference information, even though the X2-handover procedure is not used to hand over UEs between those eNodeBs.[14]

The initialization of the X2 interface starts with the identification of a suitable neighbour followed by the setting up of the TNL.

The identification of a suitable neighbour may be done by configuration, or alternatively by a self-optimizing process known as the Automatic Neighbour Relation Function (ANRF).[15] This is described in more detail in Section 25.2.

Once a suitable neighbour has been identified, the initiating eNodeB can further set up the TNL using the transport layer address of this neighbour – either as retrieved from the network or locally configured. The automatic retrieval of the X2 IP address(es) via the network and the eNodeB Configuration Transfer procedure are described in details in Section 25.3.2.

Once the TNL has been set up, the initiating eNodeB must trigger the X2 setup procedure. This procedure enables an automatic exchange of application level configuration data relevant to the X2 interface, similar to the S1 setup procedure already described in Section 2.5.2. For example, each eNodeB reports within the 'X2 SETUP REQUESTŠ message to a neighbour eNodeB information about each cell it manages, such as the cell's physical identity, the frequency band, the tracking area identity and/or the associated PLMNs.

This automatic exchange of application-level configuration data within the X2 setup procedure is also the core of two additional SON features: automatic self-configuration of the Physical Cell Identities (PCIs) and RACH self-optimization. These features both aim to avoid conflicts between cells controlled by neighbouring eNodeBs; they are explained in detail in Sections 25.4 and 25.7 respectively.

Once the X2 setup procedure has been completed, the X2 interface is operational.

[14] In such a case, the S1-handover procedure is used instead.

[15] Under this function the UEs are requested to detect neighbour eNodeBs by reading the Cell Global Identity (CGI) contained in the broadcast information.

2.6.3 Mobility over X2

Handover via the X2 interface is triggered by default unless there is no X2 interface established or the source eNodeB is configured to use the S1-handover instead.

The X2-handover procedure is illustrated in Figure 2.19. Like the S1-handover, it is also composed of a preparation phase (steps 4 to 6), an execution phase (steps 7 to 9) and a completion phase (after step 9).

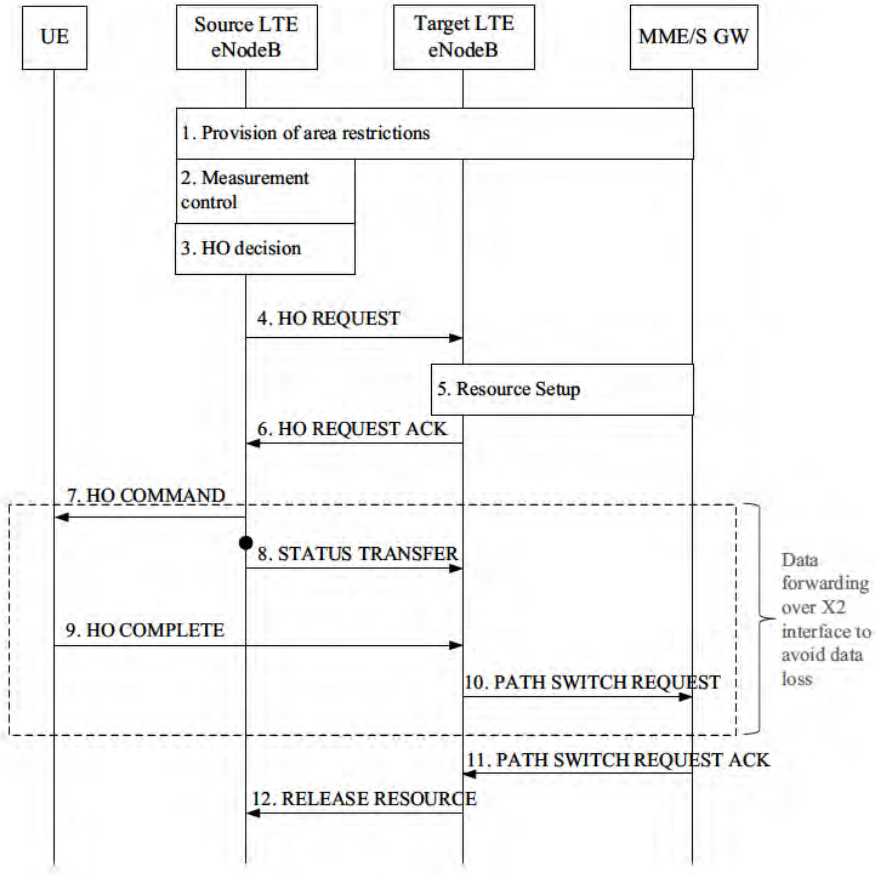


Figure 2.19: X2-based handover procedure.

The key features of the X2-handover for intra-LTE handover are:

- The handover is directly performed between two eNodeBs. This makes the preparation phase quick.

- Data forwarding may be operated per bearer in order to minimize data loss.
- The MME is only informed at the end of the handover procedure once the handover is successful, in order to trigger the path switch.
- The release of resources at the source side is directly triggered from the target eNodeB.

For those bearers for which in-sequence delivery of packets is required, the 'STATUS TRANSFER' message (step 8) provides the Sequence Number (SN) and the Hyper Frame Number (HFN) which the target eNodeB should assign to the first packet with no sequence number yet assigned that it must deliver. This first packet can either be one received over the target S1 path or one received over X2 if data forwarding over X2 is used (see below). When it sends the 'STATUS TRANSFER' message, the source eNodeB freezes its transmitter/receiver status – i.e. it stops assigning PDCP SNs to downlink packets and stops delivering uplink packets to the EPC.

Mobility over X2 can be categorized according to its resilience to packet loss: the handover can be said 'seamless' if it minimizes the interruption time during the move of the UE, or 'lossless' if it tolerates no loss of packets at all. These two modes use data forwarding of user plane downlink packets. The source eNodeB may decide to operate one of these two modes on a per-EPS-bearer basis, based on the QoS received over S1 for this bearer (see Section 2.5.4) and the service in question. These two modes are described in more detail below.

2.6.3.1 Seamless Handover

If, for a given bearer, the source eNodeB selects the seamless handover mode, it proposes to the target eNodeB in the 'HANDOVER REQUEST' message to establish a GTP tunnel to operate the downlink data forwarding. If the target eNodeB accepts, it indicates in the 'HANDOVER REQUEST ACK' message the tunnel endpoint where the forwarded data is expected to be received. This tunnel endpoint may be different from the one set up as the termination point of the new bearer established over the target S1.

Upon reception of the 'HANDOVER REQUEST ACK' message, the source eNodeB can start forwarding the data freshly arriving over the source S1 path towards the indicated tunnel endpoint in parallel with sending the handover trigger to the UE over the radio interface. This forwarded data is thus available at the target eNodeB to be delivered to the UE as early as possible.

When forwarding is in operation and in-sequence delivery of packets is required, the target eNodeB is assumed to deliver first the packets forwarded over X2 before delivering the first ones received over the target S1 path once the S1 path switch has been performed. The end of the forwarding is signalled over X2 to the target eNodeB by the reception of some 'special GTP packets' which the S-GW has inserted over the source S1 path just before switching this S1 path; these are then forwarded by the source eNodeB over X2 like any other regular packets.

2.6.3.2 Lossless Handover

If the source eNodeB selects the lossless mode for a given bearer, it will additionally forward over X2 those user plane downlink packets which it has PDCP processed but are still buffered locally because they have not yet been delivered and acknowledged by the UE. These packets

are forwarded together with their assigned PDCP SN included in a GTP extension header field. They are sent over X2 prior to the freshly arriving packets from the source S1 path. The same mechanisms described above for the seamless handover are used for the GTP tunnel establishment. The end of forwarding is also handled in the same way, since in-sequence packet delivery applies to lossless handovers. In addition, the target eNodeB must ensure that all the packets – including the ones received with sequence number over X2 – are delivered in sequence at the target side. Further details of seamless and lossless handover are described in Section 4.2.

Selective retransmission. A new feature in LTE compared to previous systems is the optimization of the radio interface usage by selective retransmission. When lossless handover is operated, the target eNodeB may, however, not deliver over the radio interface some of the forwarded downlink packets received over X2 if it is informed by the UE that those packets have already been received at the source side (see Section 4.2.6). This is called downlink selective retransmission.

Similarly in the uplink, the target eNodeB may desire that the UE does not retransmit packets already received earlier at the source side by the source eNodeB, for example to avoid wasting radio resources. To operate this uplink selective retransmission scheme for one bearer, it is necessary that the source eNodeB forwards to the target eNodeB, over another new GTP tunnel, those user plane uplink packets which it has received out of sequence. The target eNodeB must first request the source eNodeB to establish this new forwarding tunnel by including in the 'HANDOVER REQUEST ACK' message a GTP tunnel endpoint where it expects the forwarded uplink packets to be received. The source eNodeB must, if possible, then indicate in the 'STATUS TRANSFER' message for this bearer the list of SNs corresponding to the forwarded packets which are to be expected. This list helps the target eNodeB to inform the UE earlier of the packets not to be retransmitted, making the overall uplink selective retransmission scheme faster (see also Section 4.2.6).

2.6.3.3 Multiple Preparation

'Multiple preparation' is another new feature of the LTE handover procedure. This feature enables the source eNodeB to trigger the handover preparation procedure towards multiple candidate target eNodeBs. Even though only one of the candidates is indicated as target to the UE, this makes recovery faster in case the UE fails on this target and connects to one of the other prepared candidate eNodeBs. The source eNodeB receives only one 'RELEASE RESOURCE' message from the final selected eNodeB.

Regardless of whether multiple or single preparation is used, the handover can be cancelled during or after the preparation phase. If the multiple preparation feature is operated, it is recommended that upon reception of the 'RELEASE RESOURCE' message the source eNodeB triggers a 'cancel' procedure towards each of the non-selected prepared eNodeBs.

2.6.3.4 Mobility Robustness Handling

In order to detect and report the cases where the mobility is unsuccessful and results in connection failures, specific messages are available over the X2 interface from Release 9 onwards to report handovers that are triggered too late or too early or to an inappropriate cell. These scenarios are explained in detail in Section 25.6.

2.6.3.5 Mobility towards Home eNodeBs via X2

In Release 10, in order to save backhaul bandwidth reduce delays, mobility between two HeNBs does not necessarily need to use S1 handover and transit via the MME but can directly use the X2 handover. This optimization is described in detail in Section 24.2.3.

2.6.4 Load and Interference Management Over X2

The exchange of load information between eNodeBs is of key importance in the flat architecture used in LTE, as there is no central Radio Resource Management (RRM) node as was the case, for example, in UMTS with the Radio Network Controller (RNC).

The exchange of load information falls into two categories depending on the purpose it serves:

- **Load balancing.** If the exchange of load information is for the purpose of load balancing, the frequency of exchange is rather low (in the order of seconds). The objective of load balancing is to counteract local traffic load imbalance between neighbouring cells with the aim of improving the overall system capacity. The mechanisms for this are explained in detail in Section 25.5.

  In Release 10, partial reporting is allowed per cell and per measurement. Therefore, if a serving eNodeB does not support some measurements, it will still report the other measurements that it does support. For each unsupported measurement, the serving eNodeB can indicate if the lack of support is permanent or temporary.

- **Interference coordination.** If the exchange of load information is to optimize RRM processes such as interference coordination, the frequency of exchange is rather high (in the order of tens of milliseconds). A special X2 'LOAD INDICATION' message is provided over the X2 interface for the exchange of load information related to interference management. For uplink interference management, two indicators can be provided within the 'LOAD INDICATION' message: a 'High Interference Indicator' and an 'Overload Indicator'. The usage of these indicators is explained in detail in Section 12.5.

The Load Indication procedure allows an eNodeB to signal to its neighbour eNodeBs new interference coordination intentions when applicable. This can either be frequency-domain interference management, as explained in Sections 12.5.1 and 12.5.2, or time-domain interference management, as explained in Section 31.2.3.

2.6.5 UE Historical Information Over X2

The provision of UE historical information is part of the X2-handover procedure and is designed to support self-optimization of the network.

Generally, the UE historical information consists of some RRM information which is passed from the source eNodeB to the target eNodeB within the 'HANDOVER REQUEST' message to assist the RRM management of a UE. The information can be partitioned into two types:

- UE RRM-related information, passed over X2 within the RRC transparent container;

- Cell RRM-related information, passed over X2 directly as an information element of the 'X2 AP HANDOVER REQUEST' message itself.

An example of such UE historical information is the list of the last few cells visited by the UE, together with the time spent in each one. This information is propagated from one eNodeB to another and can be used to determine the occurrence of ping-pong between two or three cells for instance. The length of the history information can be configured for more flexibility.

2.7 Summary

The EPS provides UEs with IP connectivity to the packet data network. In this chapter we have seen an overview of the EPS network architecture, including the functionalities provided by the E-UTRAN access network and the evolved packet core network..

It can be seen that the concept of EPS bearers, together with their associated quality of service attributes, provide a powerful tool for the provision of a variety of simultaneous services to the end user. Depending on the nature of the application, the EPS can supply the UE with multiple data flows with different QoSs. A UE can thus be engaged in a VoIP call which requires guaranteed delay and bit rate at the same time as browsing the web with a best effort QoS.

From the perspective of the network operator, the LTE system breaks new ground in terms of its degree of support for self-optimization and self-configuration of the network via the X2, S1 and Uu interfaces; these aspects are described in more detail in Chapter 25.

References[16]

[1] 3GPP Technical Specification 24.301, 'Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS); Stage 3', www.3gpp.org.

[2] 3GPP Technical Specification 33.401, 'System Architecture Evolution (SAE): Security Architecture', www.3gpp.org.

[3] 3GPP Technical Specification 23.402, 'Architecture enhancements for non-3GPP accesses', www.3gpp.org.

[4] 3GPP Technical Specification 29.060, 'General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface', www.3gpp.org.

[5] 3GPP Technical Specification 23.203, 'Policy and charging control architecture', www.3gpp.org.

[6] 3GPP Technical Specification 36.300, 'Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2', www.3gpp.org.

[7] 3GPP Technical Specification 23.272, ' Circuit Switched (CS) fallback in Evolved Packet System (EPS); Stage 2', www.3gpp.org.

[8] 3GPP Technical Specification 23.272, 'Single Radio Voice Call Continuity (SRVCC); Stage 2', www.3gpp.org.

[9] Request for Comments 4960 The Internet Engineering Task Force (IETF), Network Working Group, 'Stream Control Transmission Protocol', http://www.ietf.org.

[16] All web sites confirmed 1st March 2011.

3

Control Plane Protocols

Himke van der Velde

3.1 Introduction

As introduced in Section 2.2.2, the Control Plane of the Access Stratum (AS) handles radio-specific functionalities. The AS interacts with the Non-Access Stratum (NAS), also referred to as the 'upper layers'. Among other functions, the NAS control protocols handle Public Land Mobile Network[1] (PLMN) selection, tracking area update, paging, authentication and Evolved Packet System (EPS) bearer establishment, modification and release.

The applicable AS-related procedures largely depend on the Radio Resource Control (RRC) state of the User Equipment (UE), which can be either RRC_IDLE or RRC_CONNECTED.

A UE in RRC_IDLE performs cell selection and reselection – in other words, it decides on which cell to camp. The cell (re)selection process takes into account the priority of each applicable frequency of each applicable Radio Access Technology (RAT), the radio link quality and the cell status (i.e. whether a cell is barred or reserved). An RRC_IDLE UE monitors a paging channel to detect incoming calls, and also acquires system information. The System Information (SI) mainly consists of parameters by which the network (E-UTRAN) can control the cell (re)selection process.

In RRC_CONNECTED, the E-UTRAN allocates radio resources to the UE to facilitate the transfer of (unicast) data via shared data channels.[2] To support this operation, the UE monitors an associated control channel[3] used to indicate the dynamic allocation of the shared transmission resources in time and frequency. The UE provides the network with reports of its

[1] The network of one operator in one country.

[2] The Physical Downlink Shared CHannel (PDSCH) and Physical Uplink Shared CHannel (PUSCH)– see Sections 9.2.2 and 16.2 respectively.

[3] The Physical Downlink Control CHannel (PDCCH) – see Section 9.3.5.

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

buffer status and of the downlink channel quality, as well as neighbouring cell measurement information to enable E-UTRAN to select the most appropriate cell for the UE. These measurement reports include cells using other frequencies or RATs. The UE also receives SI, consisting mainly of information required to use the transmission channels. To extend its battery lifetime, a UE in RRC_CONNECTED may be configured with a Discontinuous Reception (DRX) cycle.

RRC, as specified in [1], is the protocol by which the E-UTRAN controls the UE behaviour in RRC_CONNECTED. RRC also includes the control signalling applicable for a UE in RRC_IDLE, namely paging and SI. The UE behaviour in RRC_IDLE is specified in [2].

Chapter 22 gives some further details of the UE measurements which support the mobility procedures.

Functionality related to Multimedia Broadcast/Multicast Services (MBMSs) is covered separately in Chapter 13.

3.2 Radio Resource Control (RRC)

3.2.1 Introduction

The RRC protocol supports the transfer of *common* NAS information (i.e. NAS information which is applicable to all UEs) as well as *dedicated* NAS information (which is applicable only to a specific UE). In addition, for UEs in RRC_IDLE, RRC supports notification of incoming calls (via paging).

The RRC protocol covers a number of functional areas:

- **System information** handles the broadcasting of SI, which includes NAS common information. Some of the system information is applicable only for UEs in RRC_IDLE while other SI is also applicable for UEs in RRC_CONNECTED.
- **RRC connection control** covers all procedures related to the establishment, modification and release of an RRC connection, including paging, initial security activation, establishment of Signalling Radio Bearers (SRBs) and of radio bearers carrying user data (Data Radio Bearers, DRBs), handover within LTE (including transfer of UE RRC context information[4]), configuration of the lower protocol layers,[5] access class barring and radio link failure.
- **Network controlled inter-RAT mobility** includes handover, cell change orders and redirection upon connection release, security activation and transfer of UE RRC context information.
- **Measurement configuration and reporting** for intra-frequency, inter-frequency and inter-RAT mobility, includes configuration and activation of measurement gaps.
- **Miscellaneous functions** including, for example, transfer of dedicated NAS information and transfer of UE radio access capability information.

[4]This UE context information includes the radio resource configuration including local settings not configured across the radio interface, UE capabilities and radio resource management information.

[5]Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC), Medium Access Control (MAC), all of which are explained in detail in Chapter 4, and the physical layer which is explained in Chapters 5–11 and 14–18.

Dedicated RRC messages are transferred across SRBs, which are mapped via the PDCP and RLC layers onto logical channels – either the Common Control CHannel (CCCH) during connection establishment or a Dedicated Control CHannel (DCCH) in RRC_CONNECTED. System Information and Paging messages are mapped directly to logical channels – the Broadcast Control CHannel (BCCH) and Paging Control CHannel (PCCH) respectively. The various logical channels are described in more detail in Section 4.4.1.2.

SRB0 is used for RRC messages which use the CCCH, SRB1 is for RRC messages using DCCH, and SRB2 is for the (lower-priority) RRC messages using DCCH which only include NAS dedicated information.[6] All RRC messages using DCCH are integrity-protected and ciphered by the PDCP layer (after security activation) and use Automatic Repeat reQuest (ARQ) protocols for reliable delivery through the RLC layer. The RRC messages using CCCH are not integrity-protected and do not use ARQ in the RLC layer.

It should also be noted that the NAS independently applies integrity protection and ciphering.

Figure 3.1 illustrates the overall radio protocol architecture as well as the use of radio bearers, logical channels, transport channels and physical channels.

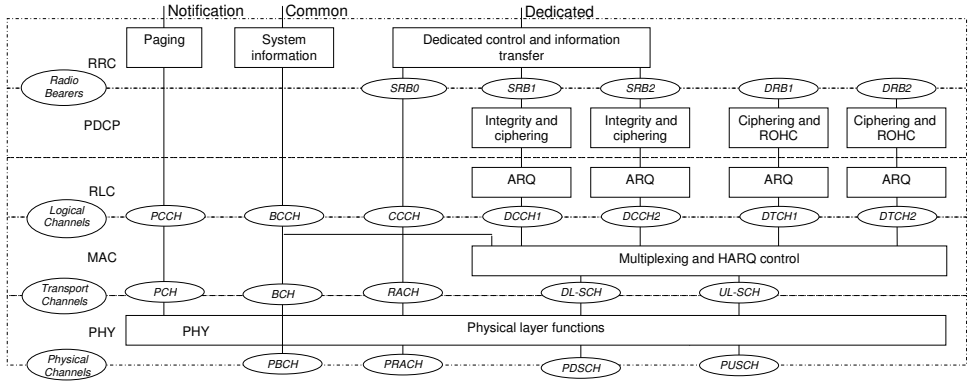


Figure 3.1: Radio architecture.

For control information for which low transfer delay is more important than reliable transfer (i.e. for which the use of ARQ is inappropriate due to the additional delay it incurs), MAC signalling is used provided that there are no security concerns (integrity protection and ciphering are not applicable for MAC signalling).

3.2.2 System Information

System information is structured by means of System Information Blocks (SIBs), each of which contains a set of functionally-related parameters. The SIB types that have been defined include:

[6]Prior to SRB2 establishment, SRB1 is also used for RRC messages which only include NAS dedicated information. In addition, SRB1 is used for higher priority RRC messages which only include NAS dedicated information.

- **The Master Information Block** (MIB), which includes a limited number of the most frequently transmitted parameters which are essential for a UE's initial access to the network.
- **System Information Block Type 1** (SIB1), which contains parameters needed to determine if a cell is suitable for cell selection, as well as information about the time-domain scheduling of the other SIBs.
- **System Information Block Type 2** (SIB2), which includes common and shared channel information.
- **SIB3–SIB8**, which include parameters used to control intra-frequency, inter-frequency and inter-RAT cell reselection.
- **SIB9**, which is used to signal the name of a Home eNodeB (HeNBs).
- **SIB10–SIB12**, which include the Earthquake and Tsunami Warning Service (ETWS) notifications and Commercial Mobile Alert System (CMAS) warning messages (See Section 13.7).
- **SIB13**, which includes MBMS related control information (See Section 13.6.3.2.

Three types of RRC message are used to transfer system information: the MIB message, the SIB1 message and SI messages. An SI message, of which there may be several, includes one or more SIBs which have the same scheduling requirements (i.e. the same transmission periodicity). Table 3.1 provides an example of a possible system information scheduling configuration, also showing which SIBs the UE has to acquire in the idle and connected states. The physical channels used for carrying the SI are explained in Section 9.2.1.

Table 3.1: Example of SI scheduling configuration.

Message	Content	Period (ms)	Applicability
MIB	Most essential parameters	40	Idle and connected
SIB1	Cell access related parameters, scheduling information	80	Idle and connected
1st SI	SIB2: Common and shared channel configuration	160	Idle and connected
2nd SI	SIB3: Common cell reselection information and intra-frequency cell reselection parameters other than the neighbouring cell information SIB4: Intra-frequency neighbouring cell information	320	Idle only
3rd SI	SIB5: Inter-frequency cell reselection information	640	Idle only
4th SI	SIB6: UTRA cell reselection information SIB7: GERAN cell reselection information	640	Idle only, depending on UE support of UMTS or GERAN

3.2.2.1 Time-Domain Scheduling of System Information

The time-domain scheduling of the MIB and SIB1 messages is fixed with a periodicities of 40 ms and 80 ms respectively, as explained in Sections 9.2.1 and 9.2.2.2.

The time-domain scheduling of the SI messages is dynamically flexible: each SI message is transmitted in a defined periodically-occurring time-domain window, while physical layer control signalling[7] indicates in which subframes[8] within this window the SI is actually scheduled. The scheduling windows of the different SI messages (referred to as SI-windows) are consecutive (i.e. there are neither overlaps nor gaps between them) and have a common length that is configurable. SI-windows can include subframes in which it is not possible to transmit SI messages, such as subframes used for SIB1, and subframes used for the uplink in TDD.

Figure 3.2 illustrates an example of the time-domain scheduling of SI, showing the subframes used to transfer the MIB, SIB1 and four SI messages. The example uses an SI-window of length 10 subframes, and shows a higher number of 'blind' Hybrid ARQ (HARQ) transmissions[9] being used for the larger SI messages.

SI messages may have different periodicities. Consequently, in some clusters of SI-windows all the SI messages are scheduled, while in other clusters only the SIs with shorter repetition periods are transmitted. For the example of Table 3.1, the cluster of SI-windows beginning at System Frame Number (SFN) 0 contains all the SI messages, the cluster starting at SFN160 contains only the first SI message, that beginning at SFN320 contains the first and second SI messages, and the one starting at SFN480 contains only the first SI message.

Note that Figure 3.2 shows a cluster of SI-windows where all the SI messages are transmitted. At occasions where a given SI is not transmitted (due to a longer repetition period), its corresponding SI-window is not used.

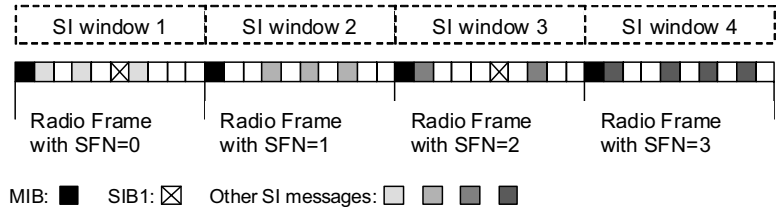


Figure 3.2: SI scheduling example.

3.2.2.2 Validity and Change Notification

SI normally changes only at specific radio frames whose System Frame Number is given by SFN mod $N = 0$, where N is configurable and defines the period between two radio frames at which a change may occur, known as the *modification period*. Prior to performing a change

[7] The Physical Downlink Control Channel – PDCCH; see Section 9.3.5.
[8] A subframe in LTE has a duration of 1 ms; see Section 6.2
[9] With blind HARQ retransmissions, there is no feedback to indicate whether the reception has been successful.

of the system information, the E-UTRAN notifies the UEs by means of a *Paging* message including a *SystemInfoModification* flag. Figure 3.3 illustrates the change of SI, with different shading indicating different content.

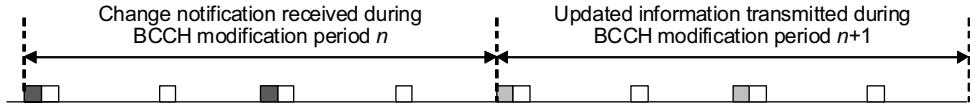


Figure 3.3: SI modification periods. Reproduced by permission of © 3GPP.

LTE provides two mechanisms for indicating that SI has changed:

1. A paging message including a flag indicating whether or not SI has changed.
2. A value tag in SIB1 which is incremented every time one or more SI message changes.

UEs in RRC_IDLE use the first mechanism, while UEs in RRC_CONNECTED can use either mechanism; the second being useful, for example, in cases when a UE was unable to receive the paging messages.

UEs in RRC_IDLE are only required to receive the paging message at their normal paging occasions – i.e. no additional wake-ups are expected to detect changes of SI. In order to ensure reliability of reception, the change notification paging message is normally repeated a number of times during the BCCH modification period preceding that in which the new system information is first transmitted. Correspondingly, the modification period is expressed as a multiple of the cell-specific default paging cycle.

UEs in RRC_CONNECTED are expected to try receiving a paging message the same number of times per modification period as UEs in RRC_IDLE using the default paging cycle. The exact times at which UEs in RRC_CONNECTED which are using this method have to try to receive a paging message are not specified; the UE may perform these tries at convenient times, such as upon wake-up from DRX, using any of the subframes which are configured for paging during the modification period. Since the eNodeB anyway has to notify all the UEs in RRC_IDLE, it has to send a paging message in all subframes which are configured for paging (up to a maximum of four subframes per radio frame) during an entire modification period. Connected mode UEs can utilize any of these subframes. The overhead of transmitting paging messages to notify UEs of a change of SI is considered marginal, since such changes are expected to be infrequent – at most once every few hours.

If the UE receives a notification of a change of SI, it starts acquiring SI from the start of the next modification period. Until the UE has successfully acquired the updated SI, it continues to use the existing parameters. If a critical parameter changes, the communication may be seriously affected, but any service interruption that may result is considered acceptable since it is short and infrequent.

If the UE returns to a cell, it is allowed to assume that the SI previously acquired from the cell remains valid if it was received less than 3 hours previously and the value tag matches.

3.2.3 Connection Control within LTE

Connection control involves:

- Security activation;
- Connection establishment, modification and release;
- DRB establishment, modification and release;
- Mobility within LTE.

3.2.3.1 Security Key Management

Security is a very important feature of all 3GPP RATs. LTE provides security in a similar way to its predecessors UMTS and GSM.

Two functions are provided for the maintenance of security: *ciphering* of both control plane (RRC) data (i.e. SRBs 1 and 2) and user plane data (i.e. all DRBs), and *integrity protection* which is used for control plane (RRC) data only. Ciphering is used in order to protect the data streams from being received by a third party, while integrity protection allows the receiver to detect packet insertion or replacement. RRC always activates both functions together, either following connection establishment or as part of the handover to LTE.

The hierarchy of keys by which the AS security keys are generated is illustrated in Figure 3.4. The process is based on a common secret key K_{ASME} (Access Security Management Entity) which is available only in the Authentication Centre in the Home Subscriber Server (HSS) (see Section 2.2.1) and in a secure part of the Universal Subscriber Identity Module (USIM) in the UE. A set of keys and checksums are generated at the Authentication Centre using this secret key and a random number. The generated keys, checksums and random number are transferred to the Mobility Management Entity (MME) (see Section 2.2.1), which passes one of the generated checksums and the random number to the UE. The USIM in the UE then computes the same set of keys using the random number and the secret key. Mutual authentication is performed by verifying the computed checksums in the UE and network using NAS protocols.

Upon connection establishment, the AS derives an *AS base-key* K_{eNB} (eNodeB-specific) and Next Hop (NH), from K_{ASME}.

K_{eNB} is used to generate three further security keys known as the *AS derived-keys*: one, called $K_{RRC\,int}$, is used for integrity protection of the RRC signalling (SRBs), one for ciphering of the RRC signalling known as $K_{RRC\,enc}$ and $K_{UP\,enc}$ used for ciphering of user data (i.e. DRBs).

NH is an intermediate key used to implement 'forward security'[10] [3]. It is derived by the UE and MME using K_{ASME} and K_{eNB} when the security context is established or using K_{ASME} and the previous NH otherwise. NH is associated with a counter called Next hop Chaining Counter (NCC) which is initially set to 0 at connection establishment.

In case of handover within E-UTRAN, a new AS base-key and new AS Derived-keys are computed from the AS base-key used in the source cell. An intermediate key, K_{eNB^*} is derived by the UE and the source eNodeB based on the Physical Cell Identity (PCI) of the target cell,

[10] Forward security refers to the property that, for an eNodeB sharing a K_{eNB} with a UE, it shall be computationally infeasible to predict any future K_{eNB}, that will be used between the same UE and another eNodeB

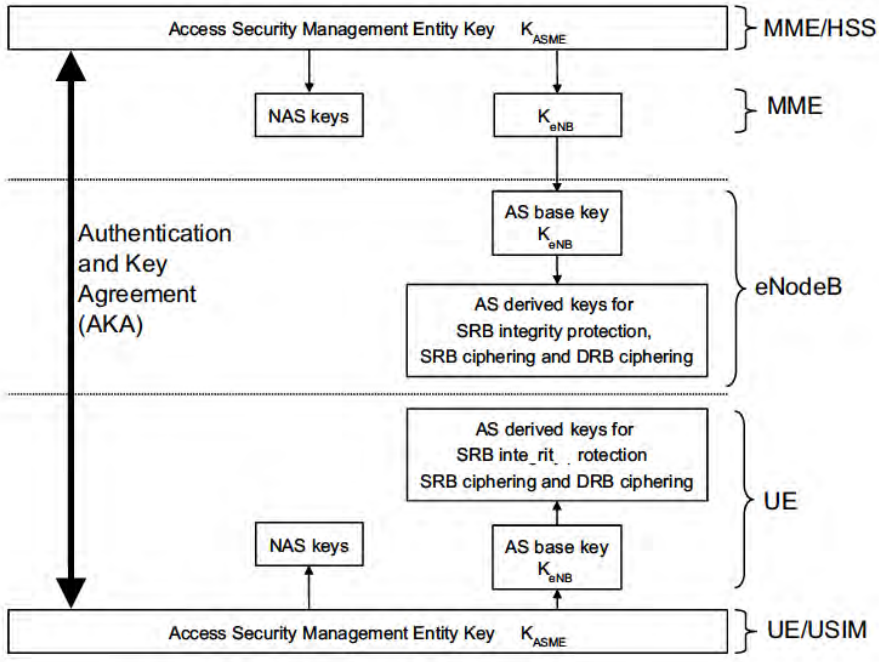


Figure 3.4: Security key derivation.

the target frequency and NH or K_{eNB}. If a fresh NH is available[11], the derivation of K_{eNB^*} is based on NH (referred to as vertical derivation). If no fresh NH is available then the K_{eNB^*} derivation is referred to as horizontal derivation and is based on K_{eNB}. K_{eNB^*} is then used at the target cell as the new K_{eNB} for RRC and data traffic.

For handover to E-UTRAN from UTRAN or GERAN, the AS base-key is derived from integrity and ciphering keys used in the UTRAN or GERAN. Handover within LTE may be used to take a new K_{ASME} into account, i.e. following a re-authentication by NAS.

The use of the security keys for the integrity protection and ciphering functions is handled by the PDCP layer, as described in Section 4.2.3.

The security functions are never deactivated, although it is possible to apply a 'NULL' ciphering algorithm. The 'NULL' algorithm may also be used in certain special cases, such as for making an emergency call without a USIM.

[11] In this case the NCC is incremented and is then larger than that of the currently active K_{eNB^*}

3.2.3.2 Connection Establishment and Release

Two levels of NAS states reflect the state of a UE in respect of connection establishment: the EPS Mobility Management (EMM) state (EMM-DEREGISTERED or EMM-REGISTERED) reflects whether the UE is registered in the MME, and the EPS Connection Management (ECM) state (ECM-IDLE or ECM-CONNECTED) reflects the connectivity of the UE with the Evolved Packet Core (EPC – see Chapter 2).

The NAS states, and their relationship to the AS RRC states, are illustrated in Figure 3.5.

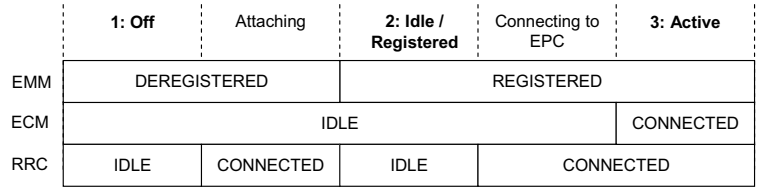


Figure 3.5: Possible combinations of NAS and AS states.

The transition from ECM-IDLE to ECM-CONNECTED not only involves establishment of the RRC connection but also includes establishment of the S1-connection (see Section 2.5). RRC connection establishment is initiated by the NAS and is completed prior to S1-connection establishment, which means that connectivity in RRC_CONNECTED is initially limited to the exchange of control information between UE and E-UTRAN.

UEs are typically moved to ECM-CONNECTED when becoming active. It should be noted, however, that in LTE the transition from ECM-IDLE to ECM-CONNECTED is performed within 100 ms. Hence, UEs engaged in intermittent data transfer need not be kept in ECM-CONNECTED if the ongoing services can tolerate such transfer delays. In any case, an aim in the design of LTE was to support similar battery power consumption levels for UEs in RRC_CONNECTED as for UEs in RRC_IDLE.

RRC connection release is initiated by the eNodeB following release of the S1-connection between the eNodeB and the Core Network (CN).

Connection establishment message sequence. RRC connection establishment involves the establishment of SRB1 and the transfer of the initial uplink NAS message. This NAS message triggers the establishment of the S1-connection, which normally initiates a subsequent step during which E-UTRAN activates AS-security and establishes SRB2 and one or more DRBs (corresponding to the default and optionally dedicated EPS bearers).

Figure 3.6 illustrates the RRC connection establishment procedure, including the subsequent step of initial security activation and radio bearer establishment.

Step 1: Connection establishment

- Upper layers in the UE trigger connection establishment, which may be in response to paging. The UE checks if access is barred (see Section 3.3.4.6). If this is not the case, the lower layers in the UE perform a contention-based random access procedure

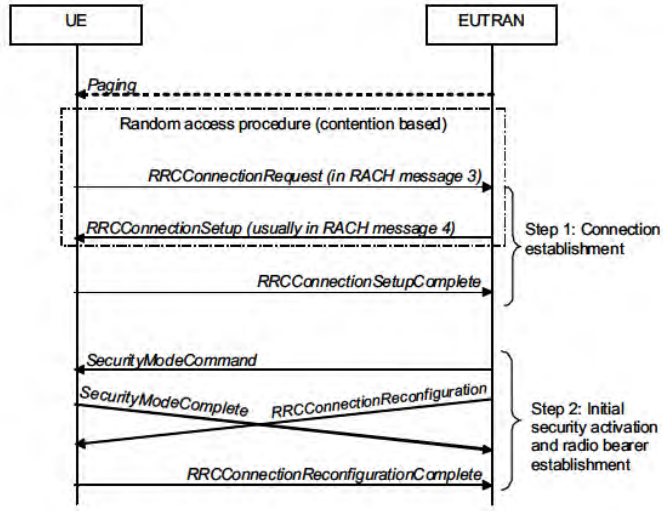


Figure 3.6: Connection establishment (see Section 17.3.1 for details of the contention-based RACH procedure).

as described in Section 17.3, and the UE starts a timer (known as *T300*) and sends the RRCConnectionRequest message. This message includes an initial identity (S-TMSI[12] or a random number) and an establishment cause.

- If E-UTRAN accepts the connection, it returns the RRCConnectionSetup message that includes the initial radio resource configuration including SRB1. Instead of signalling each individual parameter, E-UTRAN may order the UE to apply a default configuration – i.e. a configuration for which the parameter values are specified in the RRC specification [1].
- The UE returns the RRCConnectionSetupComplete message and includes the NAS message, an identifier of the selected PLMN (used to support network sharing) and, if provided by upper layers, an identifier of the registered MME. Based on the last two parameters, the eNodeB decides on the CN node to which it should establish the S1-connection.

Step 2: Initial security activation and radio bearer establishment

- E-UTRAN sends the SecurityModeCommand message to activate integrity protection and ciphering. This message, which is integrity-protected but not ciphered, indicates which algorithms shall be used.
- The UE verifies the integrity protection of the SecurityModeControl message, and, if this succeeds, it configures lower layers to apply integrity protection and ciphering to all subsequent messages (with the exception that ciphering is not applied to the

[12] S-Temporary Mobile Subscriber Identity.

response message, i.e. the SecurityModeComplete (or SecurityModeFailure) message).

- E-UTRAN sends the RRCConnectionReconfiguration message including a radio resource configuration used to establish SRB2 and one or more DRBs. This message may also include other information such as a piggybacked NAS message or a measurement configuration. E-UTRAN may send the RRCConnectionReconfiguration message prior to receiving the SecurityModeComplete message. In this case, E-UTRAN should release the connection when one (or both) procedures fail (because the two procedures result from a single S1-procedure, which does not support partial success).
- The UE finally returns the RRCConnectionReconfigurationComplete message.

A connection establishment may fail for a number of reasons, such as the following:

- Access may be barred (see Section 3.3.4.6).
- In case cell reselection occurs during connection establishment, the UE aborts the procedure and informs upper layers of the failure to establish the connection.
- E-UTRAN may temporarily reject the connection establishment by including a wait timer, in which case the UE rejects any connection establishment request until the wait time has elapsed.
- The NAS may abort an ongoing RRC connection establishment, for example upon NAS timer expiry.

3.2.3.3 DRB Establishment

To establish, modify or release DRBs, E-UTRAN applies the RRC connection reconfiguration procedure as described in Section 3.2.3.2.

When establishing a DRB, E-UTRAN decides how to transfer the packets of an EPS bearer across the radio interface. An EPS bearer is mapped (1-to-1) to a DRB, a DRB is mapped (1-to-1) to a DTCH (Dedicated Traffic CHannel – see Section 4.4.1.2) logical channel, all logical channels are mapped (n-to-1) to the Downlink or Uplink Shared Transport CHannel (DL-SCH or UL-SCH), which are mapped (1-to-1) to the corresponding Physical Downlink or Uplink Shared CHannel (PDSCH or PUSCH). This radio bearer mapping is illustrated in Figure 3.1.

The radio resource configuration covers the configuration of the PDCP, RLC, MAC and physical layers. The main configuration parameters / options include the following:

- For services using small packet sizes (e.g. VoIP), PDCP may be configured to apply indexheader!compressionheader compression to significantly reduce the signalling overhead.
- The RLC Mode is selected from those listed in Section 4.3.1. RLC Acknowledged Mode (AM) is applicable, except for services which require a very low transfer delay and for which reliable transfer is less important.
- E-UTRAN assigns priorities and Prioritized Bit-Rates (PBRs) to control how the UE divides the granted uplink resources between the different radio bearers (see Section 4.4.2.6).

- Unless the transfer delay requirements for any of the ongoing services are very strict, the UE may be configured with a DRX cycle (see Section 4.4.2.5).
- For services involving a semi-static packet rate (e.g. VoIP), semi-persistent scheduling may be configured to reduce the control signalling overhead (see Section 4.4.2.1). Specific resources may also be configured for reporting buffer status and radio link quality.
- Services tolerating higher transfer delays may be configured with a HARQ profile involving a higher average number of HARQ transmissions.

3.2.3.4 Mobility Control in RRC_IDLE and RRC_CONNECTED

Mobility control in RRC_IDLE is UE-controlled (cell-reselection), while in RRC_CONNECTED it is controlled by the E-UTRAN (handover). However, the mechanisms used in the two states need to be consistent so as to avoid ping-pong (i.e. rapid handing back and forth) between cells upon state transitions. The mobility mechanisms are designed to support a wide variety of scenarios including network sharing, country borders, home deployment and varying cell ranges and subscriber densities; an operator may, for example, deploy its own radio access network in populated areas and make use of another operator's network in rural areas.

If a UE were to access a cell which does not have the best radio link quality of the available cells on a given frequency, it may create significant interference to the other cells. Hence, as for most technologies, radio link quality is the primary criterion for selecting a cell on an LTE frequency. When choosing between cells on different frequencies or RATs the interference concern does not apply. Hence, for inter-frequency and inter-RAT cell reselection other criteria may be considered such as UE capability, subscriber type and call type. As an example, UEs with no (or limited) capability for data transmission may be preferably handled on GSM, while home customers or 'premium subscribers' might be given preferential access to the frequency or RAT supporting the highest data rates. Furthermore, in some LTE deployment scenarios, voice services may initially be provided by a legacy RAT only (as a Circuit Switching (CS) application), in which case the UE needs to be moved to the legacy RAT upon establishing a voice call (also referred to as *CS FallBack* (CSFB)).

E-UTRAN provides a list of neighbouring frequencies and cells which the UE should consider for cell reselection and for reporting of measurements. In general, such a list is referred to as a *white-list* if the UE is to consider only the listed frequencies or cells – i.e. other frequencies or cells are not available; conversely, in the case of a *black-list* being provided, a UE may consider any *un*listed frequencies or cells. In LTE, white-listing is used to indicate all the neighbouring frequencies of each RAT that the UE is to consider. On the other hand, E-UTRAN is not required to indicate all the neighbouring cells that the UE shall consider. Which cells the UE is required to detect by itself depends on the UE state as well as on the RAT, as explained below.

Note that for GERAN, typically no information is provided about individual cells. Only in specific cases, such as at country borders, is signalling[13] provided to indicate the group of cells that the UE is to consider – i.e. a white cell list.

[13] The 'NCC-permitted' parameter – see GERAN specifications.

Mobility in idle mode. In RRC_IDLE, cell reselection between frequencies is based on absolute priorities, where each frequency has an associated priority. Cell-specific default values of the priorities are provided via SI. In addition, E-UTRAN may assign UE-specific values upon connection release, taking into account factors such as UE capability or subscriber type. In case equal priorities are assigned to multiple cells, the cells are ranked based on radio link quality. Equal priorities are not applicable between frequencies of different RATs. The UE does not consider frequencies for which it does not have an associated priority; this is useful in situations such as when a neighbouring frequency is applicable only for UEs of one of the sharing networks.

Table 3.2 provides an overview of the SI parameters which E-UTRAN may use to control cell reselection . Other than the cell reselection priority of a frequency, no idle mode mobility-related parameters may be assigned via dedicated signalling. Further details of the parameters listed are provided in Section 3.3.

Mobility in connected mode. In RRC_CONNECTED, the E-UTRAN decides to which cell a UE should hand over in order to maintain the radio link. As with RRC_IDLE, E-UTRAN may take into account not only the radio link quality but also factors such as UE capability, subscriber type and access restrictions. Although E-UTRAN may trigger handover without measurement information (blind handover), normally it configures the UE to report measurements of the candidate target cells – see Section 22.3. Table 3.3 provides an overview of the frequency- and cell-specific parameters which E-UTRAN can configure for mobility-related measurement reporting.

In LTE the UE always connects to a single cell only – in other words, the switching of a UE's connection from a source cell to a target cell is a *hard* handover. The hard handover process is normally a 'backward' one, whereby the eNodeB which controls the source cell requests the target eNodeB to prepare for the handover. The target eNodeB subsequently generates the RRC message to order the UE to perform the handover, and the message is transparently forwarded by the source eNodeB to the UE. LTE also supports a kind of 'forward' handover, in which the UE by itself decides to connect to the target cell, where it then requests that the connection be continued. The UE applies this connection re-establishment procedure only after loss of the connection to the source cell; the procedure only succeeds if the target cell has been prepared in advance for the handover.

Besides the handover procedure, LTE also provides for a UE to be redirected to another frequency or RAT upon connection release. Redirection during connection establishment is not supported, since at that time the E-UTRAN may not yet be in possession of all the relevant information such as the capabilities of the UE and the type of subscriber (as may be reflected, for example, by the SPID, the Subscriber Profile ID for RAT/Frequency Priority). However, the redirection may be performed while AS-security has not (yet) been activated. When redirecting the UE to UTRAN or GERAN, E-UTRAN may provide SI for one or more cells on the relevant frequency. If the UE selects one of the cells for which SI is provided, it does not need to acquire it.

Message sequence for handover within LTE. In RRC_CONNECTED, the E-UTRAN controls mobility by ordering the UE to perform handover to another cell, which may be

Table 3.2: List of SI parameters which may be used to control cell reselection.

Parameter	Intra-Freq.	Inter-Freq.	UTRA	GERAN	CDMA2000
Common	(SIB3)	(SIB5)	(SIB6)	(SIB7)	(SIB8)
Reselection info	Q-Hyst MobilityStatePars Q-HystSF S-Search[a]		T-Reselect T-ReselectSF	T-Reselect T-ReselectSF	T-Reselect T-ReselectSF
Frequency list	(SIB3)	(SIB5)	(SIB6)	(SIB7)	(SIB8)
White frequency list	n/a	+	+	+	+
Frequency specific reselection info[b]	Priority $\text{Thresh}_{\text{Serving-Low}}$ $\text{Thresh}_{\text{Serving-LowQ}}$ T-Reselect T-ReselectSF $\text{Thresh}_{\text{X-LowQ}}$ T-Reselect T-ReselectSF	Priority Qoffset, $\text{Thresh}_{\text{X-High}}$, $\text{Thresh}_{\text{X-Low}}$ $\text{Thresh}_{\text{X-HighQ}}$,	Priority $\text{Thresh}_{\text{X-High}}$, $\text{Thresh}_{\text{X-Low}}$ $\text{Thresh}_{\text{X-HighQ}}$, $\text{Thresh}_{\text{X-LowQ}}$	Priority $\text{Thresh}_{\text{X-High}}$, $\text{Thresh}_{\text{X-Low}}$	Priority $\text{Thresh}_{\text{X-High}}$, $\text{Thresh}_{\text{X-Low}}$
Frequency specific suitability info[c]	Q-RxLevMin MaxTxPower Q-QualMin	Q-RxLevMin MaxTxPower Q-QualMin	Q-RxLevMin, MaxTxPower, Q-QualMin	Q-RxLevMin MaxTxPower	
Cell list	(SIB4)	(SIB5)	(SIB6)	(SIB7)	(SIB8)
White cell list	–	–	–	NCC permitted[d]	–
Black cell list	+	+	–	–	–
List of cells with specific info[e]	Qoffset	Qoffset	–	–	–

[a] Separate parameters for intra/ inter-frequency, both for RSRP and RSRQ.
[b] See Section 3.3.4.2.
[c] See Section 3.3.3.
[d] See GERAN specifications.
[e] See Section 3.3.4.3.

on the same frequency (‘intra-frequency’) or a different frequency (‘inter-frequency’). Inter-frequency measurements may require the configuration of measurement gaps, depending on the capabilities of the UE (e.g. whether it has a dual receiver) – see Section 22.3.

The E-UTRAN may also use the handover procedures for completely different purposes, such as to change the security keys to a new set (see Section 3.2.3.1), or to perform a ‘synchronized reconfiguration’ in which the E-UTRAN and the UE apply the new configuration simultaneously.

The message sequence for the procedure for handover within LTE is shown in Figure 3.7. The sequence is as follows:

1. The UE may send a MeasurementReport message (see Section 3.2.5).

Table 3.3: Frequency- and cell-specific information which can be configured in connected mode.

Parameter	Intra-Freq.	Inter-Freq.	UTRA	GERAN	CDMA2000
Frequency list					
White frequency list	n/a	+	+	+	+
Frequency specific info	Qoffset	Qoffset	Qoffset	Qoffset	Qoffset
Cell list					
White cell list	–	–	+	NCC permitted	+
Black cell list	+	+	–	–	–
List of cells with specific info.	Qoffset	Qoffset	–	–	–

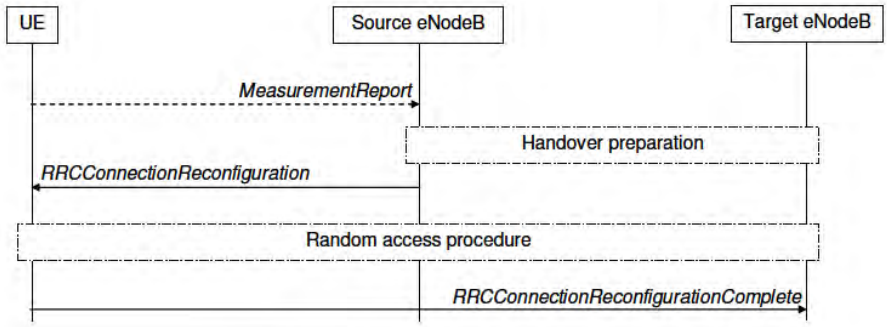


Figure 3.7: Handover within LTE.

2. Before sending the handover command to the UE, the source eNodeB requests one or more target cells to prepare for the handover. As part of this 'handover preparation request', the source eNodeB provides UE RRC context information[14] about the UE capabilities, the current AS-configuration and UE-specific Radio Resource Management (RRM) information. In response, the eNodeB controlling the target cell generates the 'handover command'. The source eNodeB will forward this command to the UE in the RRCConnectionReconfiguration message. This is done transparently (apart from performing integrity protection and ciphering) – i.e. the source eNodeB does not add or modify the protocol information contained in the message.

3. The source eNodeB sends the RRCConnectionReconfiguration message which to the UE orders it to perform handover. It includes mobility control information (namely the identity, and optionally the frequency, of the target cell) and the radio resource

[14]This UE context information includes the radio resource configuration including local settings not configured across the radio interface, UE capabilities and radio resource management information.

configuration information which is common to all UEs in the target cell (e.g. information required to perform random access. The message also includes the dedicated radio resource configuration, the security configuration and the C-RNTI[15] to be used in the target cell. Although the message may optionally include the measurement configuration, the E-UTRAN is likely to use another reconfiguration procedure for re-activating measurements, in order to avoid the RRCConnectionReconfiguration message becoming excessively large. If no measurement configuration information is included in the message used to perform inter-frequency handover, the UE stops any inter-frequency and inter-RAT measurements and deactivates the measurement gap configuration.

4. If the UE is able to comply with the configuration included in the received RRCConnectionReconfiguration message, the UE starts a timer, known as *T304*, and initiates a random access procedure (see Section 17.3), using the received Random Access CHannel (RACH) configuration, to the target cell at the first available occasion.[16] It is important to note that the UE does not need to acquire system information from the target cell prior to initiating random access and resuming data communication. However, the UE may be unable to use some parts of the physical layer configuration from the very start (e.g. semi-persistent scheduling (see Section 4.4.2.1), the PUCCH (see Section 16.3) and the Sounding Reference Signal (SRS) (see Section 15.6)). The UE derives new security keys and applies the received configuration in the target cell.

5. Upon successful completion of the random access procedure, the UE stops the timer T304. The AS informs the upper layers in the UE about any uplink NAS messages for which transmission may not have completed successfully, so that the NAS can take appropriate action.

For handover to cells broadcasting a Closed Subscriber Group (CSG) identity, normal measurement and mobility procedures are used to support handover. In addition, E-UTRAN may configure the UE to report that it is entering or leaving the proximity of cell(s) included in its CSG whitelist. Furthermore, E-UTRAN may request the UE to provide additional information broadcast by the handover candidate cell, for example the cell global identity, CSG identity or CSG membership status. E-UTRAN may use a indexproximity report‘proximity report’ to configure measurements and to decide whether or not to request the UE to provide additional information broadcast by the handover candidate cell. The additional information is used to verify whether or not the UE is authorized to access the target cell and may also be needed to identify handover candidate cells.[17] Further details of the mobility procedures for HeNBs can be found in Section 24.2.3.

[15]The Cell Radio Network Temporary Identifier is the RNTI to be used by a given UE while it is in a particular cell.

[16]The target cell does not specify when the UE is to initiate random access in that cell. Hence, the handover process is sometimes described as *asynchronous*.

[17]This may be the case if PCI confusion occurs, i.e. when the PCI that is included in the measurement report does not uniquely identify the cell.

3.2.3.5 Connection Re-Establishment Procedure

In a number of failure cases (e.g. radio link failure, handover failure, RLC unrecoverable error, reconfiguration compliance failure), the UE initiates the RRC connection re-establishment procedure, provided that security is active. If security is not active when one of the indicated failures occurs, the UE moves to RRC_IDLE instead.

To attempt RRC connection re-establishment, the UE starts a timer known as *T311* and performs cell selection. The UE should prioritize searching on LTE frequencies. However, no requirements are specified regarding for how long the UE shall refrain from searching for other RATs. Upon finding a suitable cell on an LTE frequency, the UE stops the timer T311, starts the timer T301 and initiates a contention based random access procedure to enable the RRCConnectionReestablishmentRequest message to be sent. In the RRCConnectionReestablishmentRequest message, the UE includes the identity used in the cell in which the failure occurred, the identity of that cell, a short Message Authentication Code and a cause.

The E-UTRAN uses the re-establishment procedure to continue SRB1 and to re-activate security without changing algorithms. A subsequent RRC connection reconfiguration procedure is used to resume operation on radio bearers other than SRB1 and to re-activate measurements. If the cell in which the UE initiates the re-establishment is not prepared (i.e. does not have a context for that UE), the E-UTRAN will reject the procedure, causing the UE to move to RRC_IDLE.

3.2.4 Connected Mode Inter-RAT Mobility

The overall procedure for the control of mobility is explained in this section; some further details can be found in Chapter 22.

3.2.4.1 Handover to LTE

The procedure for handover to LTE is largely the same as the procedure for handover within LTE, so it is not necessary to repeat the details here. The main difference is that upon handover to LTE the entire AS-configuration needs to be signalled, whereas within LTE it is possible to use 'delta signalling', whereby only the changes to the configuration are signalled.

If ciphering had not yet been activated in the previous RAT, the E-UTRAN activates ciphering, possibly using the NULL algorithm, as part of the handover procedure. The E-UTRAN also establishes SRB1, SRB2 and one or more DRBs (i.e. at least the DRB associated with the default EPS bearer).

3.2.4.2 Mobility from LTE

Generally, the procedure for mobility from LTE to another RAT supports both handover and Cell Change Order (CCO), possibly with Network Assistance (NACC – Network Assisted Cell Change). The CCO/NACC procedure is applicable only for mobility to GERAN. Mobility from LTE is performed only after security has been activated. When used for enhanced CSFB[18] to CDMA2000, the procedure includes support for parallel handover (i.e.

[18] See Section 2.4.2.1.

to both 1XRTT and HRPD), for handover to 1XRTT in combination with redirection to HRPD, and for redirection to HRPD only.

The procedure is illustrated in Figure 3.8.

1. The UE may send a MeasurementReport message (see Section 3.2.5 for further details).
2. In case of handover (as opposed to CCO), the source eNodeB requests the target Radio Access Network (RAN) node to prepare for the handover. As part of the 'handover preparation request' the source eNodeB provides information about the applicable inter-RAT UE capabilities as well as information about the currently-established bearers. In response, the target RAN generates the 'handover command' and returns this to the source eNodeB.
3. The source eNodeB sends a MobilityFromEUTRACommand message to the UE, which includes either the inter-RAT message received from the target (in case of handover), or the target cell/frequency and a few inter-RAT parameters (in case of CCO).
4. Upon receiving the MobilityFromEUTRACommand message, the UE starts the timer T304 and connects to the target node, either by using the received radio configuration (handover) or by initiating connection establishment (CCO) in accordance with the applicable specifications of the target RAT.

Upper layers in the UE are informed, by the AS of the target RAT, which bearers are established. From this, the UE can derive if some of the established bearers were not admitted by the target RAN node.

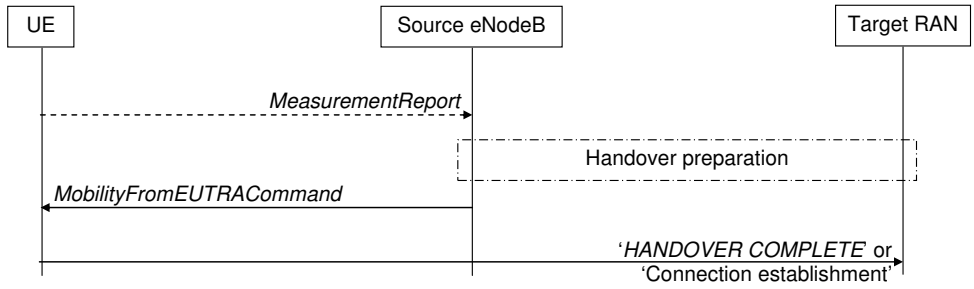


Figure 3.8: Mobility from LTE.

3.2.4.3 CDMA2000

For CDMA2000, additional procedures have been defined to support the transfer of dedicated information from the CDMA2000 upper layers, which are used to register the UE's presence in the target core network prior to performing the handover (referred to as preregistration). These procedures use SRB1.

3.2.5 Measurements

3.2.5.1 Measurement Configuration

The E-UTRAN can configure the UE to report measurement information to support the control of UE mobility. The following measurement configuration elements can be signalled via the RRCConnectionReconfiguration message.

1. **Measurement objects.** A measurement object defines on what the UE should perform the measurements – such as a carrier frequency. The measurement object may include a list of cells to be considered (white-list or black-list) as well as associated parameters, e.g. frequency- or cell-specific offsets.

2. **Reporting configurations.** A reporting configuration consists of the (periodic or event-triggered) criteria which cause the UE to send a measurement report, as well as the details of what information the UE is expected to report (e.g. the quantities, such as Received Signal Code Power (RSCP) (see Section 22.3.2.1) for UMTS or Reference Signal Received Power (RSRP) (see Section 22.3.1.1) for LTE, and the number of cells).

3. **Measurement identities.** These identify a measurement and define the applicable measurement object and reporting configuration.

4. **Quantity configurations**. The quantity configuration defines the filtering to be used on each measurement.

5. **Measurement gaps.** Measurement gaps define time periods when no uplink or downlink transmissions will be scheduled, so that the UE may perform the measurements. The measurement gaps are common for all gap-assisted measurements. Further details of the measurement gaps are discussed in Section 22.2.1.2.

The details of the above parameters depend on whether the measurement relates to an LTE, UMTS, GERAN or CDMA2000 frequency. Further details of the measurements performed by the UE are explained in Section 22.3. The E-UTRAN configures only a single measurement object for a given frequency, but more than one measurement identity may use the same measurement object. The identifiers used for the measurement object and reporting configuration are unique across all measurement types. An example of a set of measurement objects and their corresponding reporting configurations is shown in Figure 3.9.

In LTE it is possible to configure the quantity which triggers the report (RSCP or RSRP) for each reporting configuration. The UE may be configured to report either the trigger quantity or both quantities.

The RRC measurement reporting procedures include some extensions specifically to support Self-Optimizing Network (SON) functions such as the determination of Automatic Neighbour Relations (ANR) – see Section 25.2. The RRC measurement procedures also support UE positioning[19] by means of the enhanced cell identity method – see Section 19.4.

[19] See Chapter 19.

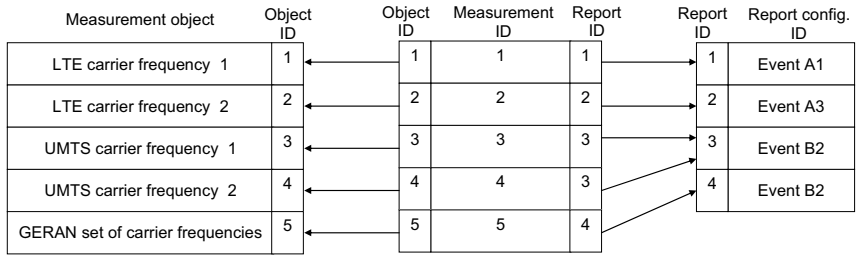


Figure 3.9: Example measurement configuration.

3.2.5.2 Measurement Report Triggering

Depending on the measurement type, the UE may measure and report any of the following:

- The serving cell;
- Listed cells (i.e. cells indicated as part of the measurement object);
- Detected cells on a listed frequency (i.e. cells which are not listed cells but are detected by the UE).

For some RATs, the UE measures and reports listed cells only (i.e. the list is a white-list), while for other RATs the UE also reports detected cells. For further details, see Table 3.3. Additionally, E-UTRAN can configure UTRAN PCI ranges for which the UE is allowed to send a measurement reports (mainly for the support of handover to UTRAN cells broadcasting a CSG identity).

For LTE, the following event-triggered reporting criteria are specified:

- **Event A1.** Serving cell becomes better than absolute threshold.
- **Event A2.** Serving cell becomes worse than absolute threshold.
- **Event A3.** Neighbour cell becomes better than an offset relative to the serving cell.
- **Event A4.** Neighbour cell becomes better than absolute threshold.
- **Event A5.** Serving cell becomes worse than one absolute threshold and neighbour cell becomes better than another absolute threshold.

For inter-RAT mobility, the following event-triggered reporting criteria are specified:

- **Event B1.** Neighbour cell becomes better than absolute threshold.
- **Event B2.** Serving cell becomes worse than one absolute threshold and neighbour cell becomes better than another absolute threshold.

The UE triggers an event when one or more cells meets a specified ‘entry condition’. The E-UTRAN can influence the entry condition by setting the value of some configurable parameters used in these conditions – for example, one or more thresholds, an offset, and/or a hysteresis. The entry condition must be met for at least a duration corresponding to a ‘timeToTrigger’ parameter configured by the E-UTRAN in order for the event to be triggered.

The UE scales the timeToTrigger parameter depending on its speed (see Section 3.3 for further detail).

Figure 3.10 illustrates the triggering of event A3 when a timeToTrigger and an offset are configured.

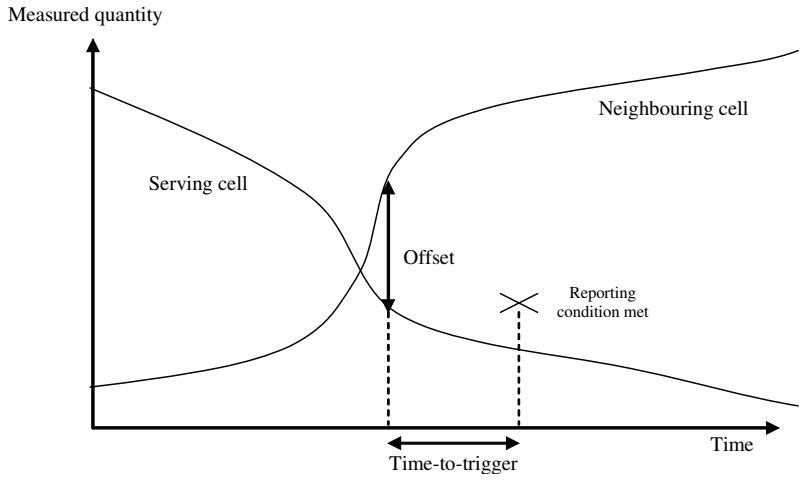


Figure 3.10: Event triggered report condition (Event A3).

The UE may be configured to provide a number of periodic reports after having triggered an event. This 'event-triggered periodic reporting' is configured by means of parameters 'reportAmount' and 'reportInterval', which specify respectively the number of periodic reports and the time period between them. If event-triggered periodic reporting is configured, the UE's count of the number of reports sent is reset to zero whenever a new cell meets the entry condition. The same cell cannot then trigger a new set of periodic reports unless it first meets a specified 'leaving condition'.

In addition to event-triggered reporting, the UE may be configured to perform periodic measurement reporting. In this case, the same parameters may be configured as for event-triggered reporting, except that the UE starts reporting immediately rather than only after the occurrence of an event.

3.2.5.3 Measurement Reporting

In a MeasurementReport message, the UE only includes measurement results related to a single measurement – in other words, measurements are not combined for reporting purposes. If multiple cells triggered the report, the UE includes the cells in order of decreasing value of the reporting quantity – i.e. the best cell is reported first. The number of cells the UE includes in a MeasurementReport may be limited by a parameter 'indexmaxReportCellsmaxReportCells'.

3.2.6 Other RRC Signalling Aspects

3.2.6.1 UE Capability Transfer

In order to avoid signalling of the UE radio access capabilities across the radio interface upon each transition from RRC_IDLE to RRC_CONNECTED, the core network stores the AS capabilities (both the E-UTRA and GERAN capabilities) while the UE is in RRC_IDLE/EMM-REGISTERED. Upon S1 connection establishment, the core network provides the capabilities to the E-UTRAN. If the E-UTRAN does not receive the (required) capabilities from the core network (e.g. due to the UE being in EMM-DEREGISTERED), it requests the UE to provide its capabilities using the UE capability transfer procedure. The E-UTRAN can indicate for each RAT (LTE, UMTS, GERAN) whether it wants to receive the associated capabilities. The UE provides the requested capabilities using a separate container for each RAT. Dynamic change of UE capabilities is not supported, except for change of the GERAN capabilities in RRC_IDLE which is supported by the tracking area update procedure.

3.2.6.2 Uplink/Downlink Information Transfer

The uplink/downlink information transfer procedures are used to transfer only upper layer information (i.e. no RRC control information is included). The procedure supports the transfer of 3GPP NAS dedicated information as well as CDMA2000 dedicated information.

In order to reduce latency, NAS information may also be included in the RRCConnection-SetupComplete and RRCConnectionReconfiguration messages. For the latter message, NAS information is only included if the AS and NAS procedures are dependent (i.e. they jointly succeed or fail). This applies for EPS bearer establishment, modification and release.

As noted earlier, some additional NAS information transfer procedures have also been defined for CDMA2000 for preregistration.

3.2.6.3 UE Information Transfer

The UE information transfer procedure was introduced in Release 9 to support SON (see Chapter 25). The procedure supports network optimization for mobility robustness by the reporting, at a later point in time, of measurement information available when a radio link failure occurs (see Section 25.6). E-UTRAN may also use the UE information transfer procedure to retrieve information regarding the last successful random access, which it may use for RACH optimization – see Section 25.7.

3.3 PLMN and Cell Selection

3.3.1 Introduction

After a UE has selected a PLMN, it performs *cell selection* – in other words, it searches for a suitable cell on which to camp (see Chapter 7). While camping on the chosen cell, the UE acquires the SI that is broadcast (see Section 9.2.1). Subsequently, the UE registers its presence in the tracking area, after which it can receive paging information which is used

to notify UEs of incoming calls. The UE may establish an RRC connection, for example to establish a call or to perform a tracking area update.

When camped on a cell, the UE regularly verifies if there is a better cell; this is known as performing *cell reselection*.

LTE cells are classified according to the service level the UE obtains on them: a *suitable cell* is a cell on which the UE obtains normal service. If the UE is unable to find a suitable cell, but manages to camp on a cell belonging to another PLMN, the cell is said to be an *acceptable cell*, and the UE enters a 'limited service' state in which it can only perform emergency calls (and receive public warning messages) – as is also the case when no USIM is present in the UE. Finally, some cells may indicate via their SI that they are barred or reserved; a UE can obtain no service on such a cell.

A category called 'operator service' is also supported in LTE, which provides normal service but is applicable only for UEs with special access rights.

Figure 3.11 provides a high-level overview of the states and the cell (re)selection procedures.

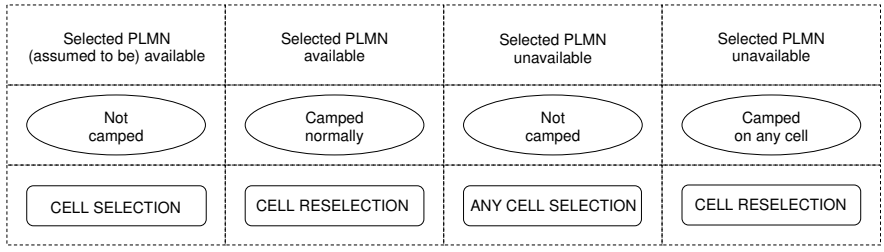


Figure 3.11: Idle mode states and procedures.

3.3.2 PLMN Selection

The NAS handles PLMN selection based on a list of available PLMNs provided by the AS. The NAS indicates the selected PLMN together with a list of equivalent PLMNs, if available. After successful registration, the selected PLMN becomes the *Registered* PLMN (R-PLMN).

The AS may autonomously indicate available PLMNs. In addition, NAS may request the AS to perform a full search for available PLMNs. In the latter case, the UE searches for the strongest cell on each carrier frequency. For these cells, the UE retrieves the PLMN identities from SI. If the quality of a cell satisfies a defined radio criterion, the corresponding PLMNs are marked as *high quality*; otherwise, the PLMNs are reported together with their quality.

3.3.3 Cell Selection

Cell selection consists of the UE searching for the strongest cell on all supported carrier frequencies of each supported RAT until it finds a suitable cell. The main requirement for cell selection is that it should not take too long, which becomes more challenging with the ever increasing number of frequencies and RATs to be searched. The NAS can speed up the

search process by indicating the RATs associated with the selected PLMN. In addition, the UE may use information stored from a previous access.

The cell selection criterion is known as the *S-criterion* and is fulfilled when the cell-selection receive level and the quality level are above a given value: Srxlev > 0 and Squal > 0, where

$$\text{Srxlev} = Q_{\text{rxlevmeas}} - (Q_{\text{rxlevmin}} - Q_{\text{rxlevminoffset}})$$

$$\text{Squal} = Q_{\text{qualmeas}}(Q_{\text{qualmin}} + Q_{\text{qualminoffset}})$$

in which $Q_{\text{rxlevmeas}}$ is the measured cell receive level value, also known as the RSRP (see Section 22.3.1.1), and Q_{rxlevmin} is the minimum required receive level in the cell. Q_{qualmeas} and Q_{qualmin} are the corresponding parameters for the quality level, also known as the RSRQ.

$Q_{\text{rxlevminoffset}}$ and $Q_{\text{qualminoffset}}$ are offsets which may be configured to prevent ping-pong between PLMNs, which may otherwise occur due to fluctuating radio conditions. The offsets are taken into account only when performing a periodic search for a higher priority PLMN while camped on a suitable cell in a visited PLMN.

The cell selection related parameters are broadcast within the SIB1 message.

For some specific cases, additional requirements are defined:

- Upon leaving connected mode, the UE should normally attempt to select the cell to which it was connected. However, the connection release message may include information directing the UE to search for a cell on a particular frequency.
- When performing 'any cell selection', the UE tries to find an acceptable cell of any PLMN by searching all supported frequencies on all supported RATs. The UE may stop searching upon finding a cell that meets the 'high quality' criterion applicable for that RAT.

Note that the UE only verifies the suitability of the strongest cell on a given frequency. In order to avoid the UE needing to acquire SI from a candidate cell that does not meet the S-criterion, suitability information is provided for inter-RAT neighbouring cells.

3.3.4 Cell Reselection

Once the UE camps on a suitable cell, it starts cell reselection. This process aims to move the UE to the 'best' cell of the selected PLMN and of its equivalent PLMNs, if any. As described in Section 3.2.3.4, cell reselection between frequencies and RATs is primarily based on absolute priorities. Hence, the UE first evaluates the frequencies of all RATs based on their priorities. Secondly, the UE compares the cells on the relevant frequencies based on radio link quality, using a ranking criterion. Finally, upon reselecting to the target cell the UE verifies the cell's accessibility. Further rules have also been defined to allow the UE to limit the frequencies to be measured, to speed up the process and save battery power, as discussed in Section 3.3.4.1. Figure 3.12 provides a high-level overview of the cell reselection procedure.

It should be noted that the UE performs cell reselection only after having camped for at least one second on the current serving cell.

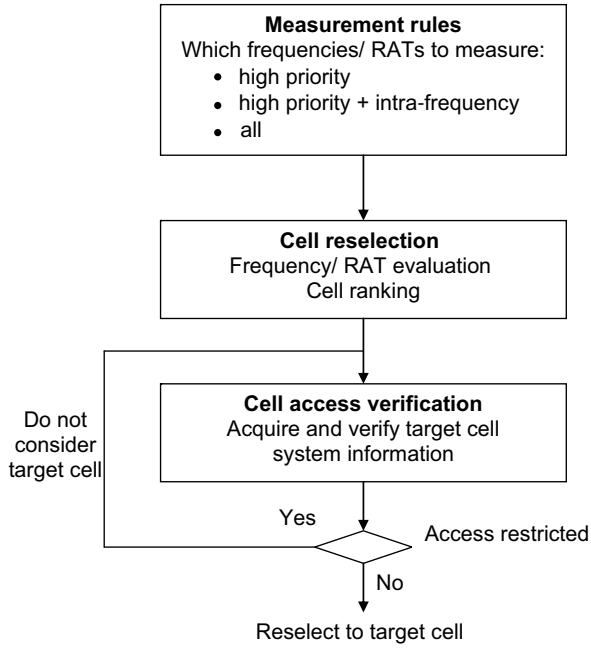


Figure 3.12: Cell reselection.

3.3.4.1 Measurement Rules

To enable the UE to save battery power, rules have been defined which limit the measurements the UE is required to perform. Firstly, the UE is required to perform intra-frequency measurements only when the quality of the serving cell is below or equal to a threshold ('SintraSearch'). Furthermore, the UE is required to measure other frequencies/RATs of lower or equal priority only when the quality of the serving cell is below or equal to another threshold ('SnonintraSearch'). The UE is always required to measure frequencies and RATs of higher priority. For both cases (i.e. intra-frequency and inter-frequency) the UE may refrain from measuring when a receive level and a quality criterion is fulfilled. The required performance (i.e. how often the UE is expected to make the measurements, and to what extent this depends on, for example, the serving cell quality) is specified in [4].

3.3.4.2 Frequency/RAT Evaluation

E-UTRAN configures an absolute priority for all applicable frequencies of each RAT. In addition to the cell-specific priorities which are optionally provided via SI, E-UTRAN can assign UE-specific priorities via dedicated signalling. Of the frequencies that are indicated in the system information, the UE is expected to consider for cell reselection only those for which it has priorities. Equal priorities are not applicable for inter-RAT cell reselection.

The UE reselects to a cell on a higher priority frequency if the S-criterion (see Section 3.3.3) of the concerned target cell exceeds a high threshold ($\text{Thresh}_{\text{X-High}}$) for longer

than a certain duration $T_{reselection}$. The UE reselects to a cell on a lower-priority frequency if the S-criterion of the serving cell is below a low threshold ($Thresh_{Serving\text{-}Low}$) while the S-criterion of the target cell on a lower-priority frequency (possibly on another RAT) exceeds a low threshold ($Thresh_{X\text{-}Low}$) during the time interval $T_{reselection}$, and in the same time no cell on a higher-priority frequency is available. The UE evaluates the thresholds either based on receive level or on quality level, depending on which parameters E-UTRAN configures. Figure 3.13 illustrates the condition(s) to be met for reselecting to a cell on a higher-priority frequency (light grey bar) and to a cell on a lower priority frequency (dark grey bars).

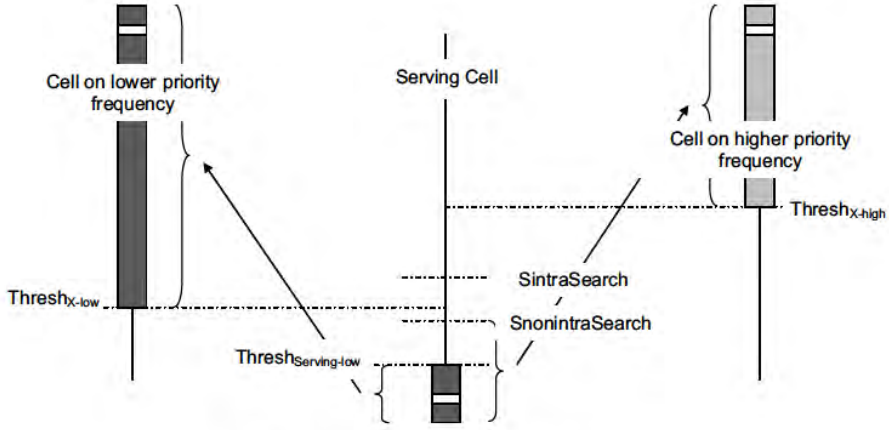


Figure 3.13: Frequency/RAT evaluation.

When reselecting to a frequency, possibly on another RAT, which has a different priority, the UE reselects to the highest-ranked cell on the concerned frequency (see Section 3.3.4.3).

Note that, as indicated in Section 3.2.3.4, thresholds and priorities are configured per frequency, while $T_{reselection}$ is configured per RAT.

From Release 8 onwards, UMTS and GERAN support the same priority-based cell reselection as provided in LTE, with a priority per frequency. Release 8 RANs will continue to handle legacy UEs by means of offset-based ranking. Likewise, Release 8 UEs should apply the ranking based on radio link quality (with offsets) unless UMTS or GERAN indicate support for priority-based reselection.

3.3.4.3 Cell Ranking

The UE ranks the intra-frequency cells and the cells on other frequencies having equal priority which fulfil the S-criterion using a criterion known as the *R-criterion*. The R-criterion generates rankings R_s and R_n for the serving cell and neighbour cells respectively:

$$\text{For the serving cell:} \quad R_s = Q_{meas,s} + Q_{hyst,s}$$

$$\text{For neighbour cells:} \quad R_n = Q_{meas,n} + Q_{off\,s,n}$$

where Q_{meas} is the measured cell received quality (RSRP) (see Section 22.3.1.1), $Q_{hyst,s}$ is a parameter controlling the degree of hysteresis for the ranking, and $Q_{off\,s,n}$ is an offset applicable between serving and neighbouring cells on frequencies of equal priority (the sum of the cell-specific and frequency-specific offsets).

The UE reselects to the highest-ranked candidate cell provided that it is better ranked than the serving cell for at least the duration of $T_{reselection}$. The UE scales the parameters $T_{reselection}$ and Q_{hyst}, depending on the UE speed (see Section 3.3.4.5 below).

3.3.4.4 Accessibility Verification

If the best cell on an LTE frequency is barred or reserved, the UE is required to exclude this cell from the list of cell reselection candidates. In this case, the UE may consider other cells on the same frequency unless the barred cell indicates (by means of field 'intraFreqReselection' within SIB1) that intra-frequency reselection is not allowed for a certain duration, unless the barred cell is an inaccessible Closed Subscriber Group (CSG) cell. If, however, the best cell is unsuitable for some other specific reason (e.g. because it belongs to a forbidden tracking area or to another non-equivalent PLMN), the UE is not permitted to consider any cell on the concerned frequency as a cell reselection candidate for a maximum of 300 s.

3.3.4.5 Speed Dependent Scaling

The UE scales the cell reselection parameters depending on its speed. This applies both in idle mode ($T_{reselection}$ and Q_{hyst}) and in connected mode (timeToTrigger). The UE speed is categorized by a mobility state (high, normal or low), which the UE determines based on the number of cell reselections/handovers which occur within a defined period, excluding consecutive reselections/handovers between the same two cells. The state is determined by comparing the count with thresholds for medium and high state, while applying some hysteresis. For idle and connected modes, separate sets of control parameters are used, signalled in SIB3 and within the measurement configuration respectively.

3.3.4.6 Cell Access Restrictions

The UE performs an access barring check during connection establishment (see Section 3.2.3.2). This function provides a means to control the load introduced by UE-originated traffic. There are separate means for controlling Mobile Originated (MO) calls and MO signalling. On top of the regular access class barring, Service Specific Access Control (SSAC) may be applied. SSAC facilitates separate control for MultiMedia TELephony (MMTEL) voice and video calls. Most of the SSAC functionality is handled by upper layers. In addition, separate access control exists to protect against E-UTRAN overload due to UEs accessing E-UTRAN merely to perform CSFB to CDMA2000.

Each UE belongs to an Access Class (AC) in the range 0–9. In addition, some UEs may belong to one or more high-priority ACs in the range 11–15, which are reserved for specific uses (e.g. security services, public utilities, emergency services, PLMN staff). AC10 is used for emergency access. Further details, for example regarding in which PLMN the high priority ACs apply, are provided in [5]. The UE considers access to be barred if access is barred for all its applicable ACs.

SIB2 may include a set of AC barring parameters for MO calls and/or MO signalling. This set of parameters comprises a probability factor and a barring timer for AC0–9 and a list of barring bits for AC11–15. For AC0–9, if the UE initiates a MO call and the relevant AC barring parameters are included, the UE draws a random number. If this number exceeds the probability factor, access is not barred. Otherwise access is barred for a duration which is randomly selected centred on the broadcast barring timer value. For AC11–15, if the UE initiates a MO call and the relevant AC barring parameters are included, access is barred whenever the bit corresponding to all of the UE's ACs is set. The behaviour is similar in the case of UE-initiated MO signalling.

For cell (re)selection, the UE is expected to consider cells which are neither barred nor reserved for operator or future use. In addition, a UE with an access class in the range 11–15 shall consider a cell that is (only) reserved for operator use and part of its home PLMN (or an equivalent) as a candidate for cell reselection. The UE is never allowed to (re)select a cell that is not a reselection candidate even for emergency access.

3.3.4.7 Any Cell Selection

When the UE is unable to find a suitable cell of the selected PLMN, it performs 'any cell selection'. In this case, the UE performs normal idle mode operation: monitoring paging, acquiring SI, performing cell reselection. In addition, the UE regularly attempts to find a suitable cell on other frequencies or RATs (i.e. not listed in SI). If a UE supporting voice services is unable to find a suitable cell, it should attempt to find an acceptable cell on any supported RAT regardless of the cell reselection priorities that are broadcast. The UE is not allowed to receive MBMS in this state.

3.3.4.8 Closed Subscriber Group

LTE supports the existence of cells which are accessible only for a limited set of UEs – a Closed Subscriber Group (CSG). In order to prevent UEs from attempting to register on a CSG cell on which they do not have access, the UE maintains a CSG white list, i.e. a list of CSG identities for which access has been granted to the UE. The CSG white list can be transferred to the UE by upper layers, or updated upon successful access to a CSG cell. To facilitate the latter, UEs support 'manual selection' of CSG cells which are not in the CSG white list. The manual selection may be requested by the upper layers, based on a text string broadcast by the cell. LTE also supports hybrid cells. Like CSG cells, hybrid cells broadcast a CSG identity; they are accessible as CSG cells by UEs whose CSG white lists include the CSG identity, and as normal cells by all other UEs (see Section 24.2.2).

3.4 Paging

To receive paging messages from E-UTRAN, UEs in idle mode monitor the PDCCH channel for an RNTI value used to indicate paging: the P-RNTI (see Section 9.2.2.2). The UE only needs to monitor the PDCCH channel at certain UE-specific occasions (i.e. at specific subframes within specific radio frames – see Section 6.2 for an introduction to the LTE radio frame structure.). At other times, the UE may apply DRX, meaning that it can switch off its receiver to preserve battery power.

The E-UTRAN configures which of the radio frames and subframes are used for paging. Each cell broadcasts a default paging cycle. In addition, upper layers may use dedicated signalling to configure a UE-specific paging cycle. If both are configured, the UE applies the lowest value. The UE calculates the radio frame (the Paging Frame (PF)) and the subframe within that PF (the Paging Occasion (PO)), which E-UTRAN applies to page the UE as follows:

$$\begin{aligned} \text{SFN mod } T &= (T/N) \times (\text{UE\_ID mod } N) \\ \text{i\_s} &= \lfloor \text{UE\_ID}/N \rfloor \text{ mod Ns} \\ T &= \text{UE DRX cycle (i.e. paging cycle)} = \min(T_c, T_{ue}) \\ N &= \min(T, \text{nB}) \\ \text{Ns} &= \max(1, \text{nB}/T) \end{aligned} \tag{3.1}$$

where:

T_c is the cell-specific default paging cycle {32, 64, 128, 256} radio frames,

T_{ue} is the UE-specific paging cycle {32, 64, 128, 256} radio frames,

N is the number of paging frames within the paging cycle of the UE,

UE_ID is the IMSI[20] mod 1024, with IMSI being the decimal rather than the binary number,

i_s is an index pointing to a pre-defined table defining the corresponding subframe,

nB is the number of 'paging subframes' per paging cycle (across all UEs in the cell),

Ns is the number of 'paging subframes' in a radio frame that is used for paging.

Table 3.4 includes a number of examples to illustrate the calculation of the paging radio frames (PF) and subframes (PO).

Table 3.4: Examples for calculation of paging frames and subframes.

Case	UE_ID	T_c	T_{ue}	T	nB	N	Ns	PF	i_s	PO
A	147	256	256	256	64	64	1	76	0	9
B	147	256	128	128	32	32	1	76	0	9
C	147	256	128	128	256	128	2	19	1	4

In cases A and B in Table 3.4, one out of every four radio frames is used for paging, using one subframe in each of those radio frames. For case B, there are 32 paging frames within the UE's paging cycle, across which the UEs are distributed based on the UE-identity. In case C, two subframes in each radio frame are used for paging, i.e. Ns = 2. In this case, there are 128 paging frames within the UE's paging cycle and the UEs are also distributed across the two subframes within the paging frame. The LTE specifications include a table that indicates the subframe applicable for each combination of Ns and i_s, which is the index that follows from Equation (3.1). Figure 3.14 illustrates cases B and C. All the shaded subframes can be used for paging; the darker ones are applicable for the UE with the indicated identity.

[20]International Mobile Subscriber Identity.

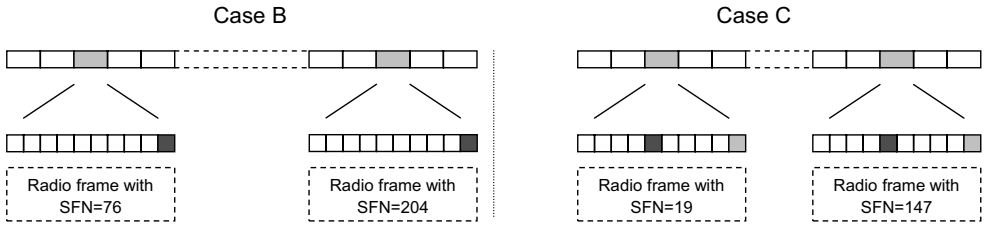


Figure 3.14: Paging frame and paging occasion examples.

3.5 Summary

The main aspects of the Control Plane protocols in LTE can be broken down into the Cell Selection and Reselection Procedures when the UE is in Idle Mode, and the RRC protocol when the UE is in Connected Mode.

The roles of these protocols include supporting security, mobility both between different LTE cells and between LTE and other radio systems, and establishment and reconfiguration of the radio bearers which carry control information and user data.

References[21]

[1] 3GPP Technical Specification 36.331, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification', www.3gpp.org.

[2] 3GPP Technical Specification 36.304, 'Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode (Release 9)', www.3gpp.org.

[3] 3GPP Technical Specification 33.401, '3GPP System Architecture Evolution; Security Architecture', www.3gpp.org.

[4] 3GPP Technical Specification 36.133, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management', www.3gpp.org.

[5] 3GPP Technical Specification 22.011, 'Service accessibility', www.3gpp.org.

[21] All web sites confirmed 1st March 2011.

4

User Plane Protocols

Patrick Fischer, SeungJune Yi, SungDuck Chun and YoungDae Lee

4.1 Introduction to the User Plane Protocol Stack

The LTE Layer 2 user-plane protocol stack is composed of three sublayers, as shown in Figure 4.1:

- **The Packet Data Convergence Protocol (PDCP) layer [1]:** This layer processes Radio Resource Control (RRC) messages in the control plane and Internet Protocol (IP) packets in the user plane. Depending on the radio bearer, the main functions of the PDCP layer are header compression, security (integrity protection and ciphering), and support for reordering and retransmission during handover. For radio bearers which are configured to use the PDCP layer, there is one PDCP entity per radio bearer.
- **The Radio Link Control (RLC) layer [2]:** The main functions of the RLC layer are segmentation and reassembly of upper layer packets in order to adapt them to the size which can actually be transmitted over the radio interface. For radio bearers which need error-free transmission, the RLC layer also performs retransmission to recover from packet losses. Additionally, the RLC layer performs reordering to compensate for out-of-order reception due to Hybrid Automatic Repeat reQuest (HARQ) operation in the layer below. There is one RLC entity per radio bearer.
- **The Medium Access Control (MAC) layer [3]:** This layer performs multiplexing of data from different radio bearers. Therefore there is only one MAC entity per UE. By deciding the amount of data that can be transmitted from each radio bearer and instructing the RLC layer as to the size of packets to provide, the MAC layer aims to achieve the negotiated Quality of Service (QoS) for each radio bearer. For the

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

uplink, this process includes reporting to the eNodeB the amount of buffered data for transmission.

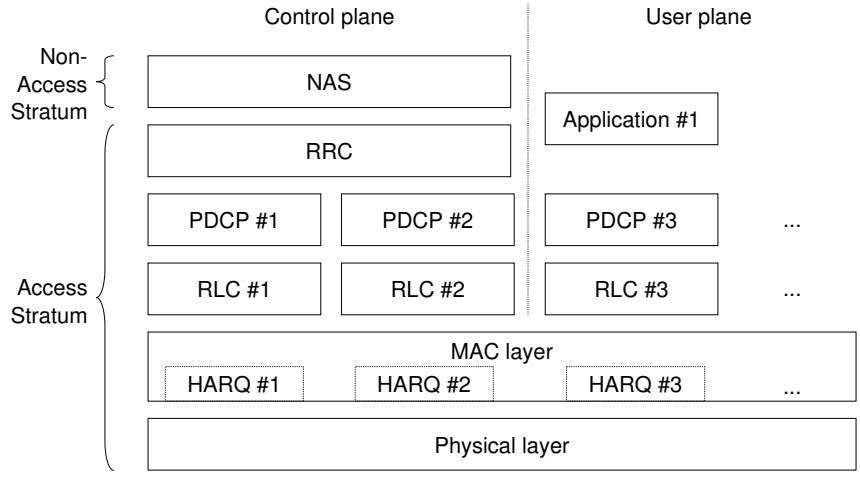


Figure 4.1: Overview of user-plane architecture.

At the transmitting side, each layer receives a Service Data Unit (SDU) from a higher layer, for which the layer provides a service, and outputs a Protocol Data Unit (PDU) to the layer below. The RLC layer receives packets from the PDCP layer. These packets are called PDCP PDUs from a PDCP point of view and represent RLC SDUs from an RLC point of view. The RLC layer creates packets which are provided to the layer below, i.e. the MAC layer. The packets provided by RLC to the MAC layer are RLC PDUs from an RLC point of view, and MAC SDUs from a MAC point of view. At the receiving side, the process is reversed, with each layer passing SDUs up to the layer above, where they are received as PDUs.

An important design feature of the LTE protocol stack is that all the PDUs and SDUs are *byte aligned*.[1] This is to facilitate handling by microprocessors, which are normally defined to handle packets in units of bytes. In order to further reduce the processing requirements of the user plane protocol stack in LTE, the headers created by each of the PDCP, RLC and MAC layers are also byte-aligned. This implies that sometimes unused padding bits are needed in the headers, and thus the cost of designing for efficient processing is that a small amount of potentially-available capacity is wasted.

[1] Byte alignment means that the lengths of the PDUs and SDUs are multiples of 8 bits.

4.2 Packet Data Convergence Protocol (PDCP)

4.2.1 Functions and Architecture

The PDCP layer performs the following functions:

- Header compression and decompression for user plane data;
- Security functions:
  - ciphering and deciphering for user plane and control plane data;
  - integrity protection and verification for control plane data;
- Handover support functions:
  - in-sequence delivery and reordering of PDUs for the layer above at handover;
  - lossless handover for user plane data mapped on RLC Acknowledged Mode (AM, see Section 4.3.1).
- Discard for user plane data due to timeout.

The PDCP layer manages data streams in the user plane, as well as in the control plane (i.e. the RRC protocol – see Section 3.2), only for the radio bearers using either a Dedicated Control CHannel (DCCH) or a Dedicated Transport CHannel (DTCH) — see Section 4.4.1.2. The architecture of the PDCP layer differs for user plane data and control plane data, as shown in Figures 4.2 and 4.3 respectively.

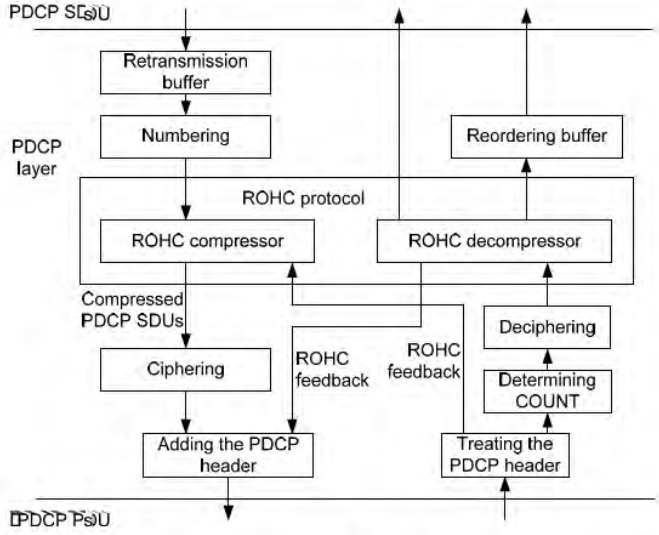


Figure 4.2: Overview of user-plane PDCP. Reproduced by permission of © 3GPP.

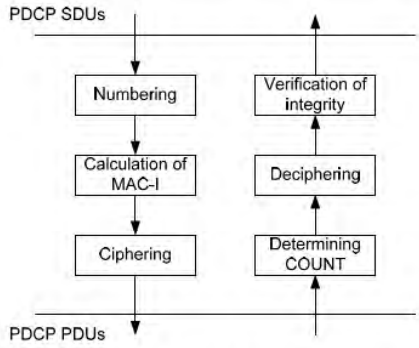


Figure 4.3: Overview of control-plane PDCP. Reproduced by permission of © 3GPP.

Each of the main functions is explained in the following subsections. Two different types of PDCP PDU are defined in LTE: PDCP Data PDUs and PDCP Control PDUs. PDCP Data PDUs are used for both control and user plane data. PDCP Control PDUs are only used to transport the feedback information for header compression, and for PDCP status reports which are used in case of handover (see Section 4.2.6) and hence are only used within the user plane.

4.2.2 Header Compression

One of the main functions of PDCP is header compression using the RObust Header Compression (ROHC) protocol defined by the IETF (Internet Engineering Task Force). In LTE, header compression is very important because there is no support for the transport of voice services via the Circuit-Switched (CS) domain.[2] Thus, in order to provide voice services on the Packet-Switched (PS) domain in a way that comes close to the efficiency normally associated with CS services it is necessary to compress the IP/UDP/RTP[3] header which is typically used for Voice over IP (VoIP) services.

The IETF specifies in 'RFC 4995'[4] a framework which supports a number of different header compression 'profiles' (i.e. sets of rules and parameters for performing the compression). The header compression profiles supported for LTE are shown in Table 4.1. This means that a UE may implement one or more of these ROHC profiles. It is important to notice that the profiles already defined in the IETF's earlier 'RFC 3095' have been redefined in RFC 4995 in order to increase robustness in some cases. The efficiency of RFC 3095 and RFC 4995 is similar, and UMTS[5] supports only RFC 3095.

[2] LTE does, however, support a CS FallBack (CSFB) procedure to allow an LTE UE to be handed over to a legacy RAT to originate a CS voice call, as well as a Single Radio Voice Call Continuity (SRVCC) procedure to hand over a Packet-Switched (PS) VoIP call to a CS voice call – see Sections 2.4.2.1 and 2.4.2.2.

[3] Internet Protocol / User Datagram Protocol / Real-Time Transport Protocol.

[4] Requests for Comments (RFCs) capture much of the output of the IETF.

[5] Universal Mobile Telecommunications System.

Table 4.1: Supported header compression protocols.

Reference	Usage
RFC 4995	No compression
RFC 3095, RFC 4815	RTP/UDP/IP
RFC 3095, RFC 4815	UDP/IP
RFC 3095, RFC 4815	ESP/IP
RFC 3843, RFC 4815	IP
RFC 4996	TCP/IP
RFC 5225	RTP/UDP/IP
RFC 5225	UDP/IP
RFC 5225	ESP/IP
RFC 5225	IP

The support of ROHC is not mandatory for the UE, except for those UEs which support VoIP. UEs which support VoIP have to support at least one profile for compression of RTP, UDP and IP.[6] The eNodeB controls by RRC signalling which of the ROHC profiles supported by the UE are allowed to be used. The ROHC compressors in the UE and the eNodeB then dynamically detect IP flows that use a certain IP header configuration and choose a suitable compression profile from the allowed and supported profiles.

ROHC header compression operates by allowing both the sender and the receiver to store the static parts of the header (e.g. the IP addresses of the sender/receiver), and to update these only when they change. Furthermore, dynamic parts (for example, the timestamp in the RTP header) are compressed by transmitting only the difference from a reference clock maintained in both the transmitter and the receiver.

As the non-changing parts of the headers are thus transmitted only once, successful decompression depends on their correct reception. Feedback is therefore used in order to confirm the correct reception of initialization information for the header decompression. Furthermore, the correct decompression of the received PDCP PDUs is confirmed periodically, depending on the experienced packet losses.

As noted above, the most important use case for ROHC is VoIP. Typically, for the transport of a VoIP packet which contains a payload of 32 bytes, the header added will be 60 bytes for the case of IPv6 and 40 bytes for the case of IPv4[7] – i.e. an overhead of 188% and 125% respectively. By means of ROHC, after the initialization of the header compression entities, this overhead can be compressed to four to six bytes, and thus to a relative overhead of 12.5–18.8%. This calculation is valid during the active periods, but during silence periods the payload size is smaller so the relative overhead is higher.

[6] ROHC is required for VoIP supported via the IP Multimedia Subsystem (IMS); in theory it could be possible to support raw IP VoIP without implementing ROHC.

[7] IPv6 is the successor to the original IPv4, for many years the dominant version of IP used on the Internet, and introduces a significantly expanded address space.

4.2.3 Security

The security architecture of LTE was introduced in Section 3.2.3.1. The implementation of security, by ciphering (of both control plane (RRC) data and user plane data) and integrity protection (for control plane (RRC) data only), is the responsibility of the PDCP layer.

A PDCP Data PDU counter (known as 'COUNT' in the LTE specifications) is used as an input to the security algorithms. The COUNT value is incremented for each PDCP Data PDU during an RRC connection; it has a length of 32 bits in order to allow an acceptable duration for an RRC connection.

During an RRC connection, the COUNT value is maintained by both the UE and the eNodeB by counting each transmitted/received PDCP Data PDU. In order to provide robustness against lost packets, each protected PDCP Data PDU includes a PDCP Sequence Number (SN) which corresponds to the least significant bits of the COUNT value.[8] Thus if one or more packets are lost, the correct COUNT value of a newly received packet can be determined using the PDCP SN. This means that the associated COUNT value is the next highest COUNT value for which the least significant bits correspond to the PDCP SN. A loss of synchronization of the COUNT value between the UE and eNodeB can then only occur if a number of packets corresponding to the maximum SN are lost consecutively. In principle, the probability of this kind of loss of synchronization occurring could be minimized by increasing the length of the SN, even to the extent of transmitting the whole COUNT value in every PDCP Data PDU. However, this would cause a high overhead, and therefore only the least significant bits are used as the SN; the actual SN length depends on the configuration and type of PDU, as explained in the description of the PDCP PDU formats in Section 4.2.6.

This use of a counter is designed to protect against a type of attack known as a *replay attack*, where the attacker tries to resend a packet that has been intercepted previously; the use of the COUNT value also provides protection against attacks which aim at deriving the used key or ciphering pattern by comparing successive patterns. Due to the use of the COUNT value, even if the same packet is transmitted twice, the ciphering pattern will be completely uncorrelated between the two transmissions, thus preventing possible security breaches.

Integrity protection is realized by adding a field known as 'Message Authentication Code for Integrity' (MAC-I)[9] to each RRC message. This code is calculated based on the Access Stratum (AS) derived keys (see Section 3.2.3.1), the message itself, the radio bearer ID, the direction (i.e. uplink or downlink) and the COUNT value.

If the integrity check fails, the message is discarded and the integrity check failure is indicated to the RRC layer so that the RRC connection re-establishment procedure can be executed (see Section 3.2.3.5).

Ciphering is realized by performing an XOR operation with the message and a ciphering stream that is generated by the ciphering algorithm based on the AS derived keys (see Section 3.2.3.1), the radio bearer ID, the direction (i.e. uplink or downlink), and the COUNT value.

Ciphering can only be applied to PDCP Data PDUs. PDCP Control PDUs (such as ROHC feedback or PDCP status reports) are neither ciphered nor integrity protected.

[8] In order to avoid excessive overhead, the most significant bits of the COUNT value, also referred to as the Hyper Frame Number (HFN), are not signalled but derived from counting overflows of the PDCP SN.

[9] Note that the MAC-I has no relation to the MAC layer.

Except for identical retransmissions, the same COUNT value is not allowed to be used more than once for a given security key. The eNodeB is responsible for avoiding reuse of the COUNT with the same combination of radio bearer ID, AS base-key and algorithm. In order to avoid such reuse, the eNodeB may for example use different radio bearer IDs for successive radio bearer establishments, trigger an intracell handover or trigger a UE state transition from connected to idle and back to connected again (see Section 3.2).

4.2.4 Handover

Handover is performed when the UE moves from the coverage of one cell to the coverage of another cell in RRC_CONNECTED state. Depending on the required QoS, either a seamless or a lossless handover is performed as appropriate for each user plane radio bearer, as explained in the following subsections.

4.2.4.1 Seamless Handover

Seamless handover is applied for user plane radio bearers mapped on RLC Unacknowledged Mode (UM, see Section 4.3.1). These types of data are typically reasonably tolerant of losses but less tolerant of delay (e.g. voice services). Seamless handover is therefore designed to minimize complexity and delay, but may result in loss of some SDUs.

At handover, for radio bearers to which seamless handover applies, the PDCP entities including the header compression contexts are reset, and the COUNT values are set to zero. As a new key is anyway generated at handover, there is no security reason to maintain the COUNT values. PDCP SDUs in the UE for which the transmission has not yet started will be transmitted after handover to the target cell. In the eNodeB, PDCP SDUs that have not yet been transmitted can be forwarded via the X2 interface[10] to the target eNodeB. PDCP SDUs for which the transmission has already started but that have not been successfully received will be lost. This minimizes the complexity because no context (i.e. configuration information) has to be transferred between the source and the target eNodeB at handover.

4.2.4.2 Lossless Handover

Based on the SN that is added to PDCP Data PDUs it is possible to ensure in-sequence delivery during handover, and even provide a fully lossless handover functionality, performing retransmission of PDCP SDUs for which reception has not yet been acknowledged prior to the handover. This lossless handover function is used mainly for delay-tolerant services such as file downloads where the loss of one PDCP SDU can result in a drastic reduction in the data rate due to the reaction of the Transmission Control Protocol (TCP).

Lossless handover is applied for user plane radio bearers that are mapped on RLC Acknowledged Mode (AM, see Section 4.3.1).

For lossless handover, the header compression protocol is reset in the UE because the header compression context is not forwarded from the source eNodeB to the target eNodeB. However, the PDCP SNs and the COUNT values associated with PDCP SDUs are maintained. For simplicity reasons, inter-eNodeB handover and intra-eNodeB handover are handled in the same way in LTE.

[10] For details of the X2 interface, see Section 2.6.

In normal transmission, while the UE is not handing over from one cell to another, the RLC layer in the UE and the eNodeB ensures in-sequence delivery. PDCP PDUs that are retransmitted by the RLC protocol, or that arrive out of sequence due to the variable delay in the HARQ transmission, are reordered based on the RLC SN. At handover, the RLC layer in the UE and in the eNodeB will deliver all PDCP PDUs that have already been received to the PDCP layer in order to have them decompressed before the header compression protocol is reset. Because some PDCP SDUs may not be available at this point, the PDCP SDUs that are not available in-sequence are not delivered immediately to higher layers in the UE or to the gateway in the network. In the PDCP layer, the PDCP SDUs received out of order are stored in the reordering buffer (see Figure 4.2). PDCP SDUs that have been transmitted but not yet been acknowledged by the RLC layer are stored in a retransmission buffer in the PDCP layer.

In order to ensure lossless handover in the *uplink*, the UE retransmits the PDCP SDUs stored in the PDCP retransmission buffer. This is illustrated in Figure 4.4. In this example the PDCP entity has initiated transmission for the PDCP SDUs with the sequence numbers 1 to 5; the packets with the sequence numbers 3 and 5 have not been received by the source eNodeB, for example due to the handover interrupting the HARQ retransmissions. After the handover, the UE restarts the transmission of the PDCP SDUs for which successful transmission has not yet been acknowledged to the target eNodeB. In the example in Figure 4.4 only the PDCP SDUs 1 and 2 have been acknowledged prior to the handover. Therefore, after the handover the UE will retransmit the packets 3, 4 and 5, although the network had already received packet 4.

In order to ensure in-sequence delivery in the uplink, the source eNodeB, after decompression, delivers the PDCP SDUs that are received in-sequence to the gateway, and forwards the PDCP SDUs that are received out-of-sequence to the target eNodeB. Thus, the target eNodeB can reorder the decompressed PDCP SDUs received from the source eNodeB and the retransmitted PDCP SDUs received from the UE based on the PDCP SNs which are maintained during the handover, and deliver them to the gateway in the correct sequence.

In order to ensure lossless handover in the *downlink*, the source eNodeB forwards the uncompressed PDCP SDUs for which reception has not yet been acknowledged by the UE to the target eNodeB for retransmission in the downlink. The source eNodeB receives an indication from the gateway that indicates the last packet sent to the source eNodeB. The source eNodeB also forwards this indication to the target eNodeB so that the target eNodeB knows when it can start transmission of packets received from the gateway. In the example in Figure 4.5, the source eNodeB has started the transmission of the PDCP SDUs 1 to 4; due to, for example, a handover occurring prior to the HARQ retransmissions of packet 3, packet 3 will not be received by the UE from the source eNodeB. Furthermore the UE has only sent an acknowledgment for packets 1 and 2, although packet 4 has been received by the UE. The target eNodeB then ensures that the PDCP SDUs that have not yet been acknowledged in the source eNodeB are sent to the UE. Thus, the UE can reorder the received PDCP SDUs and the PDCP SDUs that are stored in the reordering buffer, and deliver them to higher layers in sequential order.

The UE will expect the packets from the target eNodeB in ascending order of SNs. In the case of a packet not being forwarded from the source eNodeB to the target eNodeB, i.e. when one of the packets that the UE expects is missing during the handover operation, the UE can immediately conclude that the packet is lost and can forward the packets which have already been received in sequence to higher layers. This avoids the UE having to retain

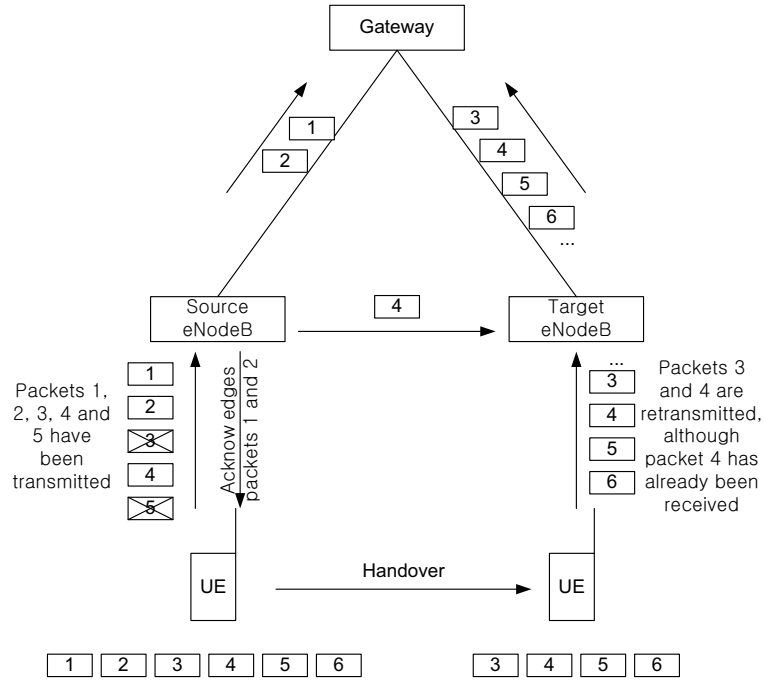


Figure 4.4: Lossless handover in the uplink.

already-received packets in order to wait for a potential retransmission. Thus the forwarding of the packets in the network can be decided without informing the UE.

In some cases it may happen that a PDCP SDU has been successfully received, but a corresponding RLC acknowledgement has not. In this case, after the handover, there may be unnecessary retransmissions initiated by the UE or the target eNodeB based on the incorrect status received by the RLC layer. In order to avoid these unnecessary retransmissions a PDCP status report can be sent from the eNodeB to the UE and from the UE to the eNodeB as described in Section 4.2.6. Additionally, a PDCP 'Status Report' can request retransmission of PDCP SDUs which were correctly received but failed in header decompression. Whether to send a PDCP status report after handover is configured independently for each radio bearer.

4.2.5 Discard of Data Packets

Typically, the data rate that is available on the radio interface is smaller than the data rate available on the network interfaces (e.g. S1[11]). Thus, when the data rate of a given service is higher than the data rate provided by the LTE radio interface, this leads to buffering in the UE and in the eNodeB. This buffering allows the scheduler in the MAC layer some freedom to vary the instantaneous data rate at the physical layer in order to adapt to the current radio

[11] For details of the S1 interface, see Section 2.5.

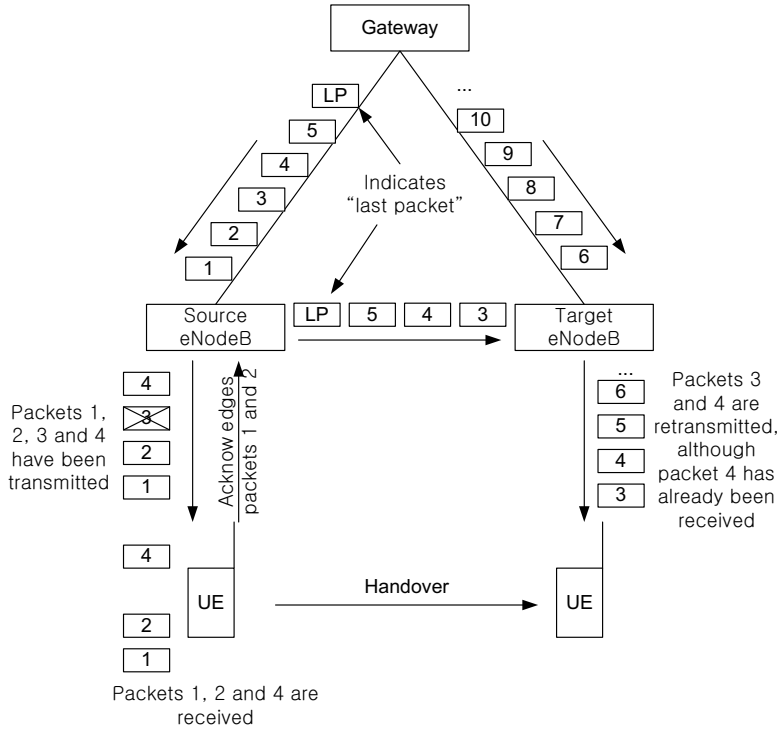


Figure 4.5: Lossless handover in the downlink.

channel conditions. Thanks to the buffering, the variations in the instantaneous data rate are then seen by the application only as some jitter in the transfer delay.

However, when the data rate provided by the application exceeds the data rate provided by the radio interface for a long period, large amounts of buffered data can result. This may lead to a large loss of data at handover if lossless handover is not applied to the bearer, or to an excessive delay for real time applications.

In the fixed internet, one of the roles typically performed by the routers is to drop packets when the data rate of an application exceeds the available data rate in a part of the internet. An application may then detect this loss of packets and adapt its data rate to the available rate. A typical example is the TCP transmit window handling, where the transmit window of TCP is reduced when a lost packet is detected, thus adapting to the available rate. Other applications such as video or voice calls via IP can also detect lost packets, for example via RTCP (Real-time Transport Control Protocol) feedback, and can adapt the data rate accordingly.

In order to allow these mechanisms to work, and to prevent excessive delay, a discard function is included in the PDCP layer for LTE. This discard function is based on a timer, where for each PDCP SDU received from the higher layers in the transmitter a timer is started, and when the transmission of the PDCP SDU has not yet been initiated in the UE at the expiry of this timer the PDCP SDU is discarded. If the timer is set to an appropriate value

for the required QoS of the radio bearer, this discard mechanism can prevent excessive delay and queuing in the transmitter.

4.2.6 PDCP PDU Formats

PDCP PDUs for user plane data comprise a 'D/C' field in order to distinguish Data and Control PDUs, the formats of which are shown in Figures 4.6 and 4.7 respectively. PDCP Data PDUs comprise a 7- or 12-bit SN as shown in Table 4.2. PDCP Data PDUs for user plane data contain either an uncompressed (if header compression is not used) or a compressed IP packet.

Figure 4.6: Key features of PDCP Data PDU format. See Table 4.2 for presence of D/C and MAC-I fields.

Table 4.2: PDCP Data PDU formats.

PDU type	D/C field	Sequence number length	MAC-I	Applicable RLC Modes (see Section 4.3.1)
User plane long SN	Present	12 bits	Absent	AM / UM
User plane short SN	Present	7 bits	Absent	UM
Control plane	Absent	5 bits	32 bits	AM

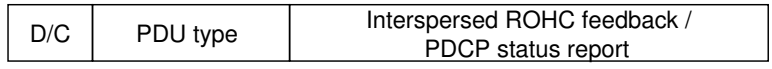


Figure 4.7: Key features of PDCP Control PDU format.

PDCP Data PDUs for control plane data (i.e. RRC signalling) comprise a MAC-I field of 32-bit length for integrity protection. PDCP Data PDUs for control plane data contain one complete RRC message.

As can be seen in Table 4.2 there are three types of PDCP Data PDU, distinguished mainly by the length of the PDCP SN and the presence of the MAC-I field. As mentioned in Section 4.2.3, the length of the PDCP SN in relation to the data rate, the packet size and the packet inter-arrival rate determines the maximum possible interruption time without desynchronizing the COUNT value which is used for ciphering and integrity protection.

The PDCP Data PDU for user plane data using the long SN allows longer interruption times, and makes it possible, when it is mapped on RLC Acknowledged Mode (AM) (see Section 4.3.1.3), to perform lossless handover as described in Section 4.2.4, but implies a higher overhead. Therefore it is mainly used for data applications with a large IP packet size where the overhead compared to the packet size is not too significant, for example for file transfer, web browsing, or e-mail traffic.

The PDCP Data PDU for user plane data using the short SN is mapped on RLC Unacknowledged Mode (UM) (see Section 4.3.1.2) and is typically used for VoIP services, where only seamless handover is used and retransmission is not necessary.

PDCP Control PDUs are used by PDCP entities handling user plane data (see Figure 4.1). There are two types of PDCP Control PDU, distinguished by the PDU Type field in the PDCP header. PDCP Control PDUs carry either PDCP 'Status Reports' for the case of lossless handover, or ROHC feedback created by the ROHC header compression protocol. PDCP Control PDUs carrying ROHC feedback are used for user plane radio bearers mapped on either RLC UM or RLC AM, while PDCP Control PDUs carrying PDCP Status Reports are used only for user plane radio bearers mapped on RLC AM.

In order to reduce complexity, a PDCP Control PDU carrying ROHC feedback carries exactly one ROHC feedback packet – there is no possibility to transmit several ROHC feedback packets in one PDCP PDU.

A PDCP Control PDU carrying a PDCP Status Report for the case of lossless handover is used to prevent the retransmission of already-correctly-received PDCP SDUs, and also to request retransmission of PDCP SDUs which were correctly received but for which header decompression failed. This PDCP Control PDU contains a bitmap indicating which PDCP SDUs need to be retransmitted and a reference SN, the First Missing SDU (FMS). In the case that all PDCP SDUs have been received in sequence this field indicates the next expected SN, and no bitmap is included.

4.3 Radio Link Control (RLC)

The RLC layer is located between the PDCP layer (the 'upper' layer) and the MAC layer (the 'lower' layer). It communicates with the PDCP layer through a Service Access Point (SAP), and with the MAC layer via logical channels. The RLC layer reformats PDCP PDUs in order to fit them into the size indicated by the MAC layer; that is, the RLC transmitter segments and/or concatenates the PDCP PDUs, and the RLC receiver reassembles the RLC PDUs to reconstruct the PDCP PDUs.

In addition, the RLC reorders the RLC PDUs if they are received out of sequence due to the HARQ operation performed in the MAC layer. This is the key difference from UMTS, where the HARQ reordering is performed in the MAC layer. The advantage of HARQ reordering in RLC is that no additional SN and reception buffer are needed for HARQ reordering. In LTE, the RLC SN and RLC reception buffer are used for both HARQ reordering and RLC-level ARQ related operations.

The functions of the RLC layer are performed by 'RLC entities'. An RLC entity is configured in one of three data transmission modes: Transparent Mode (TM), Unacknowledged Mode (UM), and Acknowledged Mode (AM). In AM, special functions are defined to support retransmission. When UM or AM is used, the choice between the two modes is made by

the eNodeB during the RRC radio bearer setup procedure (see Section 3.2.3.3), based on the QoS requirements of the EPS bearer.[12] The three RLC modes are described in detail in the following sections.

4.3.1 RLC Entities

4.3.1.1 Transparent Mode (TM) RLC Entity

As the name indicates, the TM RLC entity is transparent to the PDUs that pass through it – no functions are performed and no RLC overhead is added. Since no overhead is added, an RLC SDU is directly mapped to an RLC PDU and vice versa. Therefore, the use of TM RLC is very restricted. Only RRC messages which do not need RLC configuration can utilize the TM RLC, such as broadcast System Information (SI) messages, paging messages, and RRC messages which are sent when no Signalling Radio Bearers (SRBs) other than SRB0 (see Section 3.2.1) are available. TM RLC is not used for user plane data transmission in LTE.

TM RLC provides a unidirectional data transfer service – in other words, a single TM RLC entity is configured either as a transmitting TM RLC entity or as a receiving TM RLC entity.

4.3.1.2 Unacknowledged Mode (UM) RLC Entity

UM RLC provides a unidirectional data transfer service like TM RLC. UM RLC is mainly utilized by delay-sensitive and error-tolerant real-time applications, especially VoIP, and other delay-sensitive streaming services. Point-to-multipoint services such as MBMS (Multimedia Broadcast/Multicast Service) also use UM RLC – since no feedback path is available in the case of point-to-multipoint services, AM RLC cannot be utilized by these services.

A block diagram of the UM RLC entity is shown in Figure 4.8.

The main functions of UM RLC can be summarized as follows:

- Segmentation and concatenation of RLC SDUs;
- Reordering of RLC PDUs;
- Duplicate detection of RLC PDUs;
- Reassembly of RLC SDUs.

Segmentation and concatenation. The transmitting UM RLC entity performs segmentation and/or concatenation on RLC SDUs received from upper layers, to form RLC PDUs. The size of the RLC PDU at each transmission opportunity is decided and notified by the MAC layer depending on the radio channel conditions and the available transmission resources; therefore, the size of each transmitted RLC PDU can be different.

The transmitting UM RLC entity includes RLC SDUs into an RLC PDU in the order in which they arrive at the UM RLC entity. Therefore, a single RLC PDU can contain RLC SDUs or segments of RLC SDUs according to the following pattern:

(zero or one) SDU segment + (zero or more) SDUs + (zero or one) SDU segment.

[12] Evolved Packet System – see Section 2.

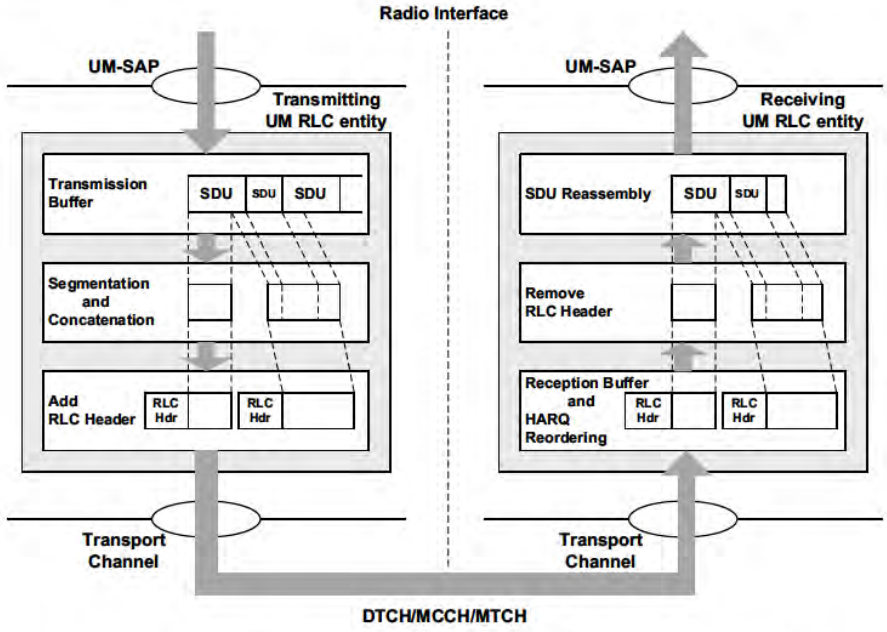


Figure 4.8: Model of UM RLC entities. Reproduced by permission of © 3GPP.

The constructed RLC PDU is always byte-aligned and has no padding.

After segmentation and/or concatenation of RLC SDUs, the transmitting UM RLC entity includes relevant UM RLC headers in the RLC PDU to indicate the sequence number[13] of the RLC PDU, and additionally the size and boundary of each included RLC SDU or RLC SDU segment.

Reordering, duplicate detection, and reassembly. When the receiving UM RLC entity receives RLC PDUs, it first reorders them if they are received out of sequence. Out-of-sequence reception is unavoidable due to the fact that the HARQ operation in the MAC layer uses multiple HARQ processes (see Section 4.4). Any RLC PDUs received out of sequence are stored in the reception buffer until all the previous RLC PDUs are received and delivered to the upper layer.

During the reordering process, any duplicate RLC PDUs received are detected by checking the SNs and discarded. This ensures that the upper layer receives upper layer PDUs only once. The most common cause of receiving duplicates is HARQ ACKs for MAC PDUs being misinterpreted as NACKs, resulting in unnecessary retransmissions of the MAC PDUs, which causes duplication in the RLC layer.

To detect reception failures and avoid excessive reordering delays, a reordering timer is used in the receiving UM RLC entity to set the maximum time to wait for the reception of RLC PDUs that have not been received in sequence. The receiving UM RLC entity starts

[13] Note that the RLC sequence number is independent from the sequence number added by PDCP.

the reordering timer when a missing RLC PDU is detected, and it waits for the missing RLC PDUs until the timer expires. When the timer expires, the receiving UM RLC entity declares the missing RLC PDUs as lost and starts to reassemble the next available RLC SDUs from the RLC PDUs stored in the reception buffer.

The reassembly function is performed on an RLC SDU basis; only RLC SDUs for which all segments are available are reassembled from the stored RLC PDUs and delivered to the upper layers. RLC SDUs that have at least one missing segment are simply discarded and not reassembled. If RLC SDUs were concatenated in an RLC PDU, the reassembly function in the RLC receiver separates them into their original RLC SDUs. The RLC receiver delivers reassembled RLC SDUs to the upper layers in increasing order of SNs.

An example scenario of a lost RLC PDU with HARQ reordering is shown in Figure 4.9. A reordering timer is started when the RLC receiver receives PDU#8. If PDU#7 has not been received before the timer expires, the RLC receiver decides that the PDU#7 is lost, and starts to reassemble RLC SDUs from the next received RLC PDU. In this example, SDU#22 and SDU#23 are discarded because they are not completely received, and SDU#24 is kept in the reception buffer until all segments are received. Only SDU#21 is completely received, so it is delivered up to the PDCP layer.

The reordering and duplicate detection functions are not applicable to RLC entities using the Multicast Control CHannel (MCCH) or Multicast Traffic CHannel (MTCH) (see Section 4.4.1.2). This is because the HARQ operation in the MAC layer is not used for these channels.

4.3.1.3 Acknowledged Mode (AM) RLC Entity

In contrast to the other RLC transmission modes, AM RLC provides a bidirectional data transfer service. Therefore, a single AM RLC entity is configured with the ability both to transmit and to receive – we refer to the corresponding parts of the AM RLC entity as the *transmitting side* and the *receiving side* respectively.

The most important feature of AM RLC is 'retransmission'. An ARQ operation is performed to support error-free transmission. Since transmission errors are corrected by retransmissions, AM RLC is mainly utilized by error-sensitive and delay-tolerant non-real-time applications. Examples of such applications include most of the interactive/background type services, such as web browsing and file downloading. Streaming-type services also frequently use AM RLC if the delay requirement is not too stringent. In the control plane, RRC messages typically utilize the AM RLC in order to take advantage of RLC acknowledgements and retransmissions to ensure reliability.

A block diagram of the AM RLC entity is shown in Figure 4.10.

Although the AM RLC block diagram looks complicated at first glance, the transmitting and receiving sides are similar to the UM RLC transmitting and receiving entities respectively, except for the retransmission-related blocks. Therefore, most of the UM RLC behaviour described in the previous section applies to AM RLC in the same manner. The transmitting side of the AM RLC entity performs segmentation and/or concatenation of RLC SDUs received from upper layers to form RLC PDUs together with relevant AM RLC headers, and the receiving side of the AM RLC entity reassembles RLC SDUs from the received RLC PDUs after HARQ reordering.

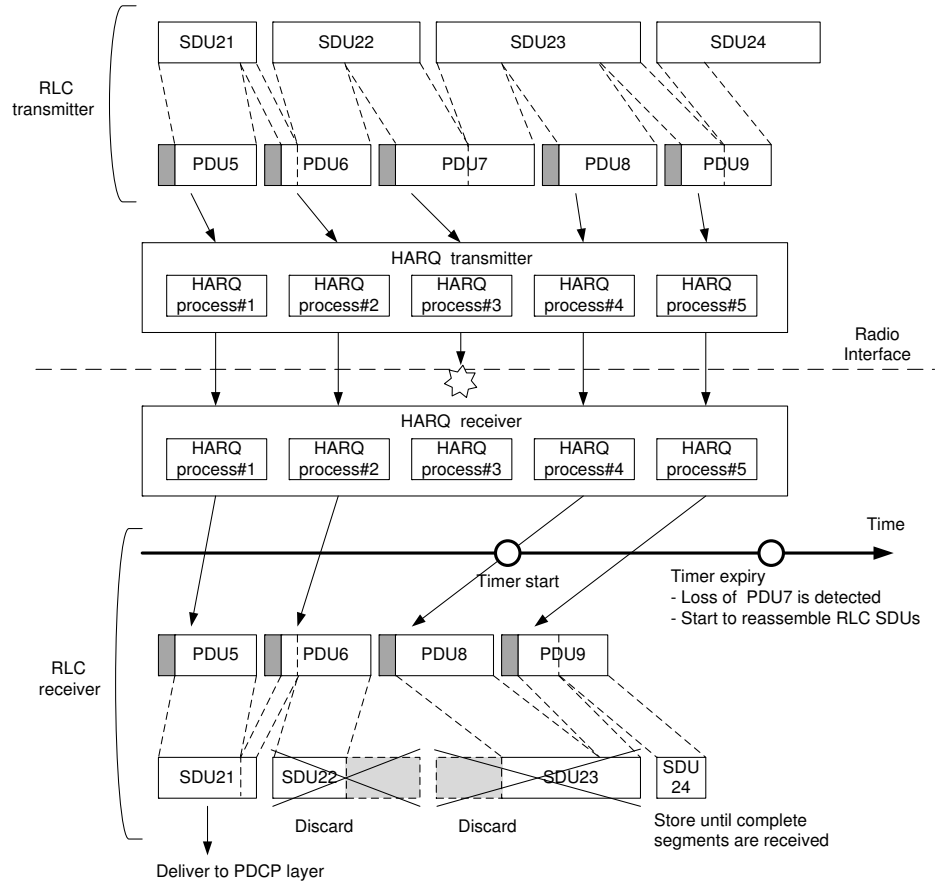


Figure 4.9: Example of PDU loss detection with HARQ reordering.

In addition to performing the functions of UM RLC, the main functions of AM RLC can be summarized as follows:

- Retransmission of RLC Data PDUs;
- Re-segmentation of retransmitted RLC Data PDUs;
- Polling;
- Status reporting;
- Status prohibit.

Retransmission and resegmentation. As mentioned before, the most important function of AM RLC is *retransmission*. In order that the transmitting side retransmits only the missing RLC PDUs, the receiving side provides a 'Status Report' to the transmitting side indicating ACK and/or NACK information for the RLC PDUs. Status reports are sent by

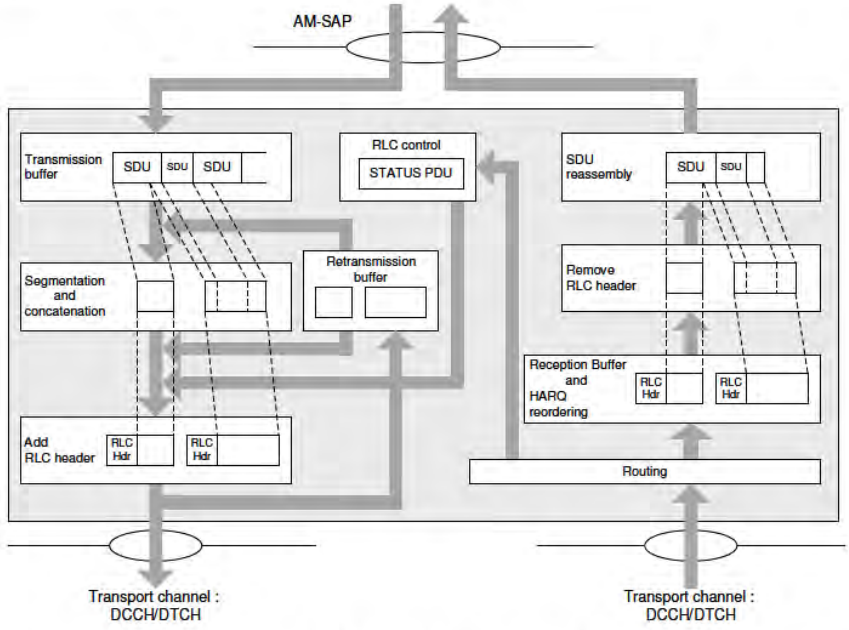


Figure 4.10: Model of AM RLC entity. Reproduced by permission of © 3GPP.

the transmitting side of the AM RLC entity whose receiving side received the corresponding RLC PDUs. Hence, the AM RLC transmitting side is able to transmit two types of RLC PDU, namely RLC Data PDUs containing data received from upper layers and RLC Control PDUs generated in the AM RLC entity itself. To differentiate between Data and Control PDUs, a 1-bit flag is included in the AM RLC header (see Section 4.3.2.3).

When the transmitting side transmits RLC Data PDUs, it stores the PDUs in the retransmission buffer for possible retransmission if requested by the receiver through a status report. In case of retransmission, the transmitter can resegment the original RLC Data PDUs into smaller PDU segments if the MAC layer indicates a size that is smaller than the original RLC Data PDU size.

An example of RLC resegmentation is shown in Figure 4.11. In this example, an original PDU of 600 bytes is resegmented into two PDU segments of 200 and 400 bytes at retransmission.

The original RLC PDU is distinguished from the retransmitted segments by another 1-bit flag in the AM RLC header: in the case of a retransmitted segment, some more fields are included in the AM RLC header to indicate resegmentation related information. The receiver can use status reports to indicate the status of individual retransmitted segments, not just full PDUs.

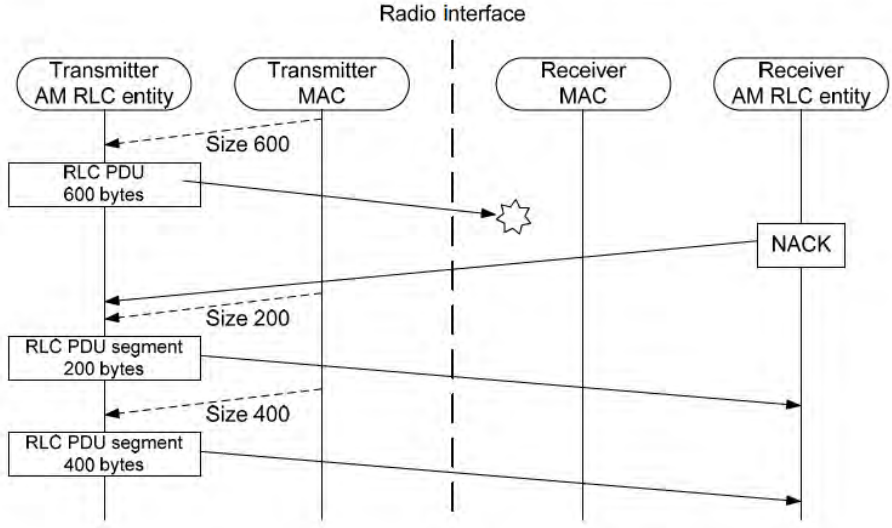


Figure 4.11: Example of RLC resegmentation.

Polling, status report and status prohibit. The transmitting side of the AM RLC entity can proactively request a status report from the peer receiving side by *polling* using a 1-bit indicator in the AM RLC header. The transmitting side can then use the status reports to select the RLC Data PDUs to be retransmitted, and manage transmission and retransmission buffers efficiently. Typical circumstances in which the transmitting side may initiate a poll include, for example, the last PDU in the transmitting side having been transmitted, or a predefined number of PDUs or data bytes having been transmitted.

When the receiving side of the AM RLC entity receives a poll from the peer transmitting side, it checks the reception buffer status and transmits a status report at the earliest transmission opportunity.

The receiving side can also generate a status report of its own accord if it detects a reception failure of an AM RLC PDU. For the detection of a reception failure, a similar mechanism is used as in the case of UM RLC in relation to the HARQ reordering delay. In AM RLC, however, the detection of a reception failure triggers a status report instead of considering the relevant RLC PDUs as permanently lost.

Note that the transmission of status reports needs to be carefully controlled according to the trade-off between transmission delay and radio efficiency. To reduce the transmission delay, status reports need to be transmitted frequently, but on the other hand frequent transmission of status reports wastes radio resources. Moreover, if further status reports are sent whilst the retransmissions triggered by a previous status report have not yet been received, unnecessary retransmissions may result, thus consuming further radio resources; in AM RLC this is in fact a second cause of duplicate PDUs occurring which have to be discarded by the duplicate-detection functionality. Therefore, to control the frequency of

status reporting in an effective way, a 'status prohibit' function is available in AM RLC, whereby the transmission of new status reports is prohibited while a timer is running.

4.3.2 RLC PDU Formats

As mentioned above, the RLC layer provides two types of PDU, namely the RLC Data PDU and the RLC Control PDU. The RLC Data PDU is used to transmit PDCP PDUs and is defined in all RLC transmission modes. The RLC Control PDU delivers control information between peer RLC entities and is defined only in AM RLC. The RLC PDUs used in each RLC transmission mode are summarized in Table 4.3.

Table 4.3: PDU types used in RLC.

RLC Mode	Data PDU	Control PDU
TM	TMD (TM Data)	N/A
UM	UMD (UM Data)	N/A
AM	AMD (AM Data)/AMD segment	STATUS

In the following subsections, each of the RLC PDU formats is explained in turn.

4.3.2.1 Transparent Mode Data PDU Format

The Transparent Mode Data (TMD) PDU consists only of a data field and does not have any RLC headers. Since no segmentation or concatenation is performed, an RLC SDU is directly mapped to a TMD PDU.

4.3.2.2 Unacknowledged Mode Data PDU Format

The Unacknowledged Mode Data (UMD) PDU (Figure 4.12) consists of a data field and UMD PDU header. PDCP PDUs (i.e. RLC SDUs) can be segmented and/or concatenated into the data field. The UMD PDU header is further categorized into a fixed part (included in each UMD PDU) and an extension part (included only when the data field contains more than one SDU or SDU segment – i.e. only when the data field contains any SDU borders).

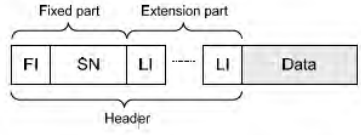


Figure 4.12: Key features of UMD PDU format.

- **Framing Info (FI).** This 2-bit field indicates whether the first and the last data field elements are complete SDUs or partial SDUs (i.e. whether the receiving RLC entity needs to receive multiple RLC PDUs in order to reassemble the corresponding SDU).
- **Sequence Number (SN).** For UMD PDUs, either a short (5 bits) or a long (10 bits) SN field can be used. This field allows the receiving RLC entity unambiguously to identify a UMD PDU, which allows reordering and duplicate-detection to take place.
- **Length Indicator (LI).** This 11-bit field indicates the length of the corresponding data field element present in the UMD PDU. There is a one-to-one correspondence between each LI and a data field element, except for the last data field element, for which the LI field is omitted because its length can be deduced from the UMD PDU size.

4.3.2.3 Acknowledged Mode Data PDU Format

In addition to the UMD PDU header fields, the Acknowledged Mode Data (AMD) PDU header (Figure 4.13) contains fields to support the RLC ARQ mechanism. The only difference in the PDU fields is that only the long SN field (10 bits) is used for AMD PDUs. The additional fields are as follows:

- **Data/Control (D/C).** This 1-bit field indicates whether the RLC PDU is an RLC Data PDU or an RLC Control PDU. It is present in all types of PDU used in AM RLC.
- **Resegmentation Flag (RF).** This 1-bit field indicates whether the RLC PDU is an AMD PDU or an AMD PDU segment.
- **Polling (P).** This 1-bit field is used to request a status report from the peer receiving side.

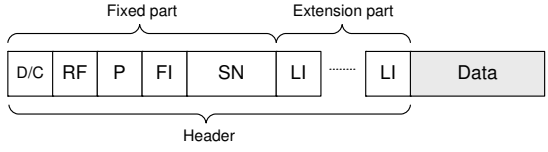


Figure 4.13: Key features of AMD PDU format.

4.3.2.4 AMD PDU Segment Format

The AMD PDU segment format (Figure 4.14) is used in case of resegmented retransmissions (when the available resource for retransmission is smaller than the original PDU size), as described in Section 4.3.1.3.

If the RF field indicates that the RLC PDU is an AMD PDU segment, the following additional resegmentation related fields are included in the fixed part of the AMD PDU header to enable correct reassembly:

- **Last Segment Flag (LSF).** This 1-bit field indicates whether or not this AMD PDU segment is the last segment of an AMD PDU.
- **Segmentation Offset (SO).** This 15-bit field indicates the starting position of the AMD PDU segment within the original AMD PDU.

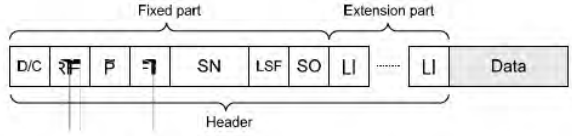


Figure 4.14: Key features of AMD PDU segment format.

4.3.2.5 STATUS PDU Format

The STATUS PDU (Figure 4.15) is designed to be very simple, as the RLC PDU error rate should normally be low in LTE due to the use of HARQ in the MAC layers. Therefore, the STATUS PDU simply lists all the missing portions of AMD PDUs by means of the following fields:

- **Control PDU Type (CPT).** This 3-bit field indicates the type of the RLC Control PDU, allowing more RLC Control PDU types to be defined in a later release of the LTE specifications. (The STATUS PDU is the only type of RLC Control PDU defined in the first version of LTE.)
- **ACK_SN.** This 10-bit field indicates the SN of the first AMD PDU which is neither received nor listed in this STATUS PDU. All AMD PDUs up to but not including this AMD PDU are correctly received by the receiver except the AMD PDUs or portions of AMD PDUs listed in the NACK_SN List.
- **NACK_SN List.** This field contains a list of SNs of the AMD PDUs that have not been completely received, optionally including indicators of which bytes of the AMD PDU are missing in the case of resegmentation.

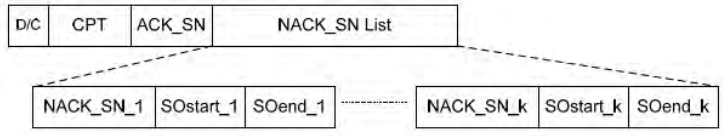


Figure 4.15: STATUS PDU format.

4.4 Medium Access Control (MAC)

The MAC layer is the lowest sublayer in the Layer 2 architecture of the LTE radio protocol stack. The connection to the physical layer below is through transport channels, and the connection to the RLC layer above is through logical channels. The MAC layer therefore performs multiplexing and demultiplexing between logical channels and transport channels: the MAC layer in the transmitting side constructs MAC PDUs, known as Transport Blocks (TBs), from MAC SDUs received through logical channels, and the MAC layer in the receiving side recovers MAC SDUs from MAC PDUs received through transport channels.

4.4.1 MAC Architecture

4.4.1.1 Overall Architecture

Figure 4.16 shows a conceptual overview of the architecture of the MAC layer.

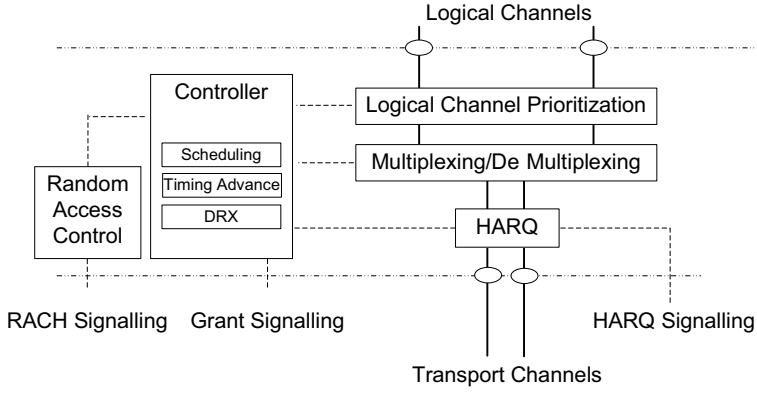


Figure 4.16: Conceptual overview of the UE-side MAC architecture.

The MAC layer consists of a HARQ entity, a multiplexing/demultiplexing entity, a logical channel prioritization entity, a random access control entity, and a controller which performs various control functions.

The HARQ entity is responsible for the transmit and receive HARQ operations. The transmit HARQ operation includes transmission and retransmission of TBs, and reception and processing of ACK/NACK signalling. The receive HARQ operation includes reception of TBs, combining of the received data and generation of ACK/NACK signalling. In order to enable continuous transmission while previous TBs are being decoded, up to eight HARQ processes in parallel are used to support multiprocess 'Stop-And-Wait' (SAW) HARQ operation.

SAW operation means that upon transmission of a TB, a transmitter stops further transmissions and awaits feedback from the receiver. When a NACK is received, or when a certain time elapses without receiving any feedback, the transmitter retransmits the TB. Such

a simple SAW HARQ operation cannot on its own utilize the transmission resources during the period between the first transmission and the retransmission. Therefore multiprocess HARQ interlaces several independent SAW processes in time so that all the transmission resources can be used. Each HARQ process is responsible for a separate SAW operation and manages a separate buffer.

In general, HARQ schemes can be categorized as either *synchronous* or *asynchronous*, with the retransmissions in each case being either *adaptive* or *non-adaptive*.

In a synchronous HARQ scheme, the retransmission(s) for each process occur at predefined times relative to the initial transmission. In this way, there is no need to signal information such as HARQ process number, as this can be inferred from the transmission timing. By contrast, in an asynchronous HARQ scheme, the retransmissions can occur at any time relative to the initial transmission, so additional explicit signalling is required to indicate the HARQ process number to the receiver, so that the receiver can correctly associate each retransmission with the corresponding initial transmission. In summary, synchronous HARQ schemes reduce the signalling overhead while asynchronous HARQ schemes allow more flexibility in scheduling.

In an adaptive HARQ scheme, transmission attributes such as the Modulation and Coding Scheme (MCS) and transmission resource allocation in the frequency domain can be changed at each retransmission in response to variations in the radio channel conditions. In a non-adaptive HARQ scheme, the retransmissions are performed without explicit signalling of new transmission attributes – either by using the same transmission attributes as those of the previous transmission, or by changing the attributes according to a predefined rule. Accordingly, adaptive schemes bring more scheduling gain at the expense of increased signalling overhead.

In LTE, asynchronous adaptive HARQ is used for the downlink, and synchronous HARQ for the uplink. In the uplink, the retransmissions may be either adaptive or non-adaptive, depending on whether new signalling of the transmission attributes is provided.

The details of the HARQ incremental redundancy schemes and timing for retransmissions are explained in Section 10.3.2.5.

In the multiplexing and demultiplexing entity, data from several logical channels can be (de)multiplexed into/from one transport channel. The multiplexing entity generates MAC PDUs from MAC SDUs when radio resources are available for a new transmission, based on the decisions of the logical channel prioritization entity. The demultiplexing entity reassembles the MAC SDUs from MAC PDUs and distributes them to the appropriate RLC entities. In addition, for peer-to-peer communication between the MAC layers, control messages called 'MAC Control Elements' can be included in the MAC PDU as explained in Section 4.4.2.7 below.

The logical channel prioritization entity prioritizes the data from the logical channels to decide how much data and from which logical channel(s) should be included in each MAC PDU, as explained in Section 4.4.2.6. The decisions are delivered to the multiplexing and demultiplexing entity.

The random access control entity is responsible for controlling the Random Access CHannel (RACH) procedure (see Section 4.4.2.3). The controller entity is responsible for a number of functions including Discontinuous Reception (DRX) , the Data Scheduling procedure, and for maintaining the uplink timing alignment. These functions are explained in the following sections.

4.4.1.2 Logical Channels

The MAC layer provides a data transfer service for the RLC layer through logical channels, which are either Control Logical Channels (for the transport of control data such as RRC signalling), or Traffic Logical Channels (for user plane data).

Control logical channels.

- **Broadcast Control CHannel (BCCH).** This is a downlink channel which is used to broadcast System Information (SI) and any Public Warning System (PWS) messages (see Section 13.7). In the RLC layer, it is associated with a TM RLC entity (see Section 4.3.1).
- **Paging Control CHannel (PCCH).** This is a downlink channel which is used to notify UEs of an incoming call or a change of SI. In the RLC layer, it is associated with a TM RLC entity (see Section 4.3.1).
- **Common Control CHannel (CCCH).** This channel is used to deliver control information in both uplink and downlink directions when there is no confirmed association between a UE and the eNodeB – i.e. during connection establishment. In the RLC layer, it is associated with a TM RLC entity (see Section 4.3.1).
- **Multicast Control CHannel (MCCH).** This is a downlink channel which is used to transmit control information related to the reception of MBMS services (see Chapter 13). In the RLC layer, it is associated with a UM RLC entity (see Section 4.3.1).
- **Dedicated Control CHannel (DCCH).** This channel is used to transmit dedicated control information relating to a specific UE, in both uplink and downlink directions. It is used when a UE has an RRC connection with eNodeB. In the RLC layer, it is associated with an AM RLC entity (see Section 4.3.1).

Traffic logical channels.

- **Dedicated Traffic CHannel (DTCH).** This channel is used to transmit dedicated user data in both uplink and downlink directions. In the RLC layer, it can be associated with either a UM RLC entity or an AM RLC entity (see Section 4.3.1).
- **Multicast Traffic CHannel (MTCH).** This channel is used to transmit user data for MBMS services in the downlink (see Chapter 13). In the RLC layer, it is associated with a UM RLC entity (see Section 4.3.1).

4.4.1.3 Transport Channels

Data from the MAC layer is exchanged with the physical layer through transport channels. Data is multiplexed into transport channels depending on how it is transmitted over the air. Transport channels are classified as downlink or uplink as follows:

Downlink transport channels.

- **Broadcast CHannel (BCH).** This channel is used to transport the parts of the SI which are essential for access the Downlink Shared CHannel (DL-SCH). The transport format is fixed and the capacity is limited.
- **Downlink Shared CHannel (DL-SCH).** This channel is used to transport downlink user data or control messages. In addition, the remaining parts of the SI that are not transported via the BCH are transported on the DL-SCH.
- **Paging CHannel (PCH).** This channel is used to transport paging information to UEs, and to inform UEs about updates of the SI (see Section 3.2.2) and PWS messages (see Section 13.7).
- **Multicast CHannel (MCH).** This channel is used to transport MBMS user data or control messages that require MBSFN combining (see Chapter 13).

The mapping of the downlink transport channels onto physical channels is explained in Section 6.4.

Uplink transport channels.

- **Uplink Shared CHannel (UL-SCH).** This channel is used to transport uplink user data or control messages.
- **Random Access CHannel (RACH).** This channel is used for access to the network when the UE does not have accurate uplink timing synchronization, or when the UE does not have any allocated uplink transmission resource (see Chapter 17).

The mapping of the uplink transport channels onto physical channels is explained in Chapter 16.

4.4.1.4 Multiplexing and Mapping between Logical Channels and Transport Channels

Figures 4.17 and 4.18 show the possible multiplexing between logical channels and transport channels in the downlink and uplink respectively.

Note that in the downlink, the DL-SCH carries information from all the logical channels except the PCCH, MCCH and MTCH.

In the uplink, the UL-SCH carries information from all the logical channels.

4.4.2 MAC Functions

4.4.2.1 Scheduling

The scheduler in the eNodeB distributes the available radio resources in one cell among the UEs, and among the radio bearers of each UE. The details of the scheduling algorithm are left to the eNodeB implementation, but the signalling to support the scheduling is standardized. Some possible scheduling algorithms are discussed in Chapter 12.

In principle, the eNodeB allocates downlink or uplink radio resources to each UE based respectively on the downlink data buffered in the eNodeB and on Buffer Status Reports

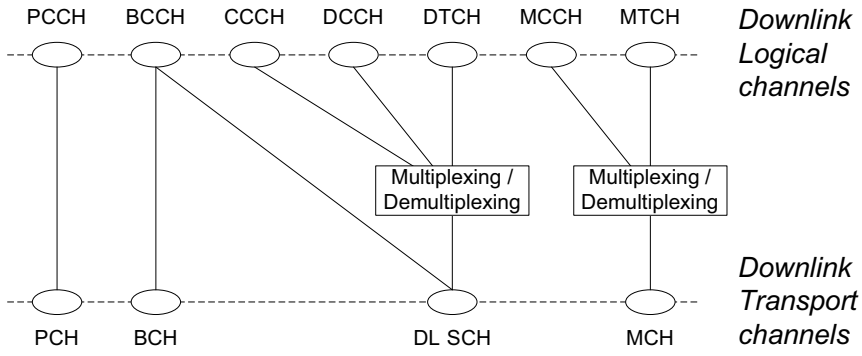


Figure 4.17: Downlink logical channel multiplexing. Reproduced by permission of © 3GPP.

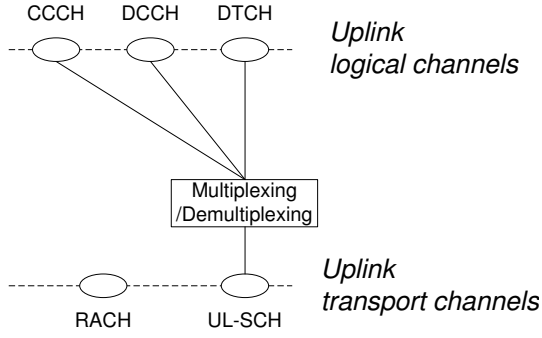


Figure 4.18: Uplink logical channel multiplexing. Reproduced by permission of © 3GPP.

(BSRs) received from the UE. In this process, the eNodeB considers the QoS requirements of each configured radio bearer, and selects the size of the MAC PDU.

The usual mode of scheduling is *dynamic scheduling*, by means of downlink assignment messages for the allocation of downlink transmission resources and uplink grant messages for the allocation of uplink transmission resources; these are valid for specific single subframes.[14] These messages also indicate whether the scheduled data is to be the first transmission of a new TB or a retransmission, by means of a 1-bit New Data Indicator (NDI); if the value of the NDI is changed relative to its previous value for the same HARQ process, the transmission is the start of a new TB. These messages are transmitted on the Physical Downlink Control CHannel (PDCCH) using a Cell Radio Network Temporary Identifier (C-RNTI) to identify the intended UE, as described in Section 9.3. This kind of scheduling is efficient for service types such as TCP or the SRBs, in which the traffic is bursty and dynamic in rate.

[14]The dynamic uplink transmission resource grants are valid for specific single subframes for initial transmissions, although they may also imply a resource allocation in later subframes for HARQ retransmissions.

In addition to the dynamic scheduling, *Semi-Persistent Scheduling* (SPS) may be used. SPS enables radio resources to be semi-statically configured and allocated to a UE for a longer time period than one subframe, avoiding the need for specific downlink assignment messages or uplink grant messages over the PDCCH for each subframe. It is useful for services such as VoIP for which the data packets are small, periodic and semi-static in size. For this kind of service the timing and amount of radio resources needed are predictable. Thus the overhead of the PDCCH is significantly reduced compared to the case of dynamic scheduling.

For the configuration of SPS, RRC signalling indicates the interval at which the radio resources are periodically assigned. Specific transmission resource allocations in the frequency domain, and transmission attributes such as the MCS, are signalled using the PDCCH. The actual transmission timing of the PDCCH messages is used as the reference timing to which the SPS interval applies. It is necessary to distinguish the PDCCH messages which apply to SPS from those used for dynamic scheduling; hence for SPS a special identity is used, known as the Semi-Persistent Scheduling C-RNTI (SPS-C-RNTI), which for each UE is different from the C-RNTI used for dynamic scheduling messages. The SPS-C-RNTI is used both for the configuration of SPS resources and for the indication of HARQ retransmissions of semi-persistently scheduled TBs. To differentiate these two cases, the NDI is used: for the configuration of the SPS resources, the SPS-C-RNTI with NDI set to 0 is used; for HARQ retransmissions, the SPS-C-RNTI with NDI set to 1 is used.

Reconfiguration of resources used for SPS can be performed for cases such as transitions between silent periods and talk spurts, or when the codec rate changes. For example, when the codec rate for a VoIP service is increased, a new downlink assignment message or uplink grant message can be transmitted to configure a larger semi-persistently scheduled radio resource for the support of bigger VoIP packets.

Allocated resources for SPS can be cancelled by an explicit scheduling message on the PDCCH using the SPS-C-RNTI indicating *SPS release*.[15] However, because there is a risk that scheduling messages can be lost in transmission, or the eNodeB's decision to release the resources may be late, an implicit mechanism to release the allocated radio resources is also specified. In the implicit mechanism, when a certain number of MAC PDUs not containing any MAC SDUs has been transmitted, the UE releases the radio resources.

Resources allocated for SPS can be temporarily overridden in a specific subframe by a scheduling message using the C-RNTI. For example, if the semi-persistently scheduled resources collide with resources configured for the Physical Random Access CHannel (PRACH) in a certain subframe, the eNodeB may choose to allocate other resources for the SPS in that subframe in order to avoid a collision with the PRACH.

Other factors potentially affecting the scheduling are uplink *Transmission Time Interval (TTI) bundling*, designed to improve uplink coverage (see Section 14.3.2), and the configuration of measurement gaps during which the UE tunes its receiver to other frequencies (see Section 22.2.1.2). In the latter case, whenever a subframe for a given HARQ process collides with a configured measurement gap, the UE can neither receive from nor transmit to a serving cell in that subframe. In such a case, if the UE cannot receive HARQ ACK/NACK feedback for an uplink TB, the UE considers that HARQ ACK is received TB and does not autonomously start a HARQ retransmission at the next transmission opportunity; to resume

[15] Explicit SPS resource release messages are positively acknowledged by the UE if they relate to downlink SPS, but not for uplink SPS.

HARQ operation, the UE has to receive a new scheduling message. If an uplink TB cannot be transmitted due to a measurement gap, the UE considers that HARQ NACK is received for that TB and transmits the TB at the next opportunity.

4.4.2.2 Scheduling Information Transfer

Buffer Status Reports (BSRs) from the UE to the eNodeB are used to assist the eNodeB's allocation of uplink radio resources. The basic assumption underlying scheduling in LTE is that radio resources are only allocated for transmission to or from a UE if data is available to be sent or received. In the downlink direction, the scheduler in the eNodeB is obviously aware of the amount of data to be delivered to each UE; however, in the uplink direction, because the scheduling decisions are performed in the eNodeB and the buffer for the data is located in the UE, BSRs have to be sent from the UE to the eNodeB to indicate the amount of data in the UE that needs to be transmitted over the UL-SCH.[16]

Two types of BSR are defined in LTE: a long BSR and a short BSR; which one is transmitted depends on the amount of available uplink transmission resources for sending the BSR, on how many groups of logical channels have non-empty buffers, and on whether a specific event is triggered at the UE. The long BSR reports the amount of data for four logical channel groups, whereas the short BSR reports the amount of data for only one logical channel group. Although the UE might actually have more than four logical channels configured, the overhead would be large if the amount of data in the UE were to be reported for every logical channel individually. Thus, grouping the logical channels into four groups for reporting purposes represents a compromise between efficiency and accuracy.

A BSR can be triggered in the following situations:

- whenever data arrives for a logical channel which has a higher priority than the logical channels whose buffers previously contained data (this is known as a Regular BSR);
- whenever data becomes available for any logical channel when there was previously no data available for transmission (a Regular BSR);
- whenever a 'retxBSR' timer expires and there is data available for transmission (a Regular BSR);
- whenever a 'periodicBSR' timer[17] expires (a Periodic BSR);
- whenever spare space in a MAC PDU can accommodate a BSR (a Padding BSR).

The 'retxBSR' timer provides a mechanism to recover from situations where a BSR is transmitted but not received. For example, if the eNodeB fails to decode a MAC PDU containing a BSR and returns a HARQ NACK, but the UE erroneously decodes the NACK as ACK, the UE will think that transmission of the BSR was successful even though it was not received by the eNodeB. In such a case, a long delay would be incurred while the UE

[16] Note that, unlike High Speed Uplink Packet Access (HSUPA), there is no possibility in LTE for a UE to transmit autonomously in the uplink by means of a transmission grant for non-scheduled transmissions. This is because the uplink transmissions from different UEs in LTE are orthogonal in time and frequency, and therefore if an uplink resource is allocated but unused, it cannot be accessed by another UE; by contrast, in HSUPA, if a UE does not use its transmission grant for non-scheduled transmissions, the resulting reduction in uplink interference can benefit other UEs. Furthermore, the short subframe length in LTE enables uplink transmission resources to be dynamically allocated more quickly than in HSUPA.

[17] Periodic BSR timer is used by RRC to control BSR reporting.

waited for an uplink resource grant that would not be forthcoming. To avoid this, the retxBSR timer is restarted whenever a uplink grant message is received; if no uplink grant is received before the timer expires, the UE transmits another BSR.

If a UE does not have enough allocated UL-SCH resources to send a BSR when a trigger for a Regular BSR occurs, the UE sends a Scheduling Request (SR) on the Physical Uplink Control CHannel (PUCCH – see Section 16.3.7) if possible; otherwise, the random access procedure (see Section 4.4.2.3) is used to request an allocation of uplink resources for sending a BSR. However, if a periodic or padding BSR is triggered when the UE does not have UL-SCH resources for a new transmission, the SR is not triggered.

Thus LTE provides suitable signalling to ensure that the eNodeB has sufficient information about the data waiting in each UE's uplink transmission buffer to allocate corresponding uplink transmission resources in a timely manner.

4.4.2.3 Random Access Procedure

The random access procedure is used when a UE is not allocated with uplink radio resources but has data to transmit, or when the UE is not time-synchronized in the uplink direction. Control of the random access procedure is an important part of the MAC layer functionality in LTE. The details are explained in Chapter 17.

4.4.2.4 Uplink Timing Alignment

Uplink timing alignment maintenance is controlled by the MAC layer and is important for ensuring that a UE's uplink transmissions arrive in the eNodeB without overlapping with the transmissions from other UEs. The details of the uplink timing advance mechanism used to maintain timing alignment are explained in Section 18.2.

The timing advance mechanism utilizes MAC Control Elements (see Section 4.4.2.7) to update the uplink transmission timing. However, maintaining the uplink synchronization in this way during periods when no data is transferred wastes radio resources and adversely impacts the UE battery life. Therefore, when a UE is inactive for a certain period of time the UE is allowed to lose uplink synchronization even in RRC_CONNECTED state. The random access procedure is then used to regain uplink synchronization when the data transfer resumes in either uplink or downlink.

4.4.2.5 Discontinuous Reception (DRX)

DRX functionality can be configured for an 'RRC_CONNECTED' UE[18] so that it does not always need to monitor the downlink channels. A DRX cycle consists of an 'On Duration' during which the UE should monitor the PDCCH and a 'DRX period' during which a UE can skip reception of downlink channels for battery saving purposes.

The parameterization of the DRX cycle involves a trade-off between battery saving and latency. On the one hand, a long DRX period is beneficial for lengthening the UE's battery life. For example, in the case of a web browsing service, it is usually a waste of resources for a UE continuously to receive downlink channels while the user is reading a downloaded

[18]Different DRX functionality applies to UEs which are in 'RRC_IDLE'. These RRC states are discussed in Chapter 3.

web page. On the other hand, a shorter DRX period is better for faster response when data transfer is resumed – for example when a user requests another web page.

To meet these conflicting requirements, two DRX cycles – a short cycle and a long cycle – can be configured for each UE, with the aim of providing a similar degree of power saving for the UE in RRC_CONNECTED as in RRC_IDLE. The transition between the short DRX cycle, the long DRX cycle and continuous reception is controlled either by a timer or by explicit commands from the eNodeB. In some sense, the short DRX cycle can be considered as a confirmation period in case a late packet arrives, before the UE enters the long DRX cycle – if data arrives at the eNodeB while the UE is in the short DRX cycle, the data is scheduled for transmission at the next wake-up time and the UE then resumes continuous reception. On the other hand, if no data arrives at the eNodeB during the short DRX cycle, the UE enters the long DRX cycle, assuming that the packet activity is finished for the time being.

Figure 4.19 shows an example of DRX operation. The UE checks for scheduling messages (indicated by its C-RNTI on the PDCCH) during the 'On Duration' period of either the long DRX cycle or the short DRX cycle depending on the currently active cycle. When a scheduling message is received during an 'On Duration', the UE starts a 'DRX Inactivity Timer' and monitors the PDCCH in every subframe while the DRX Inactivity Timer is running. During this period, the UE can be regarded as being in a continuous reception mode. Whenever a scheduling message is received while the DRX Inactivity Timer is running, the UE restarts the DRX Inactivity Timer, and when it expires the UE moves into a short DRX cycle and starts a 'DRX Short Cycle Timer'. The short DRX cycle may also be initiated by means of a MAC Control Element (see Section 4.4.2.7). When the 'DRX Short Cycle Timer' expires, the UE moves into a long DRX cycle.

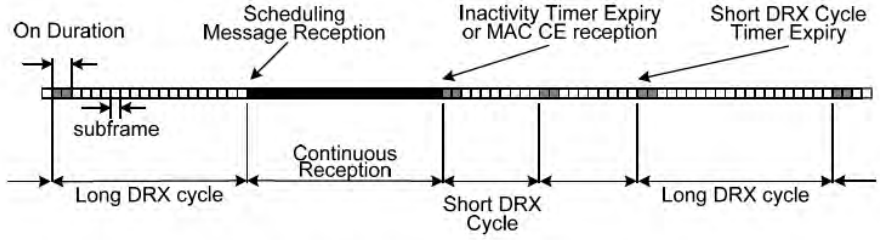


Figure 4.19: The two-level DRX procedure.

In addition to this DRX behaviour, a 'HARQ RTT (Round Trip Time) timer' is defined per downlink HARQ process with the aim of allowing the UE to sleep during the HARQ RTT. When decoding of a downlink TB for one HARQ process fails, the UE can assume that the next retransmission of the TB will occur after at least 'HARQ RTT' subframes. While the HARQ RTT timer is running, the UE does not need to monitor the PDCCH (provided that there is no other reason to be monitoring it). At the expiry of the HARQ RTT timer, if received data for a downlink HARQ process is not correctly decoded, the UE starts a 'DRX Retransmission Timer' for that HARQ process. While the timer is running, the UE

monitors the PDCCH for HARQ retransmissions. The length of the DRX Retransmission Timer is related to the degree of flexibility desired for the eNodeB's scheduler. For optimal UE battery consumption, it is desirable that eNodeB schedules a HARQ retransmission as soon as the HARQ RTT timer expires. However, this requires that eNodeB always reserve some radio resources for this, and therefore the DRX Retransmission timer can be used to relax this scheduling limitation while limiting the amount of time for which the UE has to monitor the PDCCH. The HARQ RTT is illustrated in Section 10.3.2.5.

4.4.2.6 Multiplexing and Logical Channel Prioritization

Unlike the downlink, where the multiplexing and logical channel prioritization is left to the eNodeB implementation, for the uplink the process by which a UE creates a MAC PDU to transmit using the allocated radio resources is fully standardized; this is designed to ensure that the UE satisfies the QoS of each configured radio bearer in a way which is optimal and consistent between different UE implementations. Based on the uplink transmission resource grant message signalled on the PDCCH, the UE has to decide on the amount of data for each logical channel to be included in the new MAC PDU, and, if necessary, also to allocate space for a MAC Control Element.

One simple way to meet this purpose is to serve radio bearers in order of their priority. Following this principle, the data from the logical channel of the highest priority is the first to be included into the MAC PDU, followed by data from the logical channel of the next highest priority, continuing until the MAC PDU size allocated by the eNodeB is completely filled or there is no more data to transmit.

Although this kind of priority-based multiplexing is simple and favours the highest priorities, it sometimes leads to starvation of low-priority bearers. Starvation occurs when the logical channels of the lower priority cannot transmit any data because the data from higher priority logical channels always takes up all the allocated radio resources.

To avoid starvation, while still serving the logical channels according to their priorities, in LTE a Prioritized Bit Rate (PBR) is configured by the eNodeB for each logical channel. The PBR is the data rate provided to one logical channel before allocating any resource to a lower-priority logical channel.

In order to take into account both the PBR and the priority, each logical channel is served in decreasing order of priority, but the amount of data from each logical channel included into the MAC PDU is initially limited to the amount corresponding to the configured PBR. Only when all logical channels have been served up to their PBR, then if there is still room left in the MAC PDU, each logical channel is served again in decreasing order of priority. In this second round, each logical channel is served only if all logical channels of higher priority have no more data for transmission.

In most cases, a MAC Control Element has higher priority than any other logical channel because it controls the operation of a MAC entity. Thus, when a MAC PDU is composed and there is a MAC Control Element to send, the MAC Control Element is generally included first and the remaining space is used to include data from logical channels. However, since the padding BSR is used to fill up remaining space in a MAC PDU, it is included into a MAC PDU after other logical channels. Among the various types of MAC Control Element (see Section 4.4.2.7) and logical channel, the 'C-RNTI' MAC Control Element and CCCH (CCCH) have the highest priority because they are used for either contention resolution or

RRC connection management. For example, the 'RRCConnectionReestablishmentRequest' message (see Section 3.2.3.5) on the uplink CCCH is used to recover a lost RRC connection, and it is more important to complete the connection reestablishment procedure as soon as possible than to inform the eNodeB of the UE's buffer status; otherwise, the data transfer interruption time would be longer and the probability of call failure would increase due to the delayed signalling. Likewise, a BSR of an unknown user is useless until the eNodeB knows which UE transmitted the BSR. Thus, the C-RNTI MAC Control Element has higher priority than the BSR MAC Control Element.

Figure 4.20 illustrates the LTE MAC multiplexing by way of example. First, channel 1 is served up to its PBR, channel 2 up to its PBR and then channel 3 with as much data as is available (since in this example the amount of data available is less than would be permitted by the PBR configured for that channel). After that, the remaining space in the MAC PDU is filled with data from the channel 1 which is of the highest priority until there is no further room in the MAC PDU or there is no further data from channel 1. If there is still a room after serving the channel 1, channel 2 is served in a similar way.

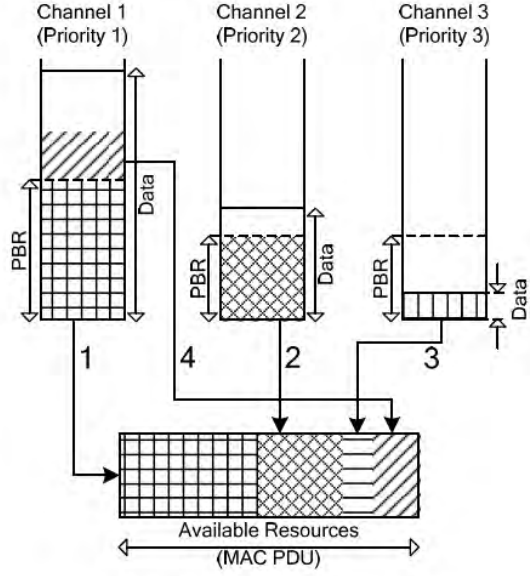


Figure 4.20: Example of MAC multiplexing.

4.4.2.7 MAC PDU Formats

When the multiplexing is done, the MAC PDU itself can be composed. The general MAC PDU format is shown in Figure 4.21. A MAC PDU primarily consists of the MAC header

and the MAC payload. The MAC header is further composed of MAC subheaders, while the MAC payload is composed of MAC control elements, MAC SDUs and padding.

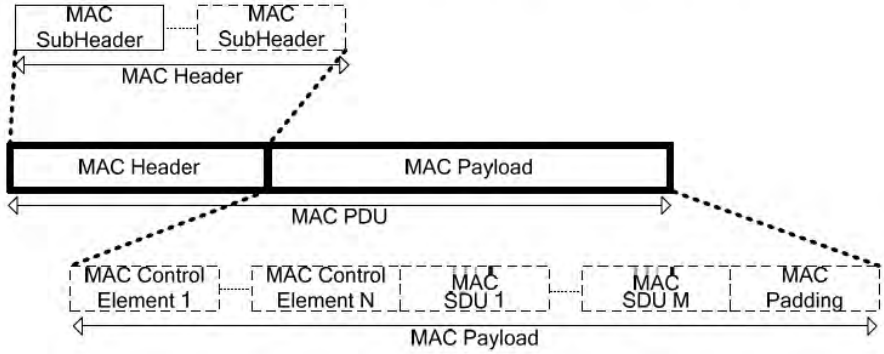


Figure 4.21: General MAC PDU format.

Each MAC subheader consists of a Logical Channel ID (LCID) and a Length (L) field. The LCID indicates whether the corresponding part of the MAC payload is a MAC Control Element, and if not, to which logical channel the related MAC SDU belongs. The L field indicates the size of the related MAC SDU or MAC Control Element.

MAC Control Elements are used for MAC-level peer-to-peer Signalling. The available types of MAC Control Element include the following:

- **Buffer Status Report MAC Control Element** for delivery of BSR information (see Section 4.4.2.2);
- **Power Headroom MAC Control Element** for the UE to report available power headroom (see Section 18.3.3);
- **DRX Command MAC Control Element** to transmit the downlink DRX commands to the UEs (see Section 4.4.2.5);
- **Timing Advance Command MAC Control Element** to transmit timing advance commands to the UEs for uplink timing alignment (see Sections 4.4.2.4 and 18.2.2);
- **C-RNTI MAC Control Element** for the UE to transmit its own C-RNTI during the random access procedure for the purpose of contention resolution (see Section 17.3.1);
- **UE Contention Resolution Identity MAC Control Element** for the eNodeB to transmit the uplink CCCH SDU that the UE has sent during the random access procedure for the purpose of contention resolution when the UE has no C-RNTI (see Section 17.3.1);
- **MBMS Dynamic Scheduling Information MAC Control Element** transmitted for each MCH to inform MBMS-capable UEs about scheduling of data transmissions on MTCH (see Section 13.6).

For each type of MAC Control Element, one special LCID is allocated.

When a MAC PDU is used to transport data from the PCCH or BCCH , the MAC PDU includes data from only one logical channel. In this case, because multiplexing is not applied, there is no need to include the LCID field in the header. In addition, if there is a one-to-one correspondence between a MAC SDU and a MAC PDU, the size of the MAC SDU can be known implicitly from the TB size. Thus, for these cases a headerless MAC PDU format is used as a transparent MAC PDU.

When a MAC PDU is used to transport the Random Access Response (RAR – see Section 17.3.1.2), a special MAC PDU format is applied with a MAC header and zero or more RARs. The MAC header consists of one or more MAC subheaders which include either a random access preamble identifier or a backoff indicator. Each MAC subheader including the Random Access Preamble IDentifier (RAPID) corresponds to one RAR in the MAC PDU (see Section 17.3.1).

4.5 Summary of the User Plane Protocols

The LTE Layer 2 protocol stack, consisting of the PDCP, RLC and MAC sublayers, acts as the interface between the radio access technology-agnostic sources of packet data traffic and the LTE physical layer. By providing functionality such as IP packet header compression, security, handover support, segmentation/concatenation, retransmission and reordering of packets, and transmission scheduling, the protocol stack enables the physical layer to be used efficiently for packet data traffic.

References[19]

[1] 3GPP Technical Specification 36.323, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Packet Data Convergence Protocol (PDCP) Specification', www.3gpp.org.

[2] 3GPP Technical Specification 36.322, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Link Control (RLC) Protocol Specification', www.3gpp.org.

[3] 3GPP Technical Specification 36.321, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification', www.3gpp.org.

[19] All web sites confirmed 1st March 2011.

Part II

Physical Layer for Downlink

5

Orthogonal Frequency Division Multiple Access (OFDMA)

Andrea Ancora, Issam Toufik, Andreas Bury and Dirk Slock

5.1 Introduction

The choice of an appropriate modulation and multiple-access technique for mobile wireless data communications is critical to achieving good system performance. In particular, typical mobile radio channels tend to be dispersive and time-variant, and this has generated interest in multicarrier modulation.

In general, multicarrier schemes subdivide the used channel bandwidth into a number of parallel subchannels as shown in Figure 5.1(a). Ideally the bandwidth of each subchannel is such that they are, ideally, each non-frequency-selective (i.e. having a spectrally flat gain); this has the advantage that the receiver can easily compensate for the subchannel gains individually in the frequency domain.

Orthogonal Frequency Division Multiplexing (OFDM) is a special case of multicarrier transmission where the non-frequency-selective narrowband subchannels, into which the frequency-selective wideband channel is divided, are overlapping but orthogonal, as shown in Figure 5.1(b). This avoids the need to separate the carriers by means of guard-bands, and therefore makes OFDM highly spectrally efficient. The spacing between the subchannels in OFDM is such that they can be perfectly separated at the receiver. This allows for a low-complexity receiver implementation, which makes OFDM attractive for high-rate mobile data transmission such as the LTE downlink.

It is worth noting that the advantage of separating the transmission into multiple narrowband subchannels cannot itself translate into robustness against time-variant channels if no channel coding is employed. The LTE downlink combines OFDM with channel coding

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

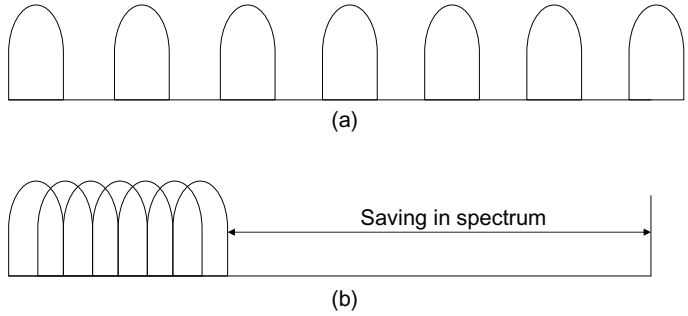


Figure 5.1: Spectral efficiency of OFDM compared to classical multicarrier modulation: (a) classical multicarrier system spectrum; (b) OFDM system spectrum.

and Hybrid Automatic Repeat reQuest (HARQ) to overcome the deep fading which may be encountered on the individual subchannels. These aspects are considered in Chapter 10 and lead to the LTE downlink falling under the category of system often referred to as 'Coded OFDM' (COFDM).

5.1.1 History of OFDM Development

Multicarrier communication systems were first introduced in the 1960s [1, 2], with the first OFDM patent being filed at Bell Labs in 1966. Initially only analogue design was proposed, using banks of sinusoidal signal generators and demodulators to process the signal for the multiple subchannels. In 1971, the use of the Discrete Fourier Transform (DFT) was proposed [3], which made OFDM implementation cost-effective. Further complexity reductions were realized in 1980 by the application of the Winograd Fourier Transform (WFT) or the Fast Fourier Transform (FFT) [4].

OFDM then became the modulation of choice for many applications for both wired systems (such as Asymmetric Digital Subscriber Line (ADSL)) and wireless systems. Wireless applications of OFDM tended to focus on broadcast systems, such as Digital Video Broadcasting (DVB) and Digital Audio Broadcasting (DAB), and relatively low-power systems such as Wireless Local Area Networks (WLANs). Such applications benefit from the low complexity of the OFDM receiver, while not requiring a high-power transmitter in the consumer terminals. This avoids one of the main disadvantages of OFDM, namely that the transmitters tend to be more expensive because of the high Peak to Average Power Ratio (PAPR); this aspect is discussed in Section 5.2.2.

The first cellular mobile radio system based on OFDM was proposed in [5]. Since then, the processing power of modern Digital Signal Processors (DSPs) has increased remarkably, paving the way for OFDM to be used in the LTE downlink. Here, the key benefits of OFDM which come to the fore are not only the low-complexity receiver but also the ability of OFDM to be adapted in a straightforward manner to operate in different channel bandwidths according to spectrum availability.

5.2 OFDM

5.2.1 Orthogonal Multiplexing Principle

A high-rate data stream typically faces the problem of having a symbol period T_s much smaller than the channel delay spread T_d if it is transmitted serially. This generates Inter-Symbol Interference (ISI) which can only be undone by means of a complex equalization procedure. In general, the equalization complexity grows with the square of the channel impulse response length.

In OFDM, the high-rate stream of data symbols is first Serial-to-Parallel (S/P) converted for modulation onto M parallel subcarriers as shown in Figure 5.2. This increases the symbol duration on each subcarrier by a factor of approximately M, such that it becomes significantly longer than the channel delay spread.

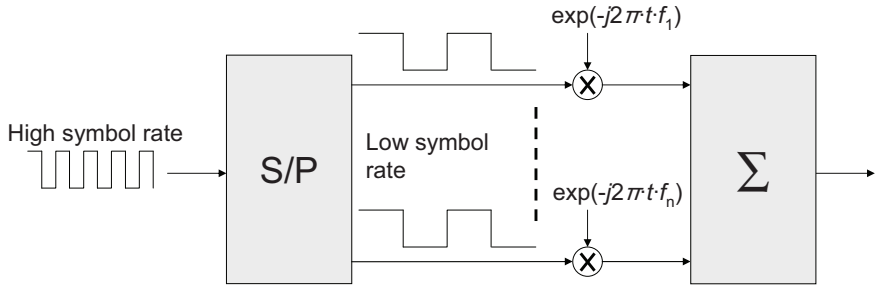


Figure 5.2: Serial-to-Parallel (S/P) conversion operation for OFDM.

This operation has the important advantage of requiring a much less complex equalization procedure in the receiver, under the assumption that the time-varying channel impulse response remains substantially constant during the transmission of each modulated OFDM symbol. Figure 5.3 shows how the resulting long symbol duration is virtually unaffected by ISI compared to the short symbol duration, which is highly corrupted.

Figure 5.4 shows the typical block diagram of an OFDM system. The signal to be transmitted is defined in the frequency domain. An S/P converter collects serial data symbols into a data block $\mathbf{S}\,[k] = [S_0\,[k]\,, S_1\,[k]\,, \ldots, S_{M\text{-}1}\,[k]]^T$ of dimension M, where k is the index of an OFDM symbol (spanning the M subcarriers). The M parallel data streams are first independently modulated resulting in the complex vector $\mathbf{X}\,[k] = [X_0\,[k]\,, X_1\,[k]\,, \ldots, X_{M\text{-}1}\,[k]]^T$. Note that in principle it is possible to use different modulations (e.g. QPSK or 16QAM) on each subcarrier; due to channel frequency selectivity, the channel gain may differ between subcarriers, and thus some subcarriers can carry higher data-rates than others. The vector $\mathbf{X}\,[k]$ is then used as input to an N-point Inverse FFT (IFFT) resulting in a set of N complex time-domain samples $\mathbf{x}\,[k] = [x_0[k], \ldots, x_{N\text{-}1}[k]]^T$. In a practical OFDM system, the number of processed subcarriers is greater than the number of modulated subcarriers (i.e. $N \geq M$), with the un-modulated subcarriers being padded with zeros.

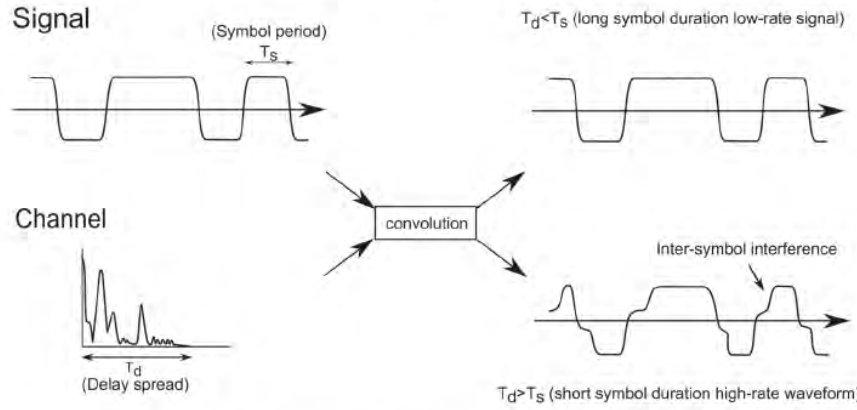


Figure 5.3: Effect of channel on signals with short and long symbol duration.

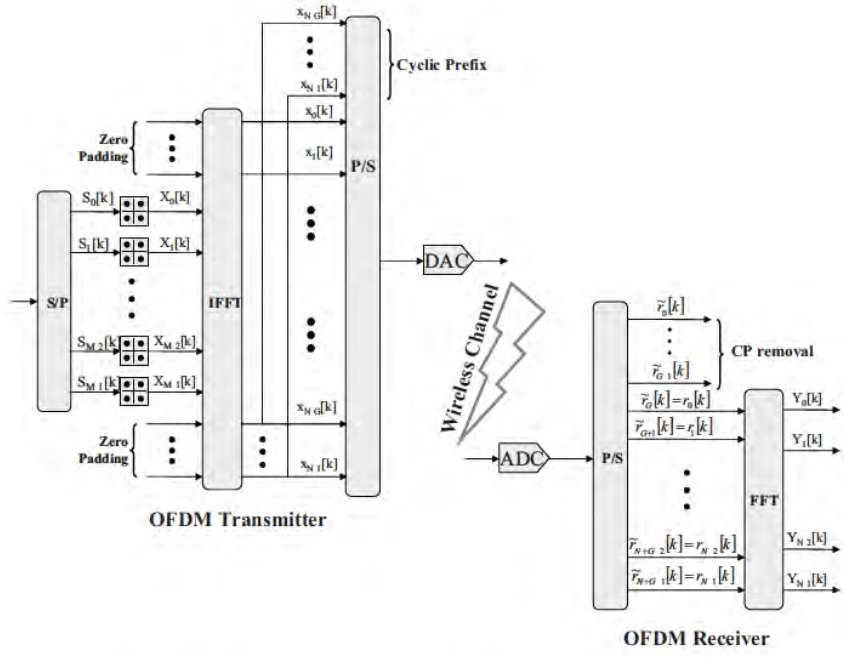


Figure 5.4: OFDM system model: (a) transmitter; (b) receiver.

The next key operation in the generation of an OFDM signal is the creation of a guard period at the beginning of each OFDM symbol $\mathbf{x}[k]$ by adding a Cyclic Prefix (CP), to eliminate the remaining impact of ISI caused by multipath propagation. The CP is generated by duplicating the last G samples of the IFFT output and appending them at the beginning of $\mathbf{x}[k]$. This yields the time domain OFDM symbol $[x_{N-G}[k], \ldots, x_{N-1}[k], x_0[k], \ldots, x_{N-1}[k]]^T$, as shown in Figure 5.5.

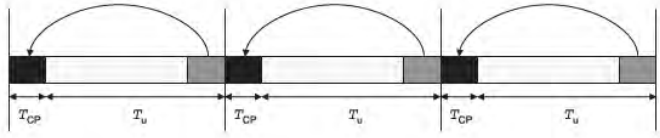


Figure 5.5: OFDM Cyclic Prefix (CP) insertion.

To avoid ISI completely, the CP length G must be chosen to be longer than the longest channel impulse response to be supported. The CP converts the linear (i.e. aperiodic) convolution of the channel into a circular (i.e. periodic) one which is suitable for DFT processing. The insertion of the CP into the OFDM symbol and its implications are explained more formally later in this section.

The output of the IFFT is then Parallel-to-Serial (P/S) converted for transmission through the frequency-selective channel.

At the receiver, the reverse operations are performed to demodulate the OFDM signal. Assuming that time- and frequency-synchronization is achieved (see Chapter 7), a number of samples corresponding to the length of the CP are removed, such that only an ISI-free block of samples is passed to the DFT. If the number of subcarriers N is designed to be a power of 2, a highly efficient FFT implementation may be used to transform the signal back to the frequency domain. Among the N parallel streams output from the FFT, the modulated subset of M subcarriers are selected and further processed by the receiver.

Let $x(t)$ be the symbol transmitted at time instant t. The received signal in a multipath environment is then given by

$$r(t) = x(t) * h(t) + z(t) \tag{5.1}$$

where $h(t)$ is the continuous-time impulse response of the channel, $*$ represents the convolution operation and $z(t)$ is the additive noise. Assuming that $x(t)$ is band-limited to $[-\frac{1}{2T_s}, \frac{1}{2T_s}]$, the continuous-time signal $x(t)$ can be sampled at sampling rate T_s such that the Nyquist criterion is satisfied. As a result of the multipath propagation, several replicas of the transmitted signals arrive at the receiver at different delays.

The received discrete-time OFDM symbol k including CP, under the assumption that the channel impulse response has a length smaller than or equal to G, can be expressed as

$$\tilde{\mathbf{r}} = \begin{bmatrix} \tilde{r}_0[k] \\ \tilde{r}_1[k] \\ \vdots \\ \tilde{r}_{\text{G-2}}[k] \\ \tilde{r}_{\text{G-1}}[k] \\ \tilde{r}_{\text{G}}[k] \\ \vdots \\ \tilde{r}_{\text{N+G-1}}[k] \end{bmatrix} = \mathbf{A} \cdot \begin{bmatrix} h_0[k] \\ h_1[k] \\ \vdots \\ h_{\text{G-1}}[k] \end{bmatrix} + \begin{bmatrix} z_0[k] \\ z_1[k] \\ \vdots \\ z_{\text{G-2}}[k] \\ z_{\text{G-1}}[k] \\ z_{\text{G}}[k] \\ \vdots \\ z_{\text{N+G-1}}[k] \end{bmatrix} \tag{5.2}$$

where

$$\mathbf{A} = \begin{bmatrix} x_{\text{N-G}}[k] & x_{\text{N-1}}[k-1] & x_{\text{N-2}}[k-1] & \cdots & x_{\text{N-G-1}}[k-1] \\ x_{\text{N-G+1}}[k] & x_{\text{N-G}}[k] & x_{\text{N-1}}[k-1] & \cdots & x_{\text{N-G+2}}[k-1] \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ x_{\text{N-2}}[k] & x_{\text{N-3}}[k] & \ddots & x_{\text{N-G}}[k] & x_{\text{N-1}}[k-1] \\ x_{\text{N-1}}[k] & x_{\text{N-2}}[k] & \ddots & x_{\text{N-G+1}}[k] & x_{\text{N-G}}[k] \\ x_0[k] & x_{\text{N-1}}[k] & \ddots & x_{\text{N-G+2}}[k] & x_{\text{N-G+1}}[k] \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ x_{\text{N-1}}[k] & x_{\text{N-2}}[k] & \cdots & \cdots & x_{\text{N-G}}[k] \end{bmatrix}$$

In general broadband transmission systems, one of the most complex operations the receiver has to handle is the equalization process to recover $x_n[k]$ (from Equation (5.2)).

Equation (5.2) can be written as the sum of intra-OFDM symbol interference (generated by the frequency-selective behaviour of the channel within an OFDM symbol) and the inter-OFDM symbol interference (between two consecutive OFDM block transmissions at time k and time $(k-1)$). This can be expressed as

$$\tilde{\mathbf{r}} = \mathbf{A}_{\text{Intra}} \cdot \begin{bmatrix} h_0[k] \\ h_1[k] \\ \vdots \\ h_{\text{G-1}}[k] \end{bmatrix} + \mathbf{A}_{\text{Inter}} \cdot \begin{bmatrix} h_0[k] \\ h_1[k] \\ \vdots \\ h_{\text{G-1}}[k] \end{bmatrix} + \begin{bmatrix} z_0[k] \\ z_1[k] \\ \vdots \\ z_{\text{N+G-1}}[k] \\ z_{\text{N-G}}[k] \\ \vdots \\ z_{\text{N+G-1}}[k] \end{bmatrix} \tag{5.3}$$

where

$$\mathbf{A}_{\text{Intra}} = \begin{bmatrix} x_{\text{N-G}}[k] & 0 & \cdots & 0 \\ x_{\text{N-G+1}}[1] & x_{\text{N-G}}[k] & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ x_{\text{N-1}}[k] & \cdots & \ddots & x_{\text{N-G}}[k] \\ x_0[k] & x_{\text{N-1}}[k] & \cdots & x_{\text{N-G+1}}[k] \\ \vdots & \ddots & \ddots & \vdots \\ x_{\text{N-1}}[k] & x_{\text{N-2}}[k] & \cdots & x_{\text{N-G}}[k] \end{bmatrix}$$

and

$$\mathbf{A}_{\text{Inter}} = \begin{bmatrix} 0 & x_{\text{N-1}}[k-1] & x_{\text{N-2}}[k-1] & \cdots & x_{\text{N-G+1}}[k-1] \\ 0 & 0 & x_{\text{N-1}}[k-1] & \cdots & x_{\text{N-G+2}}[N-1] \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & x_{\text{N-1}}[k-1] \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 \end{bmatrix}$$

In order to suppress the inter-OFDM-symbol interference, the first G samples of the received signal are discarded, eliminating the contribution of the matrix $\mathbf{A}_{\text{Inter}}$ and the G first rows of $\mathbf{A}_{\text{Intra}}$. This can be expressed as

$$\mathbf{r}[k] = \begin{bmatrix} r_0[k] \\ r_1[k] \\ \vdots \\ r_{\text{N-1}}[k] \end{bmatrix} = \begin{bmatrix} \tilde{r}_{\text{G}}[k] \\ \tilde{r}_{\text{G+1}}[k] \\ \vdots \\ \tilde{r}_{\text{N+G-1}}[k] \end{bmatrix}$$

$$= \begin{bmatrix} x_0[k] & x_{\text{N-1}}[k] & \cdots & x_{\text{N-G+1}}[k] \\ x_1[k] & x_0[k] & \cdots & x_{\text{N-G+2}}[k] \\ \vdots & \ddots & \ddots & \vdots \\ x_{\text{N-1}}[k] & x_{\text{N-2}}[k] & \cdots & x_{\text{N-G}}[k] \end{bmatrix} \cdot \begin{bmatrix} h_0[k] \\ h_1[k] \\ \vdots \\ h_{\text{G-1}}[k] \end{bmatrix} + \begin{bmatrix} z_{\text{G}}[k] \\ z_{\text{G+1}}[k] \\ \vdots \\ z_{\text{N+G-1}}[k] \end{bmatrix}$$

Adding zeros to the channel vector can extend the signal matrix without changing the output vector. This can be expressed as

$$\begin{bmatrix} r_0[k] \\ r_1[k] \\ \vdots \\ r_{\text{N-1}}[k] \end{bmatrix} = \mathbf{B} \cdot \begin{bmatrix} h_0[k] \\ h_1[k] \\ \vdots \\ h_{\text{G-1}}[k] \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} z_{\text{G}}[k] \\ z_{\text{G+1}}[k] \\ \vdots \\ z_{\text{N+G-1}}[k] \end{bmatrix}$$

where matrix **B** is given by

$$\mathbf{B} = \begin{bmatrix} x_0[k] & x_{\text{N-1}}[k] & \cdots & x_{\text{N-G+1}}[k] & x_{\text{N-G}}[k] & \cdots & x_1[k] \\ x_1[k] & x_0[k] & \cdots & x_{\text{N-G+2}}[k] & x_{\text{N-G+1}}[k] & \cdots & x_2[k] \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ x_{\text{N-1}}[k] & x_{\text{N-2}}[k] & \cdots & x_{\text{N-G}}[k] & x_{\text{N-G-1}}[k] & \cdots & x_0[k] \end{bmatrix}$$

The matrix **B** is circulant and thus its Fourier transform is diagonal with eigenvalues given by the FFT of its first row [6, 7]. It can then be written as $\mathbf{B} = \mathbf{F}^H\mathbf{X}\mathbf{F}$, with **X** diagonal, and the equivalent received signal can be expressed as follows:

$$\begin{bmatrix} r_0[k] \\ r_{\text{N-G+1}}[k] \\ \vdots \\ r_{\text{N-1}}[k] \end{bmatrix} = \mathbf{F}^H \cdot \begin{bmatrix} X_0[k] & 0 & \cdots & 0 \\ 0 & X_1[k] & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{\text{N-1}}[k] \end{bmatrix} \cdot \mathbf{F} \cdot \begin{bmatrix} h_0[k] \\ h_1[k] \\ \vdots \\ h_{\text{G-1}}[k] \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$+ \begin{bmatrix} z_{\text{N-G}}[k] \\ z_{\text{N-G+1}}[k] \\ \vdots \\ z_{\text{N+G-1}}[k] \end{bmatrix} \tag{5.4}$$

(5.5)

where **F** is the Fourier transform matrix whose elements are

$$(\mathbf{F})_{n,m} = \frac{1}{\sqrt{N}} \exp\left(- \frac{j2\pi}{N}(nm)\right)$$

for $0 \leq n \leq N-1$ and $0 \leq m \leq N-1$ and N is the length of the OFDM symbol. The elements on the diagonal are the eigenvalues of the matrix **B**, obtained as

$$X_{\text{m}}[k] = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} x_{\text{n}}[k] \exp\left(-2j\pi m \frac{n}{N}\right) \tag{5.6}$$

In reverse, the time-domain signal $x_{\text{n}}[k]$ can be obtained by

$$x_{\text{n}}[k] = \frac{1}{\sqrt{N}} \sum_{m=1}^{N} X_{\text{m}}[k] \exp\left(2j\pi m \frac{n}{N}\right) \tag{5.7}$$

By applying the Fourier transform to Equation (5.4), the equivalent received signal in the frequency domain can be obtained,

$$\begin{bmatrix} R_0[k] \\ \vdots \\ R_{\text{N-1}}[k] \end{bmatrix} = \begin{bmatrix} X_0[0] & 0 & \cdots & 0 \\ 0 & X_1[k] & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{\text{N-1}}[k] \end{bmatrix} \begin{bmatrix} H_0[k] \\ H_1[k] \\ \vdots \\ H_{\text{N-1}}[k] \end{bmatrix} + \begin{bmatrix} Z_0[k] \\ \vdots \\ Z_{\text{N-1}}[k] \end{bmatrix}$$

In summary, the CP of OFDM changes the linear convolution into a circular one. The circular convolution is very efficiently transformed by means of an FFT into a multiplicative operation in the frequency domain. Hence, the transmitted signal over a frequency-selective (i.e. multipath) channel is converted into a transmission over N parallel flat-fading channels in the frequency domain:

$$R_{\mathrm{m}}[k] = X_{\mathrm{m}}[k] \cdot H_{\mathrm{m}}[k] + Z_{\mathrm{m}}[k] \tag{5.8}$$

As a result, the equalization is much simpler than for single-carrier systems and consists of just one complex multiplication per subcarrier.

5.2.2 Peak-to-Average Power Ratio and Sensitivity to Non-Linearity

While the previous section shows the advantages of OFDM, this section highlights its major drawback: the Peak-to-Average Power Ratio (PAPR).

In the general case, the OFDM transmitter can be seen as a linear transform performed over a large block of independent identically distributed (i.i.d) QAM[1]-modulated complex symbols (in the frequency domain). From the central limit theorem [8, 9], the time-domain OFDM symbol may be approximated as a Gaussian waveform. The amplitude variations of the OFDM modulated signal can therefore be very high. However, practical Power Amplifiers (PAs) of RF transmitters are linear only within a limited dynamic range. Thus, the OFDM signal is likely to suffer from non-linear distortion caused by clipping. This gives rise to out-of-band spurious emissions and in-band corruption of the signal. To avoid such distortion, the PAs have to operate with large power back-offs, leading to inefficient amplification or expensive transmitters.

The PAPR is one measure of the high dynamic range of the input amplitude, and hence a measure of the expected degradation. To analyse the PAPR mathematically, let x_{n} be the signal after IFFT as given by Equation (5.7) where the time index k can be dropped without loss of generality. The PAPR of an OFDM symbol is defined as the square of the peak amplitude divided by the mean power, i.e.

$$\mathrm{PAPR} = \frac{\max_n\{|x_{\mathrm{n}}|^2\}}{E\{|x_{\mathrm{n}}|^2\}} \tag{5.9}$$

Under the hypothesis that the Gaussian approximation is valid, the amplitude of x_{n} has a Rayleigh distribution, while its power has a central chi-square distribution with two degrees of freedom. The Cumulative Distribution Function (CDF) $F_x(\alpha)$ of the normalized power is given by

$$F_x(\alpha) = \Pr\left(\frac{|x_{\mathrm{n}}|^2}{E\{|x_{\mathrm{n}}|^2\}} < \alpha\right) = 1 - e^{-\alpha} \tag{5.10}$$

If there is no oversampling, the time-domain samples are mutually uncorrelated and the probability that the PAPR is above a certain threshold PAPR_0 is given by[2]

$$\Pr(\mathrm{PAPR} > \mathrm{PAPR}_0) = 1 - F_x(\mathrm{PAPR}_0)^N = 1 - (1 - e^{-\mathrm{PAPR}_0})^N \tag{5.11}$$

Figure 5.6 plots the distribution of the PAPR given by Equation (5.11) for different values of the number of subcarriers N. The figure shows that a high PAPR does not occur very often. However, when it does occur, degradation due to PA non-linearities may be expected.

[1] Quadrature Amplitude Modulation.

[2] Note that the CDF of the PAPR is $F_{\mathrm{PAPR}}(\eta) = [F_x(\eta)]^N$ for N i.i.d. samples.

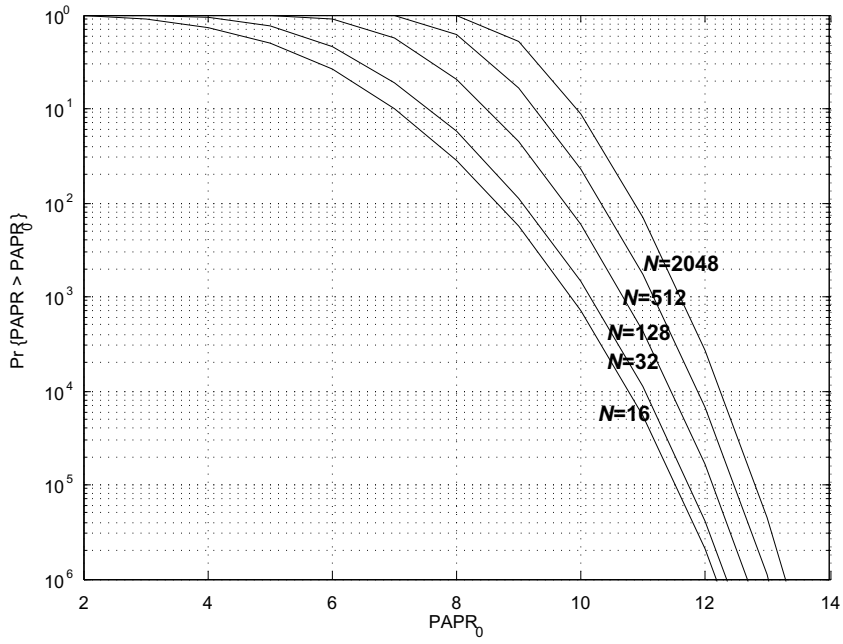


Figure 5.6: PAPR distribution for different numbers of OFDM subcarriers.

In order to evaluate the impacts of distortion on the OFDM signal reception, a useful framework for modelling non-linearities is developed in Section 21.5.2.

5.2.2.1 PAPR Reduction Techniques

Many techniques have been studied for reducing the PAPR of a transmitted OFDM signal. Although no such techniques are specified for the LTE downlink signal generation, an overview of the possibilities is provided below. In general, in LTE the cost and complexity of generating the OFDM signal with acceptable Error Vector Magnitude (EVM) (see Section 21.3.1.1) is left to the eNodeB implementation. As OFDM is not used for the LTE uplink (see Section 14.2), such considerations do not directly apply to the transmitter in the User Equipment (UE).

Techniques for PAPR reduction of OFDM signals can be broadly categorized into three main concepts: clipping and filtering [10–12], selected mapping [13] and coding techniques [14, 15]. The most potentially relevant from the point of view of LTE would be clipping and filtering, whereby the time-domain signal is clipped to a predefined level. This causes spectral leakage into adjacent channels, resulting in reduced spectral efficiency as well as in-band noise degrading the bit error rate performance. Out-of-band radiation caused by the clipping process can, however, be reduced by filtering. If discrete signals are clipped directly, the resulting clipping noise will all fall in band and thus cannot be reduced by filtering. To avoid this problem, one solution consists of oversampling the original signal by

padding the input signal with zeros and processing it using a longer IFFT. The oversampled signal is clipped and then filtered to reduce the out-of-band radiation.

5.2.3 Sensitivity to Carrier Frequency Offset and Time-Varying Channels

The orthogonality of OFDM relies on the condition that transmitter and receiver operate with exactly the same frequency reference. If this is not the case, the perfect orthogonality of the subcarriers is lost, causing subcarrier leakage, also known as Inter-Carrier Interference (ICI), as can be seen in Figure 5.7.

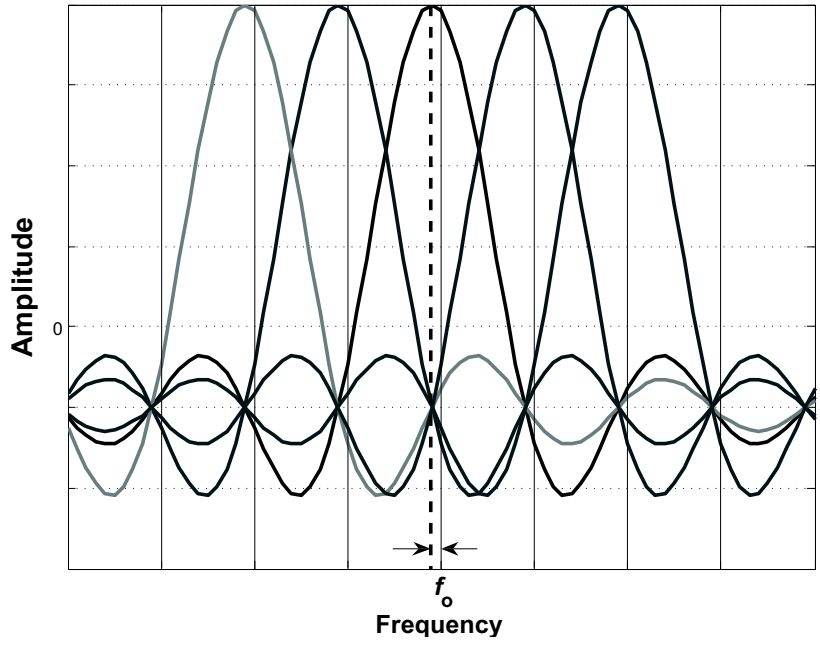


Figure 5.7: Loss of orthogonality between OFDM subcarriers due to frequency offset.

Frequency errors typically arise from a mismatch between the reference frequencies of the transmitter and the receiver local oscillators. On the receiver side in particular, due to the importance of using low-cost components in the mobile handset, local oscillator frequency drifts are usually greater than in the eNodeB and are typically a function of parameters such as temperature changes and voltage variations. This difference between the reference frequencies is widely referred to as Carrier Frequency Offset (CFO). Phase noise in the UE receiver may also result in frequency errors.

The CFO can be several times larger than the subcarrier spacing. It is usually divided into an integer part and a fractional part. Thus the frequency error can be written as

$$f_o = (\Gamma + \epsilon)\Delta f \tag{5.12}$$

where Δf is the subcarrier spacing, Γ is an integer and $-0.5 < \epsilon < 0.5$. If $\Gamma \neq 0$, then the modulated data are in the wrong positions with respect to the subcarrier mapping performed at the transmitter. This simply results in a Bit Error Rate (BER) of 0.5 if the frequency offset is not compensated at the receiver, independently of the value of ϵ. In the case of $\Gamma = 0$ and $\epsilon \neq 0$, the perfect subcarrier orthogonality is lost, resulting in ICI which can degrade the BER. Typically only synchronization errors of up to a few percent of the subcarrier spacing are tolerable in OFDM systems.

Even in an ideal case where the local oscillators are perfectly aligned, the relative speed between transmitter and receiver also generates a frequency error due to Doppler.

In the case of a single-path channel, UE mobility in a constant direction with respect to the angle of arrival of the signal results in a Doppler shift f_d, while in a scattering environment this becomes a Doppler spread with spectral density $P(f)$ as discussed further in Section 8.3.1.

It can be shown [16, 17] that, for both flat and dispersive channels, the ICI power can be computed as a function of the generic Doppler spectral density $P(f)$ as follows:

$$P_{\text{ICI}} = \int_{-f_{d_{\max}}}^{f_{d_{\max}}} P(f)(1 - \text{sinc}^2(T_s f))\, \mathrm{d}f \tag{5.13}$$

where $f_{d_{\max}}$ is the maximum Doppler frequency, and the transmitted signal power is normalized.

ICI resulting from a mismatch f_o between the transmitter and receiver oscillator frequencies can be modelled as a Doppler shift arising from single-path propagation:

$$P(f) = \delta(f - f_o) \tag{5.14}$$

Hence, substituting (5.14) into (5.13), the ICI power in the case of a deterministic CFO is given by

$$P_{\text{ICI,CFO}} = 1 - \text{sinc}^2(f_o T_s) \tag{5.15}$$

For the classical Jakes model of Doppler spread (see Section 8.3.1), Equation (5.13) can be written as

$$P_{\text{ICI,Jakes}} = 1 - 2\int_0^1 (1 - f) J_0(2\pi f_{d_{\max}} T_s f)\, \mathrm{d}f \tag{5.16}$$

where J_0 is the zero-th order Bessel function.

When no assumptions on the shape of the Doppler spectrum can be made, an upper bound on the ICI given by Equation (5.13) can be found by applying the Cauchy–Schwartz inequality, leading to [17]

$$P_{\text{ICI}} \leq \frac{\int_0^1 [1 - \text{sinc}^2(f_d T_s f)]^2\, \mathrm{d}f}{\int_0^1 1 - \text{sinc}^2(f_d T_s f)\, \mathrm{d}f} \tag{5.17}$$

This upper bound on P_{ICI} is valid only in the case of frequency spread and does not cover the case of a deterministic CFO.

Using Equations (5.17), (5.16) and (5.15), the Signal-to-Interference Ratio (SIR) in the presence of ICI can be expressed as

$$SIR_{\mathrm{ICI}} = \frac{1 - P_{\mathrm{ICI}}}{P_{\mathrm{ICI}}} \tag{5.18}$$

Figures 5.8 and 5.9 plot these P_{ICI} and SIR_{ICI} for the cases provided. These figures show that the highest ICI is introduced by a constant frequency offset. In the case of a Doppler spread, the ICI impairment is lower. Figure 5.9 shows that, in the absence of any other impairment such as interference ($SIR = \infty$), the SIR_{ICI} rapidly decays as a function of frequency misalignments.

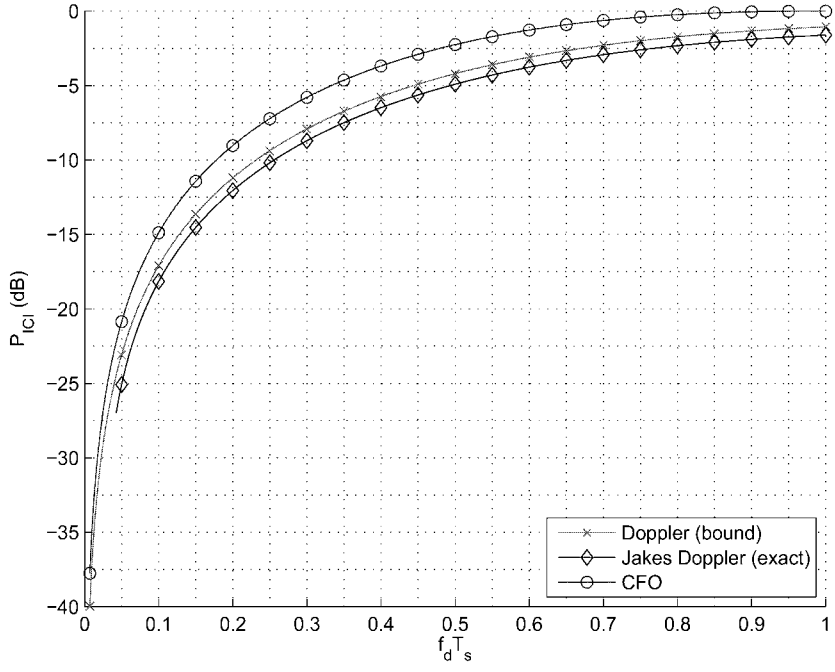


Figure 5.8: P_{ICI} for the case of a classical Doppler distribution and a deterministic CFO.

The sensitivity of the BER depends on the modulation order. It is shown in [18] that QPSK modulation can tolerate up to $\epsilon_{\max} = 0.05$ whereas 64-QAM requires $\epsilon \leq 0.01$.

5.2.4 Timing Offset and Cyclic Prefix Dimensioning

In the case of a memoryless channel (i.e. no delay spread), OFDM is insensitive to timing synchronization errors provided that the misalignment remains within the CP duration. In other words, if $T_{\mathrm{o}} \leq T_{\mathrm{CP}}$ (with T_{o} being the timing error), then orthogonality is maintained thanks to the cyclic nature of the CP. Any symbol timing delay only introduces a constant

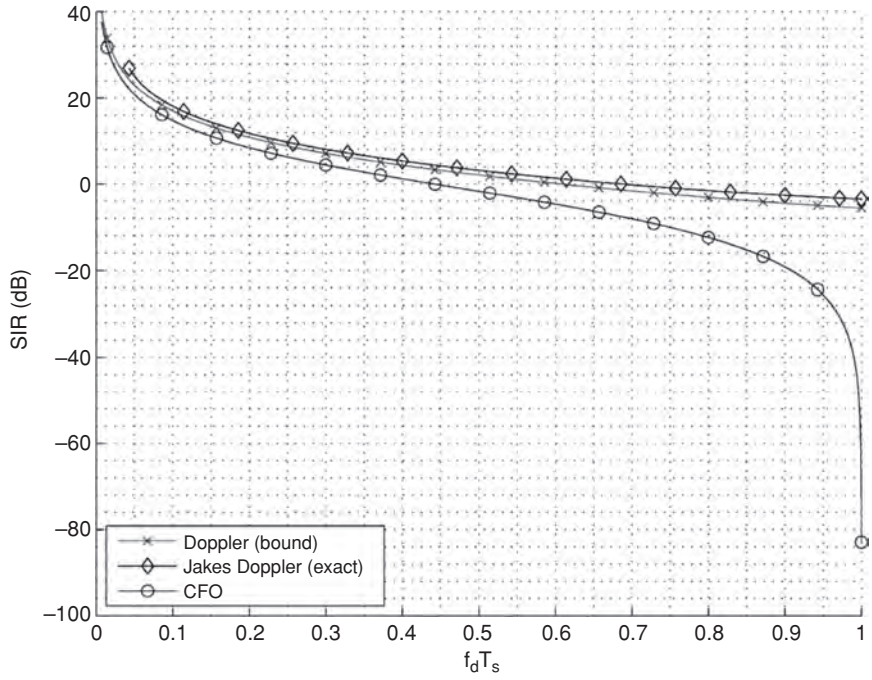


Figure 5.9: SIR_{ICI} for classical Doppler distribution and deterministic CFO.

phase shift from one subcarrier to another. The received signal at the m-th subcarrier is given by

$$R_{\mathrm{m}}[k] = X_{\mathrm{m}}[k] \exp\left(j2\pi \frac{dm}{N}\right) \tag{5.19}$$

where d is the timing offset in samples corresponding to a duration equal to T_{o}. This phase shift can be recovered as part of the channel estimation operation.

It is worth highlighting that the insensitivity to timing offsets would not hold for any kind of guard period other than a cyclic prefix; for example, zero-padding would not exhibit the same property, resulting in a portion of the useful signal power being lost.

In the general case of a channel with delay spread, for a given CP length, the maximum tolerated timing offset without degrading the OFDM reception is reduced by an amount equal to the length of the channel impulse response: $T_{\mathrm{o}} \leq T_{\mathrm{CP}} - T_{\mathrm{d}}$. For greater timing errors, ISI and ICI occur. The effect caused by an insufficient CP is discussed in the following section. Timing synchronization hence becomes more critical in long-delay-spread channels.

Initial timing acquisition in LTE is normally achieved by the cell-search and synchronization procedures (see Chapter 7). Thereafter, for continuous tracking of the timing-offset, two classes of approach exist, based on either CP correlation or Reference Signals (RSs). A combination of the two is also possible. The reader is referred to [19] for a comprehensive survey of OFDM synchronization techniques.

5.2.4.1 Effect of Insufficient Cyclic Prefix Length

As already explained, if an OFDM system is designed with a CP of length G samples such that $L < G$ where L is the length of the channel impulse response (in number of samples), the system benefits from turning the linear convolution into a circular one to keep the subcarriers orthogonal. The condition of a sufficient CP is therefore strictly related to the orthogonality property of OFDM.

As shown in [20], for an OFDM symbol consisting of $N + G$ samples where N is the FFT size, the power of the ICI and ISI can be computed as

$$P_{\text{ICI}} = 2 \sum_{k=G}^{N+G-1} |h[k]|^2 \frac{N(k-G) - (k-G)^2}{N^2} \tag{5.20}$$

$$P_{\text{ISI}} = \sum_{k=G}^{N+G-1} |h[k]|^2 \frac{(k-G)^2}{N^2} \tag{5.21}$$

The signal power P_{S} can be written as

$$P_{\text{S}} = \sum_{k=0}^{G-1} |h(k)|^2 + \sum_{k=G}^{N+G-1} |h(k)|^2 \frac{(N-k+G)^2}{N^2} \tag{5.22}$$

The resulting SIR due to the CP being too short can then be written as

$$SIR_{\text{og}} = \frac{P_{\text{S}}}{P_{\text{ISI}} + P_{\text{ICI}}} \tag{5.23}$$

Figures 5.10 and 5.11 plot Equations (5.21) to (5.23) for the case of the normal CP length in LTE (see Section 5.4) assuming a channel with a uniform and normalized Power-Delay Profile (PDP) of length $L < N + G$, where the dashed line marks the boundary of the CP ($L = G$).

5.3 OFDMA

Orthogonal Frequency Division Multiple Access (OFDMA) is an extension of OFDM to the implementation of a multiuser communication system. In the discussion above, it has been assumed that a single user receives data on all the subcarriers at any given time. OFDMA distributes subcarriers to different users at the same time, so that multiple users can be scheduled to receive data simultaneously. Usually, subcarriers are allocated in contiguous groups for simplicity and to reduce the overhead of indicating which subcarriers have been allocated to each user.

OFDMA for mobile communications was first proposed in [21] based on multicarrier FDMA (Frequency Division Multiple Access), where each user is assigned to a set of randomly selected subchannels.

OFDMA enables the OFDM transmission to benefit from multi-user diversity, as discussed in Chapter 12. Based on feedback information about the frequency-selective channel conditions from each user, adaptive user-to-subcarrier assignment can be performed, enhancing considerably the total system spectral efficiency compared to single-user OFDM systems.

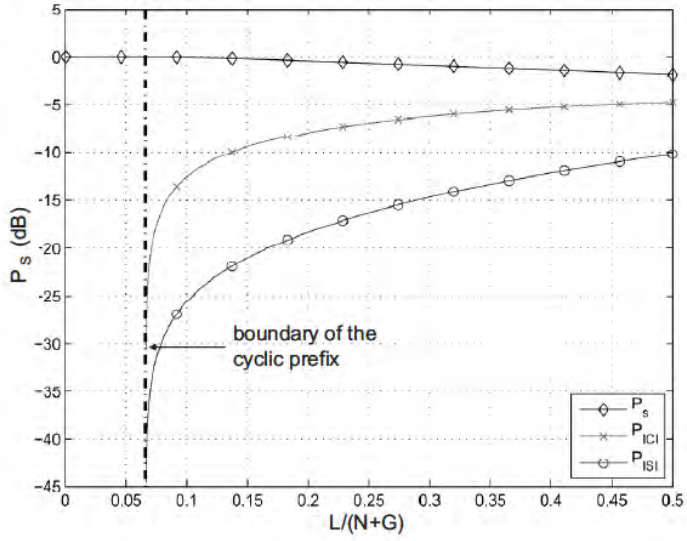


Figure 5.10: Power of signal, ICI and ISI in case of a too-short CP.

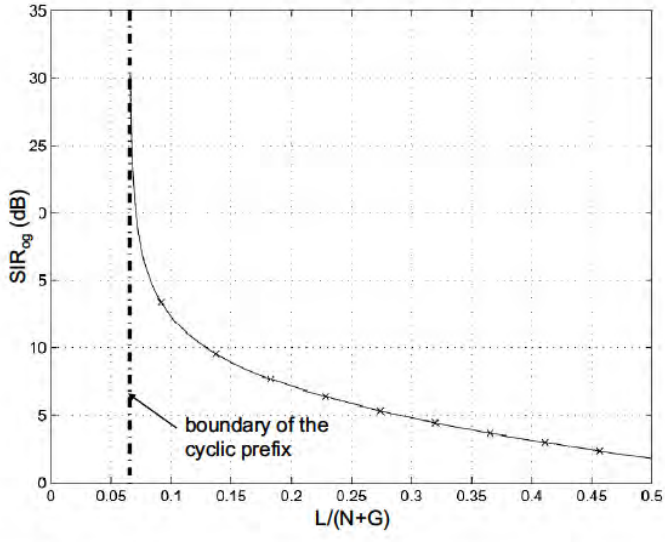


Figure 5.11: Effective SIR as a function of channel impulse response length for a given CP length.

OFDMA can also be used in combination with Time Division Multiple Access (TDMA), such that the resources are partitioned in the time-frequency plane – i.e. groups of subcarriers for a specific time duration. In LTE, such time-frequency blocks are known as Resource Blocks (RBs), as explained in Section 6.2. Figure 5.12 depicts such an OFDMA/TDMA mixed strategy as used in LTE.

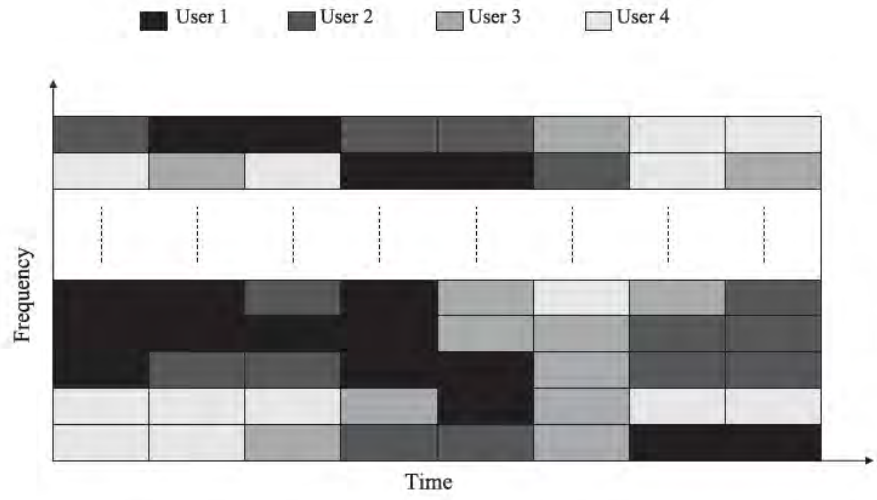


Figure 5.12: Example of resource allocation in a combined OFDMA/TDMA system.

5.4 Parameter Dimensioning

As highlighted in the previous sections, certain key parameters determine the performance of OFDM and OFDMA systems. Inevitably, some compromises have to be made in defining these parameters appropriately to maximize system spectral efficiency while maintaining robustness against propagation impairments.

For a given system, the main propagation characteristics which should be taken into account when designing an OFDM system are the expected delay spread T_d, the maximum Doppler frequency $f_{d_{max}}$, and, in the case of cellular systems, the targeted cell size.

The propagation characteristics impose constraints on the choice of the CP length and of the subcarrier spacing.

As already mentioned, the CP should be longer than the Channel Impulse Response (CIR) in order to ensure robustness against ISI. For cellular systems, and especially for large cells, longer delay spreads may typically be experienced than those encountered, for example, in WLAN systems, implying the need for a longer CP. On the other hand, a longer CP for a given OFDM symbol duration corresponds to a larger overhead in terms of energy per transmitted

bit. Out of the $N + G$ transmitted symbols, only N convey information, leading to a rate loss. This reduction in bandwidth efficiency can be expressed as a function of the CP duration $T_{\mathrm{CP}} = GT_{\mathrm{s}}$ and the OFDM symbol period $T_{\mathrm{u}} = NT_{\mathrm{s}}$ (where T_{s} is the sampling period), as follows:

$$\beta_{\mathrm{overhead}} = \frac{T_{\mathrm{CP}}}{T_{\mathrm{u}} + T_{\mathrm{CP}}} \tag{5.24}$$

It is clear that to maximize spectral efficiency, T_{u} should be chosen to be large relative to the CP duration, but small enough to ensure that the channel does not vary within one OFDM symbol.

Further, the OFDM symbol duration T_{u} is related to the subcarrier spacing by $\Delta f = 1/T_{\mathrm{u}}$. Choosing a large T_{u} leads to a smaller subcarrier separation Δf, which has a direct impact on the system sensitivity to Doppler and other sources of frequency offset, as explained in Section 5.2.3.

Thus, in summary, the following three design criteria can be identified:

$$\begin{aligned} &T_{\mathrm{CP}} \geq T_{\mathrm{d}} && \text{to prevent ISI,} \\ &\frac{f_{\mathrm{d_{max}}}}{\Delta f} \ll 1 && \text{to keep ICI due to Doppler sufficiently low,} \\ &T_{\mathrm{CP}}\Delta f \ll 1 && \text{for spectral efficiency.} \end{aligned} \tag{5.25}$$

5.4.1 Physical Layer Parameters for LTE

LTE aims at supporting a wide range of cellular deployment scenarios, including indoor, urban, suburban and rural situations covering both low and high UE mobility conditions (up to 350 or even 500 km/h). The cell sizes may range from home networks only a few metres across to large cells with radii of 100 kilometres or more.

Typical deployed carrier frequencies are in the range 400 MHz to 4 GHz, with bandwidths ranging from 1.4 to 20 MHz. All these cases imply different delay spreads and Doppler frequencies.

The 'normal' parameterization of the LTE downlink uses a $\Delta f = 15$ kHz subcarrier spacing with a CP length of approximately 5 μs. This subcarrier spacing is a compromise between the percentage overhead of the CP and the sensitivity to frequency offsets. A 15 kHz subcarrier spacing is sufficiently large to allow for high mobility and to avoid the need for closed-loop frequency adjustments.

In addition to the normal parameterization, it is possible to configure LTE with an extended CP of length approximately 17 μs.[3] This is designed to ensure that even in large suburban and rural cells, the delay spread should be contained within the CP. This, however, comes at the expense of a higher overhead from the CP as a proportion of the total system transmission resources. This is particularly useful for the multi-cell broadcast transmission mode supported by LTE known as Multimedia Broadcast Single Frequency Network (MBSFN) (see Section 13.4), in which the UE receives and combines synchronized signals from multiple cells. In this case the relative timing offsets from the multiple cells must all be received at the UE's receiver within the CP duration, if ISI is to be avoided, thus also requiring a long CP.

[3] The length of the extended CP is $\frac{1}{4}$ of the OFDM symbol.

In case MBSFN transmission were, in a future release of LTE, to be configured on a dedicated carrier rather than sharing the carrier with unicast data (this is not possible in current releases of LTE), a further set of parameters is defined with the subcarrier spacing halved to 7.5 kHz; this would allow the OFDM symbol length to be doubled to provide an extended CP of length approximately 33 μs while remaining $\frac{1}{4}$ of the OFDM symbol length. This would come at the expense of increasing the sensitivity to mobility and frequency errors.

These modes and their corresponding parameters are summarized in Figure 5.13. It is worth noting that when LTE is configured with the normal CP length, the CP length for the first OFDM symbol in each 0.5 ms interval is slightly longer than that of the next six OFDM symbols (i.e. 5.2 μs compared to 4.7 μs). This characteristic is due to the need to accommodate an integer number of OFDM symbols, namely 7, into each 0.5 ms interval, with assumed FFT block-lengths of 2048.

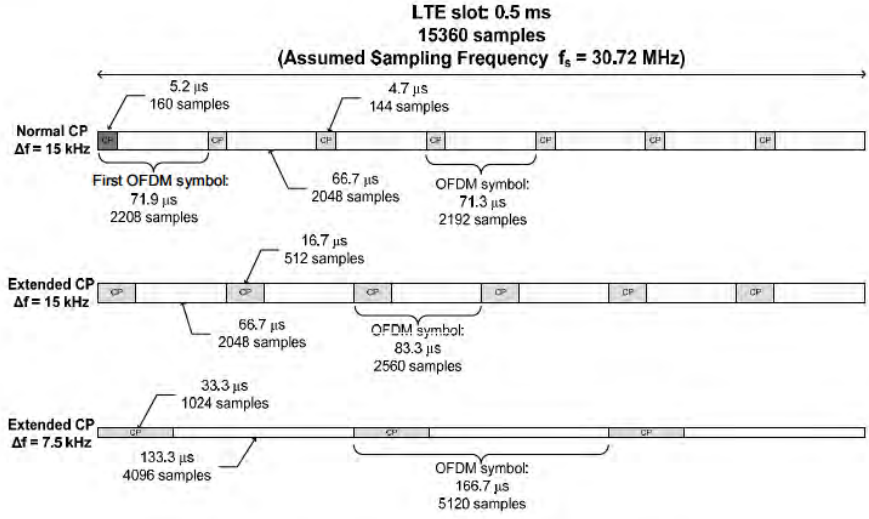


Figure 5.13: LTE OFDM symbol and CP lengths.

The actual FFT size and sampling frequency for the LTE downlink are not specified. However, the above parameterizations are designed to be compatible with a sampling frequency of 30.72 MHz. Thus, the basic unit of time in the LTE specifications, of which all other time periods are a multiple, is defined as $T_s = 1/30.72$ μs. This is itself chosen for backward compatibility with UMTS, for which the chip rate is 3.84 MHz – exactly one eighth of the assumed LTE sampling frequency.

In the case of a 20 MHz system bandwidth, an FFT order of 2048 may be assumed for efficient implementation. However, in practice the implementer is free to use other DFT sizes.

Lower sampling frequencies (and proportionally lower FFT orders) are always possible to reduce RF and baseband processing complexity for narrower bandwidth deployments: for

example, for a 5 MHz system bandwidth the FFT order and sampling frequency could be scaled down to 512 and $f_s = 7.68$ MHz respectively, while only 300 subcarriers are actually modulated with data.

For the sake of simplifying terminal implementation, the direct current (d.c.) subcarrier is left unused, in order to avoid d.c. offset errors in direct conversion receivers.

The OFDMA parameters used in the downlink are defined in the 3GPP Technical Specification 36.211 [22, Section 6].

5.5 Summary

In this chapter, we have reviewed the key features, benefits and sensitivities of OFDM and OFDMA systems. In summary, it can be noted that:

- OFDM is a mature technology.
- It is already widely deployed and is especially suited for broadcast or downlink applications because of the low receiver complexity while requiring a higher transmitter complexity (expensive PA). The low receiver complexity also makes it well-suited to MIMO schemes.
- It benefits from efficient implementation by means of the FFT.
- It achieves the high transmission rates of broadband transmission, with low receiver complexity.
- It makes use of a CP to avoid ISI, enabling block-wise processing.
- It exploits orthogonal subcarriers to avoid the spectrum wastage associated with inter-subcarrier guard-bands.
- The parameterization allows the system designer to balance tolerance of Doppler and delay spread depending on the deployment scenario.
- It can be extended to a multiple-access scheme, OFDMA, in a straightforward manner.

These factors together have made OFDMA the technology of choice for the LTE downlink.

References[4]

[1] R. W. Chang, 'Synthesis of Band-limited Orthogonal Signals for Multichannel Data Transmission'. *Bell Systems Technical Journal*, Vol. 46, pp. 1775–1796, December 1966.

[2] B. R. Saltzberg, 'Performance of an Efficient Parallel Data Transmission System'. *IEEE Trans. on Communications*, Vol. 15, pp. 805–811, December 1967.

[3] S. B. Weinstein and P. M. Ebert, 'Data Transmission by Frequency-Division Multiplexing using the Discrete Fourier Transform'. *IEEE Trans. on Communications*, Vol. 19, pp. 628–634, October 1971.

[4] A. Peled and A. Ruiz, 'Frequency Domain Data Transmission using Reduced Computational Complexity Algorithms' in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 5, pp. 964–967, April 1980.

[4] All web sites confirmed 1st March 2011.

[5] L. J. Cimini, 'Analysis and Simulation of Digital Mobile Channel using Orthogonal Frequency Division Multiplexing'. *IEEE Trans. on Communications*, Vol. 33, pp. 665–675, July 1985.

[6] G. H. Golub and C. F. Van Loan, *Matrix Computations*. John Hopkins University Press, 1996.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[8] S. N. Bernstein, 'On the Work of P. L. Chebyshev in Probability Theory'. *The Scientific Legacy of P. L. Chebyshev. First Part: Mathematics*, edited by S. N. Bernstein. Academiya Nauk SSSR, Moscow-Leningrad, p. 174, 1945.

[9] T. Henk, *Understanding Probability: Chance Rules in Everyday Life*. Cambridge University Press, 2004.

[10] X. Li and L. J. Cimini, 'Effects of Clipping and Filtering on the Performance of OFDM'. *IEEE Comm. Lett.*, Vol. 2, pp. 131–133, May 1998.

[11] L. Wane and C. Tellambura, 'A Simplified Clipping and Filtering Technique for PAPR Reduction in OFDM Systems'. *IEEE Sig. Proc. Lett.*, Vol. 12, pp. 453–456, June 2005.

[12] J. Armstrong, 'Peak to Average Power Reduction for OFDM by Repeated Clipping and Frequency Domain Filtering'. *Electronics Letters*, Vol. 38, pp. 246–247, February 2002.

[13] A. Jayalath, 'OFDM for Wireless Broadband Communications (Peak Power Reduction, Spectrum and Coding)'. PhD thesis, School of Computer Science and Software Engineering, Monash University, 2002.

[14] A. Jones, T. Wilkinson, and S. Barton, 'Block Coding Scheme for Reduction of Peak to Mean Envelope Power Ratio of Multicarrier Transmission Schemes'. *Electronics Letters*, Vol. 30, pp. 2098–2099, December 1994.

[15] C. Tellambura, 'Use of M-sequence for OFDM Peak-to-Average Power Ratio Reduction'. *Electronics Letters*, Vol. 33, pp. 1300–1301, July 1997.

[16] Y. Li and L. J. Cimini, 'Bounds on the Interchannel Interference of OFDM in Time-Varying Impairments'. *IEEE Trans. on Communications*, Vol. 49, pp. 401–404, March 2001.

[17] X. Cai and G. B. Giannakis, 'Bounding Performance and Suppressing Intercarrier Interference in Wireless Mobile OFDM'. *IEEE Trans. on Communications*, Vol. 51, pp. 2047–2056, March 2003.

[18] J. Heiskala and J. Terry, *OFDM Wireless LANs: A Theoretical and Practical Guide*. SAMS Publishing, 2001.

[19] L. Hanzo and T. Keller, *OFDM And MC-CDMA: A Primer*. Wiley-IEEE Press, 2006.

[20] A. G. Burr, 'Irreducible BER of COFDM on IIR channel'. *Electronics Letters*, Vol. 32, pp. 175–176, February 1996.

[21] R. Nogueroles, M. Bossert, A. Donder, and V. Zyablov, 'Improved Performance of a Random OFDMA Mobile Communication System' in *Proc. IEEE Vehicular Technology Conference*, Vol. 3, pp. 2502–2506, May 1998.

[22] 3GPP Technical Specification 36.311, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation', www.3gpp.org.

6

Introduction to Downlink Physical Layer Design

Matthew Baker

6.1 Introduction

The LTE downlink transmissions from the eNodeB consist of user-plane and control-plane data from the higher layers in the protocol stack (as described in Chapters 3 and 4) multiplexed with physical layer signalling to support the data transmission. The multiplexing of these various parts of the downlink signal is facilitated by the Orthogonal Frequency Division Multiple Access (OFDMA) structure described in Chapter 5, which enables the downlink signal to be subdivided into small units of time and frequency.

This subdivided structure is introduced below, together with an outline of the general steps in forming the transmitted downlink signal in the physical layer.

6.2 Transmission Resource Structure

The downlink transmission resources in LTE possess dimensions of time, frequency and space. The spatial dimension, measured in 'layers', is accessed by means of multiple 'antenna ports' at the eNodeB; for each antenna port a Reference Signal (RS) is provided to enable the User Equipment (UE) to estimate the radio channel (see Section 8.2); the techniques for using multiple antenna ports to exploit multiple spatial layers are explained in Section 11.2.

The time-frequency resources for each transmit antenna port are subdivided according to the following structure: the largest unit of time is the 10 ms radio frame, which is subdivided into ten 1 ms subframes, each of which is split into two 0.5 ms slots. Each slot comprises seven OFDM symbols in the case of the normal Cyclic Prefix (CP) length, or six if the

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

extended CP is configured in the cell as explained in Section 5.4.1. In the frequency domain, resources are grouped in units of 12 subcarriers (thus occupying a total of 180 kHz with a subcarrier spacing of 15 kHz), such that one unit of 12 subcarriers for a duration of one slot is termed a Resource Block (RB).[1]

The smallest unit of resource is the Resource Element (RE), which consists of one subcarrier for a duration of one OFDM symbol. An RB thus comprises 84 REs in the case of the normal cyclic prefix length, and 72 REs in the case of the extended cyclic prefix.

The detailed resource structure is shown in Figure 6.1 for the normal cyclic prefix length.

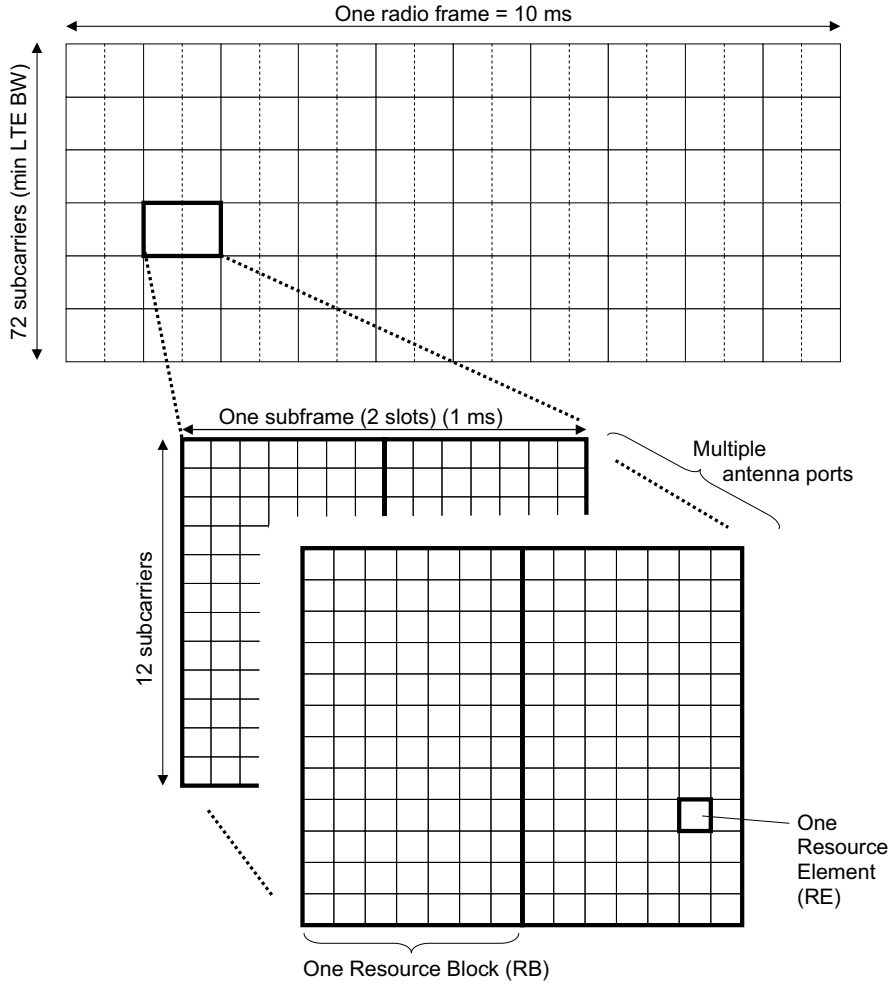


Figure 6.1: Basic time-frequency resource structure of LTE (normal cyclic prefix case).

[1]In the case of the 7.5 kHz subcarrier spacing which may be available for Multimedia Broadcast Multicast Service (MBMS) transmission in later releases of LTE (see Section 5.4.1 and Chapter 13), one RB consists of 24 subcarriers for a duration of one slot.

Within certain RBs, some REs are reserved for special purposes: synchronization signals (Chapter 7), Reference Signals (RSs – see Chapter 8), control signalling and critical broadcast system information (Chapter 9). The remaining REs are used for data transmission, and are usually allocated in pairs of RBs (the pairing being in the time domain).

The structure shown in Figure 6.1 assumes that all subframes are available for downlink transmission. This is known as 'Frame Structure Type 1' and is applicable for Frequency Division Duplexing (FDD) in paired radio spectrum, or for a standalone downlink carrier. For Time Division Duplexing (TDD) in unpaired radio spectrum, the basic structure of RBs and REs remains the same, but only a subset of the subframes are available for downlink transmission; the remaining subframes are used for uplink transmission, or for special subframes which contain a guard period to allow for switching between downlink and uplink transmission. The guard period allows the uplink transmission timing to be advanced as described in Section 18.2. This TDD structure is known as 'Frame Structure Type 2', of which seven different configurations are defined, as shown in Figure 6.2; these allow a variety of downlink-uplink ratios and switching periodicities. Further details of TDD operation using this frame structure are described in Chapter 23.

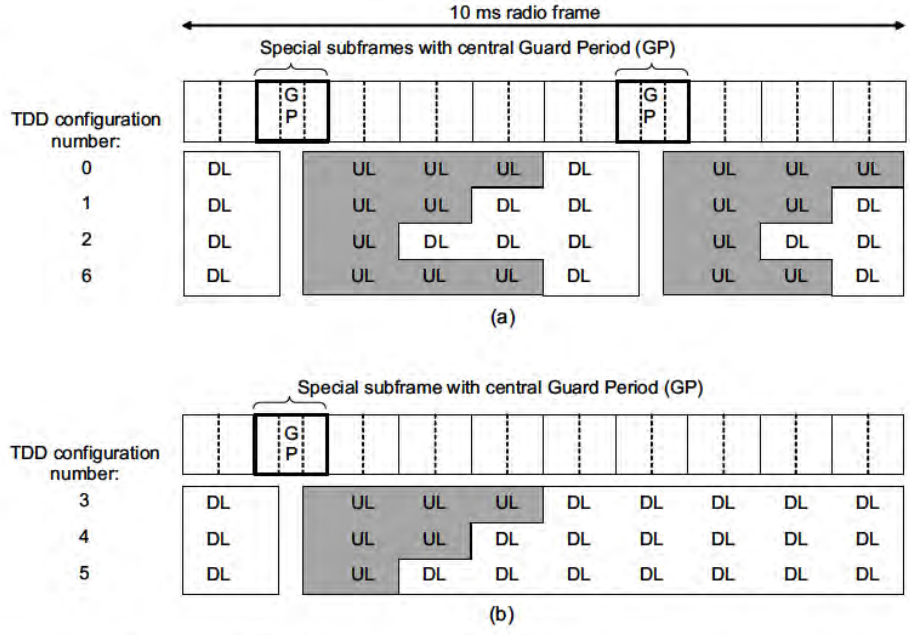


Figure 6.2: LTE subframe structure for TDD operation:
(a) configurations with 5 ms periodicity of switching from downlink (DL) to uplink (UL);
(b) configurations with 10 ms periodicity of switching from downlink (DL) to uplink (UL).

6.3 Signal Structure

The role of the physical layer is primarily to translate data into a reliable signal for transmission across the radio interface between the eNodeB and the User Equipment (UE). Each block of data is first protected against transmission errors, usually first with a Cyclic Redundancy Check (CRC), and then with channel coding, to form a *codeword* (see Chapter 10). After channel coding, the steps in the formation of the downlink LTE signal on a given carrier are illustrated in Figure 6.3.

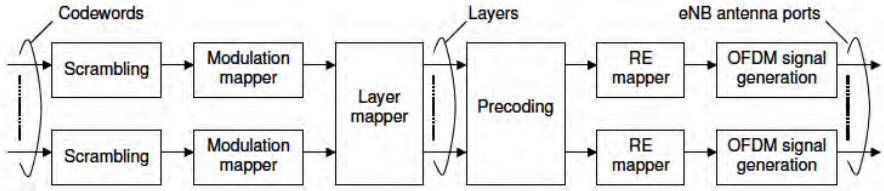


Figure 6.3: General signal structure for LTE downlink.
Reproduced by permission of © 3GPP.

The initial scrambling stage is applied to all downlink physical channels, and serves the purpose of interference rejection. The scrambling sequence in all cases uses an order-31 Gold code, which can provide 2^{31} sequences which are not cyclic shifts of each other. Gold codes [1, 2] also possess the attractive feature that they can be generated with very low implementation complexity, as they can be derived from the modulo-2 addition of two maximum-length sequences (otherwise known as *M-sequences*), which can be generated from a simple shift-register.[2] A shift-register implementation of the LTE scrambling sequence generator is illustrated in Figure 6.4.

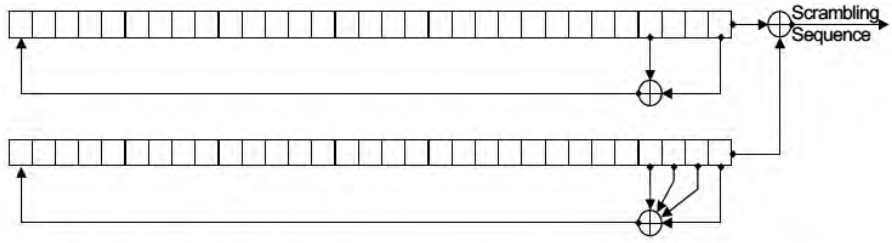


Figure 6.4: Shift-register implementation of scrambling sequence generator.

The scrambling sequence generator is re-initialized every subframe (except for the Physical Broadcast CHannel (PBCH), which is discussed in Section 9.2.1), based on the

[2]Gold Codes were also used in WCDMA (Wideband Code Division Multiple Access), for the uplink long scrambling codes.

identity of the cell, the subframe number (within a radio frame) and the UE identity. This randomizes interference between cells and between UEs. In addition, in cases where multiple data streams (codewords) are transmitted via multiple layers (see Table 11.2), the identity of the codeword is also used in the initialization.

As a useful feature for avoiding unnecessary complexity, the scrambling sequence generator described here is the same as for the pseudo-random sequence used for the RSs as described in Chapter 8, the only difference being in the method of initialization; in all cases, however, a fast-forward of 1600 places is applied at initialization, in order to ensure low cross-correlation between sequences used in adjacent cells.

Following the scrambling stage, the data bits from each channel are mapped to complex-valued modulation symbols depending on the relevant modulation scheme, then mapped to layers and precoded as explained in Chapter 11, mapped to REs, and finally translated into a complex-valued OFDM signal by means of an Inverse Fast Fourier Transform (IFFT).

6.4 Introduction to Downlink Operation

In order to communicate with an eNodeB supporting one or more cells, the UE must first identify the downlink transmission from one of these cells and synchronize with it. This is achieved by means of special synchronization signals which are embedded into the OFDM structure described above. The procedure for cell search and synchronization is described in Chapter 7.

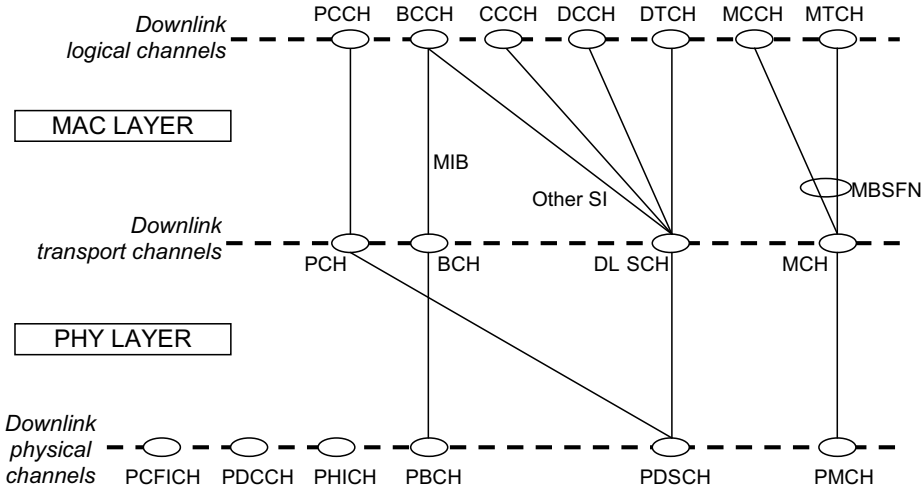


Figure 6.5: Summary of downlink physical channels and mapping to higher layers.

The next step for the UE is to estimate the downlink radio channel in order to be able to perform coherent demodulation of the information-bearing parts of the downlink signal. Some suitable techniques for channel estimation are described in Chapter 8, based on the reference signals which are inserted into the downlink signal.

Chapter 9 describes the parts of the downlink signal which carry data originating from higher protocol layers, including the PBCH, the Physical Downlink Shared Channel (PDSCH) and, in the case of MBMS transmission from Release 9 onwards, the Physical Multicast CHannel (PMCH). In addition, the design of the downlink control signalling is explained, including its implications for the ways in which downlink transmission resources may be allocated to different users for data transmission.

The downlink physical channels described in the following chapters are summarized in Figure 6.5, together with their relationship to the higher-layer channels.

The subsequent chapters explain the key techniques which enable these channels to make efficient use of the radio spectrum: channel coding and link adaptation are explained in Chapter 10, the LTE schemes for exploiting multiple antennas are covered in Chapter 11, and techniques for effective scheduling of transmission resources to multiple users are described in Chapter 12.

References[3]

[1] W. Huffman and V. Pless, *Handbook of Coding Theory, vol II*. Amsterdam: North-Holland, 1998.

[2] R. McEliece, *Finite Fields for Computer Scientists and Engineers*. Boston, MA: Kluwer Academic Publishers, 2003.

[3] All web sites confirmed 1st March 2011.

7

Synchronization and Cell Search

Fabrizio Tomatis and Stefania Sesia

7.1 Introduction

A User Equipment (UE) wishing to access an LTE cell must first undertake a *cell search procedure*. This chapter focuses on the aspects of the physical layer that are designed to facilitate cell search. The way this relates to the overall mobility functionality and protocol aspects for cell reselection and handover is explained in Chapters 3 and 22. The performance requirements related to cell search and synchronization are explained in Section 22.2.

At the physical layer, the cell search procedure consists of a series of synchronization stages by which the UE determines time and frequency parameters that are necessary to demodulate the downlink and to transmit uplink signals with the correct timing. The UE also acquires some critical system parameters.

Three major synchronization requirements can be identified in the LTE system:

1. Symbol and frame timing acquisition, by which the correct symbol start position is determined, for example to set the Discrete Fourier Transform (DFT) window position;
2. Carrier frequency synchronization, which is required to reduce or eliminate the effect of frequency errors[1] arising from a mismatch of the local oscillators between the transmitter and the receiver, as well as the Doppler shift caused by any UE motion;
3. Sampling clock synchronization.

7.2 Synchronization Sequences and Cell Search in LTE

The cell search procedure in LTE begins with a synchronization procedure which makes use of two specially designed physical signals that are broadcast in each cell: the Primary Synchronization Signal (PSS) and the Secondary Synchronization Signal (SSS). The detection of

[1] Frequency offsets may arise from factors such as temperature drift, ageing and imperfect calibration.

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

these two signals not only enables time and frequency synchronization, but also provides the UE with the physical layer identity of the cell and the cyclic prefix length, and informs the UE whether the cell uses Frequency Division Duplex (FDD) or Time Division Duplex (TDD).

In the case of initial synchronization (when the UE is not already camping on or connected to an LTE cell) after detecting the synchronization signals, the UE decodes the Physical Broadcast CHannel (PBCH), from which critical system information is obtained (see Section 9.2.1). In the case of neighbour cell identification, the UE does not need to decode the PBCH; it simply makes quality-level measurements based on the reference signals (see Chapter 8) transmitted from the newly detected cell and uses them for cell reselection (in RRC_IDLE state) or handover (in RRC_CONNECTED state); in the latter case, the UE reports these measurements to its serving cell.

The cell search and synchronization procedure is summarized in Figure 7.1, showing the information ascertained by the UE at each stage. The PSS and SSS structure is specifically designed to facilitate this acquisition of information.

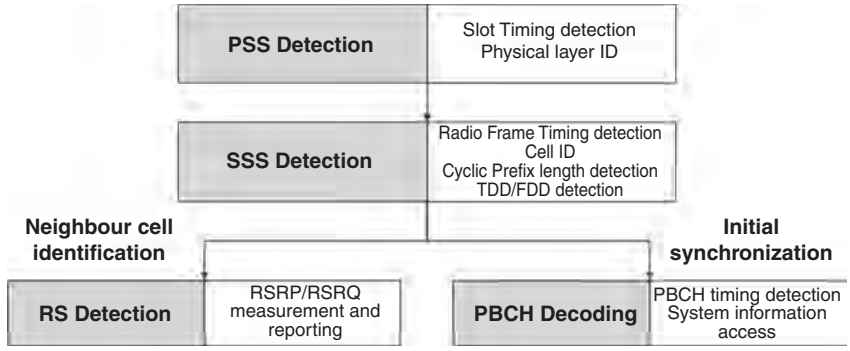


Figure 7.1: Information acquired at each step of the cell search procedure.

The PSS and SSS structure in time is shown in Figure 7.2 for the FDD case and in Figure 7.3 for TDD: the synchronization signals are transmitted periodically, twice per 10 ms radio frame. In an FDD cell, the PSS is always located in the last OFDM (Orthogonal Frequency Division Multiplexing) symbol of the first and 11th slots of each radio frame (see Chapter 6), thus enabling the UE to acquire the slot boundary timing independently of the Cyclic Prefix (CP) length. The SSS is located in the symbol immediately preceding the PSS, a design choice enabling coherent detection[2] of the SSS relative to the PSS, based on the assumption that the channel coherence duration is significantly longer than one OFDM symbol. In a TDD cell, the PSS is located in the third symbol of the 3rd and 13th slots, while the SSS is located three symbols earlier; coherent detection can be used under the assumption that the channel coherence time is significantly longer than four OFDM symbols.

The precise position of the SSS changes depending on the length of the CP configured for the cell. At this stage of the cell detection process, the CP length is unknown a priori

[2]See Section 7.3.

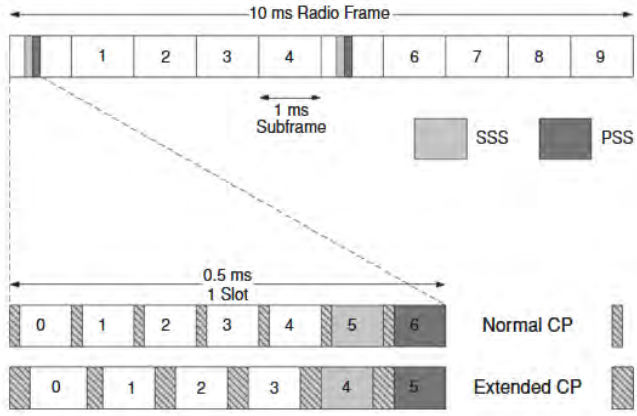


Figure 7.2: PSS and SSS frame and slot structure in time domain in the FDD case.

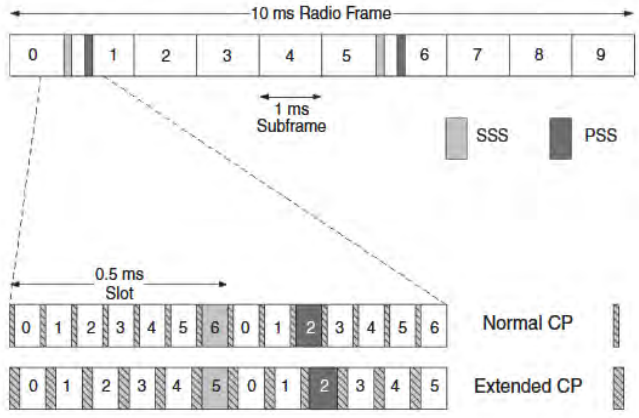


Figure 7.3: PSS and SSS frame and slot structure in time domain in the TDD case.

to the UE, and it is, therefore, blindly detected by checking for the SSS at the two possible positions.[3]

[3]Hence a total of four possible SSS positions must be checked if the UE is searching for both FDD and TDD cells.

While the PSS in a given cell is the same in every subframe in which it is transmitted, the two SSS transmissions in each radio frame change in a specific manner as described in Section 7.2.3, thus enabling the UE to establish the position of the 10 ms radio frame boundary.

In the frequency domain, the mapping of the PSS and SSS to subcarriers is shown in Figure 7.4. The PSS and SSS are transmitted in the central six Resource Blocks[4] (RBs), enabling the frequency mapping of the synchronization signals to be invariant with respect to the system bandwidth (which can in principle vary from 6 to 110 RBs to suit channel bandwidths between around 1.4 MHz and 20 MHz); this allows the UE to synchronize to the network without any a priori knowledge of the allocated bandwidth. The PSS and SSS are each comprised of a sequence of length 62 symbols, mapped to the central 62 subcarriers around the d.c. subcarrier, which is left unused. This means that the five resource elements at each extremity of each synchronization sequence are not used. This structure enables the UE to detect the PSS and SSS using a size-64 Fast Fourier Transform (FFT) and a lower sampling rate than would have been necessary if all 72 subcarriers were used in the central six resource blocks. The shorter length for the synchronization sequences also avoids the possibility in a TDD system of a high correlation with the uplink demodulation reference signals which use the same kind of sequence as the PSS (see Chapter 15).

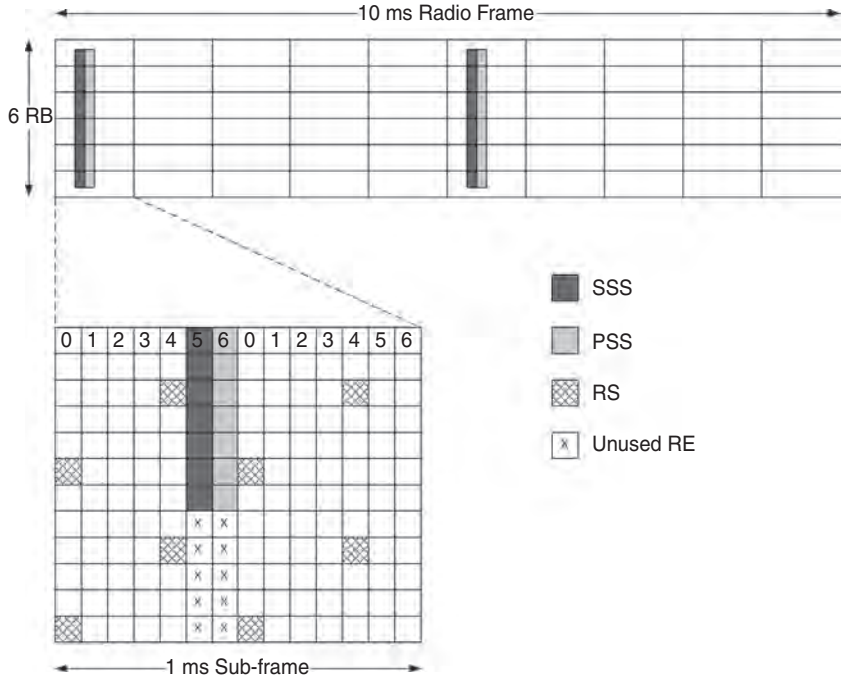


Figure 7.4: PSS and SSS frame structure in frequency and time domain for an FDD cell.

[4]Six Resource Blocks correspond to 72 subcarriers.

In the case of multiple transmit antennas being used at the eNodeB, the PSS and SSS are always transmitted from the same antenna port[5] in any given subframe, while between different subframes they may be transmitted from different antenna ports in order to benefit from time-switched antenna diversity.

The particular sequences which are transmitted for the PSS and SSS in a given cell are used to indicate the physical layer cell identity to the UE. There are 504 unique Physical Cell Identities (PCIs) in LTE, grouped into 168 groups of three identities. The three identities in a group would usually be assigned to cells under the control of the same eNodeB. Three PSS sequences are used to indicate the cell identity within the group, and 168 SSS sequences are used to indicate the identity of the group.[6]

The PSS uses sequences known as *Zadoff–Chu*. This kind of sequence is widely used in LTE, including for the uplink reference signals and random access preambles (see Chapters 15 and 17) in addition to the PSS. Therefore, Section 7.2.1 is devoted to an explanation of the fundamental principles behind Zadoff–Chu sequences, before discussing the specific constructions of the PSS and SSS sequences in the subsequent sections.

7.2.1 Zadoff–Chu Sequences

Zadoff–Chu (ZC) sequences (also known as *Generalized Chirp-Like* (GCL) sequences) are named after the authors of [1] and [2]. ZC sequences are non-binary unit-amplitude sequences [3], which satisfy a Constant Amplitude Zero Autocorrelation (CAZAC) property. CAZAC sequences are complex signals of the form $e^{j\alpha_k}$. The ZC sequence of odd-length N_{ZC} is given by

$$a_q(n) = \exp\left[-j2\pi q \frac{n(n+1)/2 + ln}{N_{\mathrm{ZC}}}\right] \tag{7.1}$$

where $q \in \{1, \ldots, N_{\mathrm{ZC}} - 1\}$ is the ZC sequence root index, $n = 0,\ 1, \ldots, N_{\mathrm{ZC}} - 1$, $l \in \mathbb{N}$ is any integer. In LTE, $l = 0$ is used for simplicity.

ZC sequences have the following three important properties:

Property 1. A ZC sequence has constant amplitude, and its N_{ZC}-point DFT also has constant amplitude. The constant amplitude property limits the Peak-to-Average Power Ratio (PAPR) and generates bounded and time-flat interference to other users. It also simplifies the implementation as only phases need to be computed and stored, not amplitudes.

Property 2. ZC sequences of any length have 'ideal' cyclic autocorrelation (i.e. the correlation with its circularly shifted version is a delta function). The zero autocorrelation property may be formulated as:

$$r_{kk}(\sigma) = \sum_{n=0}^{N_{\mathrm{ZC}}-1} a_k(n) a_k^*[(n+\sigma)] = \delta(\sigma) \tag{7.2}$$

where $r_{kk}(\cdot)$ is the discrete periodic autocorrelation function of a_k at lag σ. This property is of major interest when the received signal is correlated with a reference sequence and the received reference sequences are misaligned. As an example, Figure 7.5 shows the difference

[5] The concept of an antenna port in LTE is explained in Section 8.2.

[6] A subset of the available PCIs can be reserved for Closed Subscriber Group (CSG) cells (e.g. Home eNodeBs). If this is the case, the reserved PCIs are indicated in System Information Block 4 (SIB4), enabling the UE to eliminate those PCIs from its search if CSG cells are not applicable for it.

between the periodic autocorrelation of a truncated Pseudo-Noise (PN) sequence (as used in WCDMA [4]) and a ZC sequence. Both are 839 symbols long in this example. The ZC periodic autocorrelation is exactly zero for $\sigma \neq 0$ and it is non-zero for $\sigma = 0$, whereas the PN periodic autocorrelation shows significant peaks, some above 0.1, at non-zero lags.

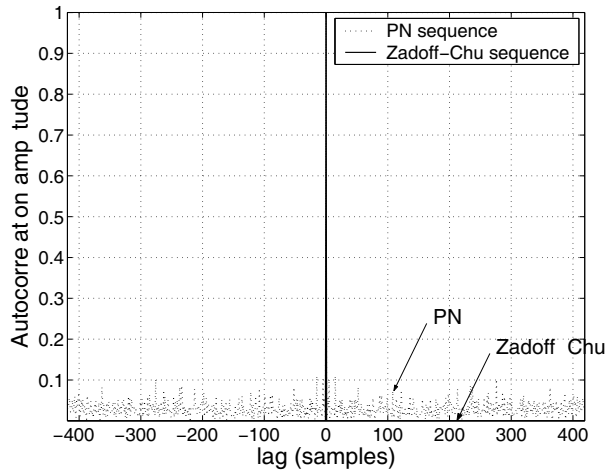


Figure 7.5: Zadoff–Chu versus PN sequence: periodic autocorrelation.

The main benefit of the CAZAC property is that it allows multiple orthogonal sequences to be generated from the same ZC sequence. Indeed, if the periodic autocorrelation of a ZC sequence provides a single peak at the zero lag, the periodic correlation of the same sequence against its cyclic shifted replica provides a peak at lag N_{CS}, where N_{CS} is the number of samples of the cyclic shift. This creates a Zero-Correlation Zone (ZCZ) between the two sequences. As a result, as long as the ZCZ is dimensioned to cope with the largest possible expected time misalignment between them, the two sequences are orthogonal for all transmissions within this time misalignment.

Property 3. The absolute value of the cyclic cross-correlation function between any two ZC sequences is constant and equal to $1/\sqrt{N_{\mathrm{ZC}}}$,[7] if $|q_1 - q_2|$ (where q_1 and q_2 are the sequence indices) is relatively prime with respect to N_{ZC} (a condition that can be easily guaranteed if N_{ZC} is a prime number). The cross-correlation of $\sqrt{N_{\mathrm{ZC}}}$ at all lags achieves the theoretical minimum cross-correlation value for any two sequences that have ideal autocorrelation.

Selecting N_{ZC} as a prime number results in $N_{\mathrm{ZC}} - 1$ ZC sequences which have the optimal cyclic cross-correlation between any pair. However, it is not always convenient to use sequences of prime length. In general, a sequence of non-prime length may be generated by either cyclic extension or truncation of a prime-length ZC sequence. A further useful property of ZC sequences is that the DFT of a ZC sequence $x_u(n)$ (in Equation (7.1)) is a weighted cyclicly shifted ZC sequence $X_{\mathrm{w}}(k)$ such that $w = -1/u \bmod N_{\mathrm{ZC}}$. This means that a ZC sequence can be generated directly in the frequency domain without the need for a DFT operation.

[7] Note that this value corresponds to the normalized cross-correlation function.

7.2.2 Primary Synchronization Signal (PSS) Sequences

The PSS is constructed from a frequency-domain ZC sequence of length 63, with the middle element punctured to avoid transmitting on the d.c. subcarrier.

The mapping of the PSS sequence to the subcarriers is shown in Figure 7.6.

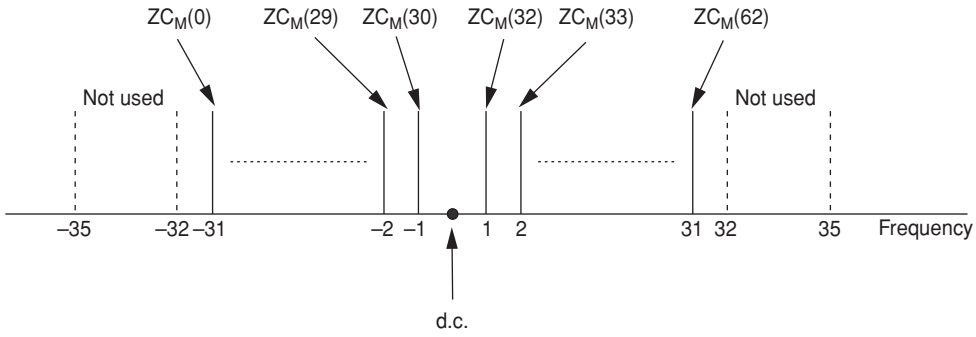


Figure 7.6: PSS sequence mapping in the frequency domain.

Three PSS sequences are used in LTE, corresponding to the three physical layer identities within each group of cells. The selected roots for the three ZC PSS sequences are $M = 29,\ 34,\ 25$, such that the frequency-domain length-63 sequence for root M is given by

$$\mathrm{ZC}_M^{63}(n) = \exp\left[-j\frac{\pi M n(n+1)}{63}\right], \quad n = 0,\ 1, \ldots, 62 \tag{7.3}$$

This set of roots for the ZC sequences was chosen for its good periodic autocorrelation and cross-correlation properties. In particular, these sequences have a low frequency-offset sensitivity, defined as the ratio of the maximum undesired autocorrelation peak in the time domain to the desired correlation peak computed at a certain frequency offset. This allows a certain robustness of the PSS detection during the initial synchronization, as shown in Figure 7.7.

Figures 7.8 and 7.9 show respectively the cross-correlation (for roots 29 and 25) as a function of timing and frequency offset and the autocorrelation as a function of timing offset (for root 29). It can be seen that the average and peak values of the cross-correlation are low relative to the autocorrelation. Furthermore, the ZC sequences are robust against frequency drifts as shown in Figure 7.10. Thanks to the flat frequency-domain autocorrelation property and to the low frequency offset sensitivity, the PSS can be easily detected during the initial synchronization with a frequency offset up to ±7.5 kHz.

From the UE's point of view, the selected root combination satisfies time-domain root-symmetry, in that sequences 29 and 34 are complex conjugates of each other and can be detected with a single correlator, thus allowing for some complexity reduction.

The UE must detect the PSS without any a priori knowledge of the channel, so non-coherent correlation[8] is required for PSS timing detection. A maximum likelihood detector,

[8] See Section 7.3.

as explained in Section 7.3, finds the timing offset m_M^* that corresponds to the maximum correlation, i.e.

$$m_M^* = \text{argmax}_m \left| \sum_{i=0}^{N-1} Y[i+m] S_M^*[i] \right|^2 \tag{7.4}$$

where i is time index, m is the timing offset, N is the PSS time-domain signal length, $Y[i]$ is the received signal at time instant i and $S_M[i]$ is the PSS with root M replica signal at time i as defined in Equation (7.3).

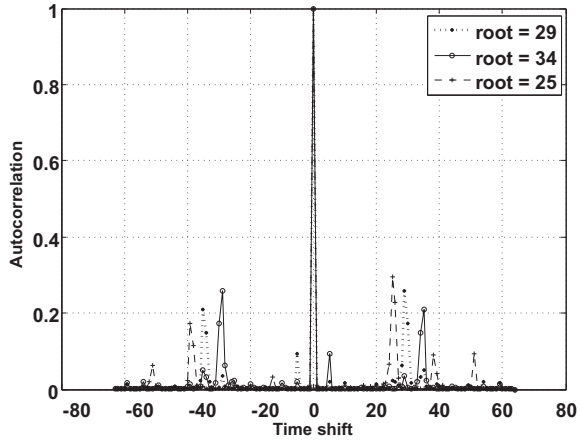


Figure 7.7: Autocorrelation profile at 7.5 kHz frequency offset for roots = 25, 29, 34.

7.2.3 Secondary Synchronization Signal (SSS) Sequences

The SSS sequences are based on maximum length sequences, known as M-sequences, which can be created by cycling through every possible state of a shift register of length n. This results in a sequence of length $2^n - 1$.

Each SSS sequence is constructed by interleaving, in the frequency domain, two length-31 BPSK[9]-modulated secondary synchronization codes, denoted here SSC1 and SSC2, as shown in Figure 7.11.

These two codes are two different cyclic shifts of a single length-31 M-sequence. The cyclic shift indices of the M-sequences are derived from a function of the PCI group (as given in Table 6.11.2.1-1 in [5]). The two codes are alternated between the first and second SSS transmissions in each radio frame. This enables the UE to determine the 10 ms radio frame timing from a single observation of an SSS, which is important for UEs handing over to LTE from another Radio Access Technology (RAT). For each transmission, SSC2 is scrambled by a sequence that depends on the index of SSC1. The sequence is then scrambled by a code that depends on the PSS. The scrambling code is one-to-one mapped to the physical layer

[9]Binary Phase Shift Keying.

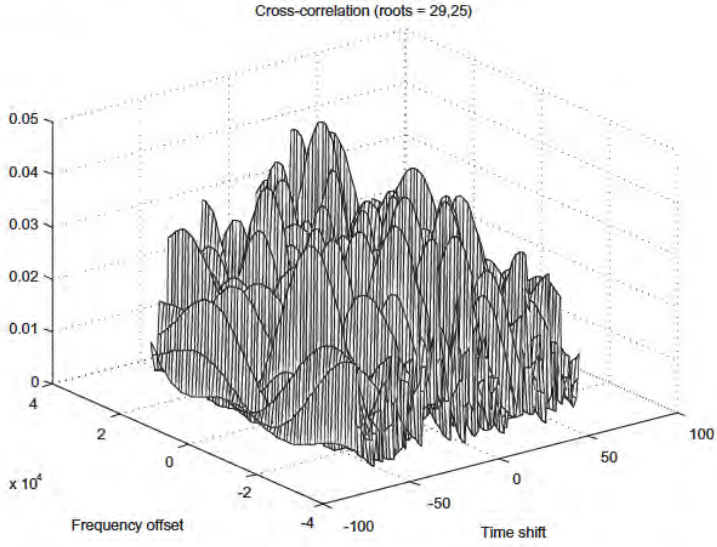


Figure 7.8: Cross-correlation of the PSS sequence pair 25 and 29.

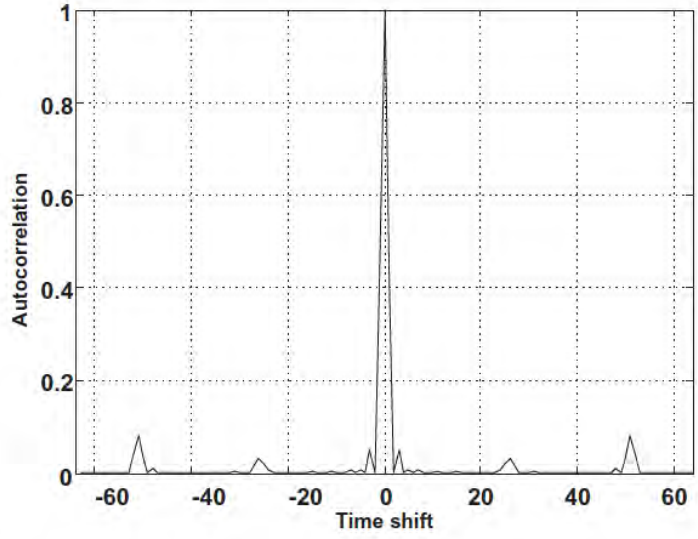


Figure 7.9: Autocorrelation of the PSS sequence 29 as a function of time offset.

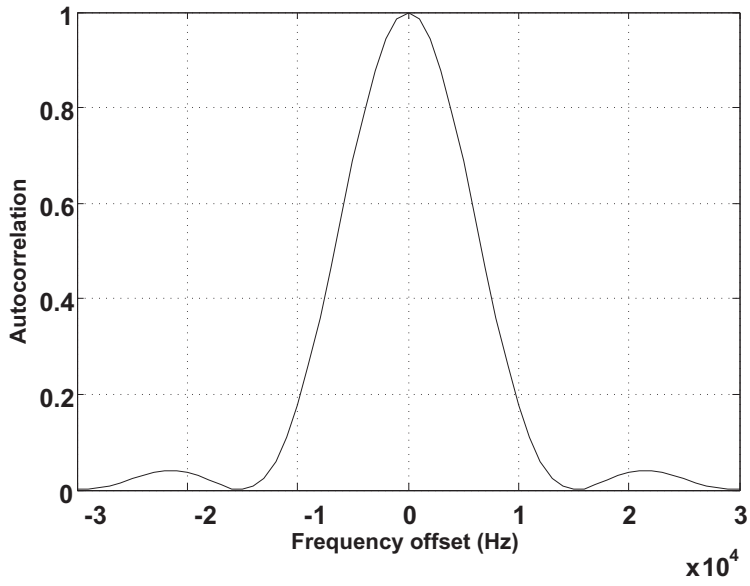


Figure 7.10: Autocorrelation of the PSS sequence as a function of frequency offset.

identity within the group corresponding to the target eNodeB. The sequence construction is illustrated in Figure 7.12; details of the scrambling operations are given in [5].

The SSS sequences have good frequency-domain properties, as shown in Figure 7.13. As in the case of the PSS, the SSS can be detected with a frequency offset up to ±7.5 kHz. In the time domain, the cross-correlation between any cyclic shifts of an SSS sequence is not as good as for classical M-sequences (for which the cross-correlation is known to be −1), owing to the effects of the scrambling operations.

From the UE's point of view, the SSS detection is done after the PSS detection, and the channel can therefore be assumed to be known (i.e. estimated based on the detected PSS sequence). It follows that a coherent detection method, as described in Section 7.3 (Equation (7.8)), can be applied:

$$\hat{\mathbf{S}}_m = \underset{\mathbf{S}}{\operatorname{argmin}}\left(\sum_{n=1}^{N} |\mathbf{y}[n] - \mathbf{S}[n, n]\hat{h_n}|^2\right) \tag{7.5}$$

where the symbols $\mathbf{S}[n, n]$ represent the SSS sequences and $\hat{h_n}$ are the estimated channel coefficients.

However, in the case of synchronized neighbouring eNodeBs, the performance of a coherent detector can be degraded. This is because if an interfering eNodeB employs the same PSS as the one used by the target cell, the phase difference between the two eNodeBs can have an impact on the quality of the estimation of the channel coefficients. On the other hand, the performance of a non-coherent detector degrades if the coherence bandwidth of the channel is less than the six resource blocks occupied by the SSS.

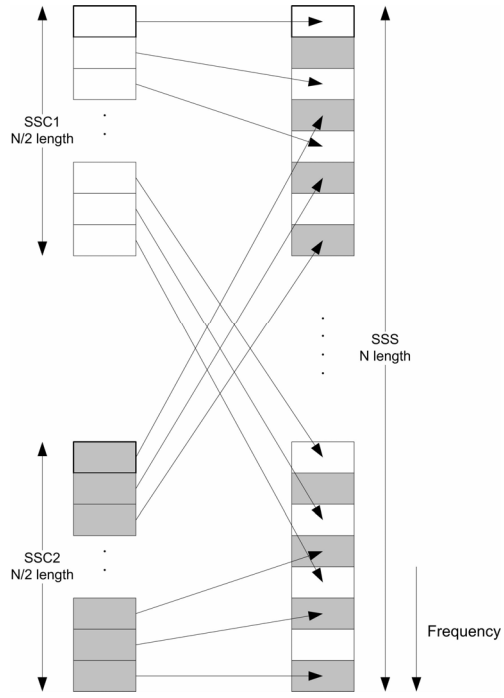


Figure 7.11: SSS sequence mapping.

In order to reduce the complexity of the SSS detector, the equivalence between M-sequence matrices and Walsh–Hadamard matrices can be exploited [6]. Using this property, a fast M-sequence transform is equivalent to a fast Walsh–Hadamard transform with index remapping. Thanks to this property the complexity of the SSS detector is reduced to $N \log_2 N$ where $N = 32$.

7.3 Coherent Versus Non-Coherent Detection

This section gives some theoretical background to the difference between coherent and non-coherent detection.

Both coherent and non-coherent detection may play a part in the synchronization procedures: in the case of the PSS, non-coherent detection is used, while for SSS sequence detection, coherent or non-coherent techniques can be used. From a conceptual point of view, a coherent detector takes advantage of knowledge of the channel, while a non-coherent detector uses an optimization metric corresponding to the average channel statistics.

We consider a generic system model where the received sequence $\mathbf{y}_m$ at time instant m is given by

$$\mathbf{y}_m = \mathbf{S}_m \mathbf{h} + \boldsymbol{\nu}_m \tag{7.6}$$

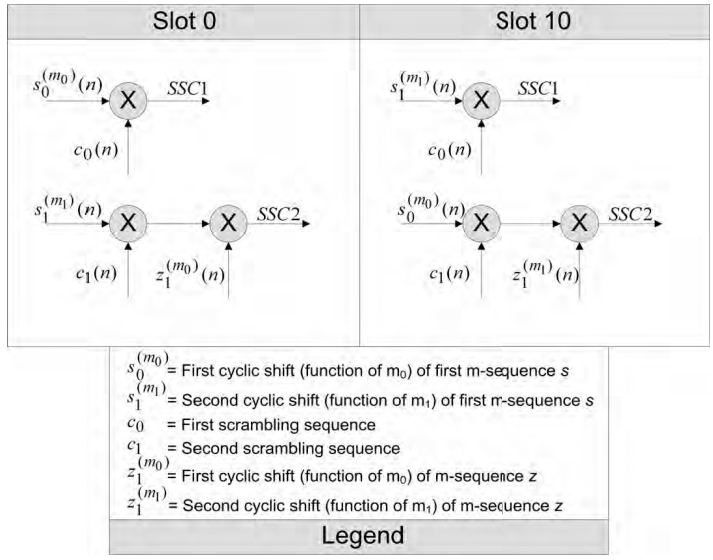


Figure 7.12: SSS sequence generation.

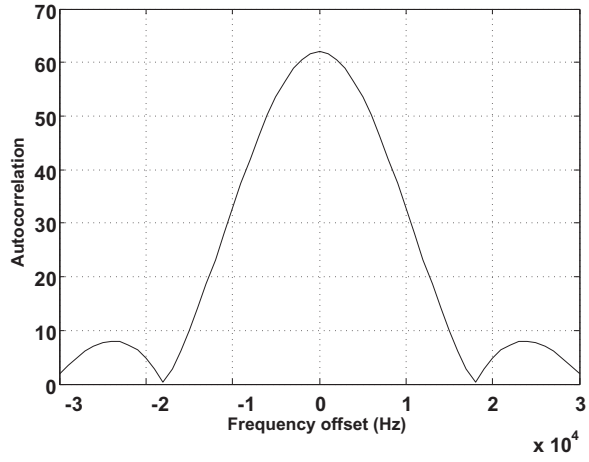


Figure 7.13: Autocorrelation of an SSS sequence as a function of frequency offset.

where the matrix $\mathbf{S}_m = \text{diag}\,[s_{m,1}, \ldots, s_{m,N}]$ represents the transmitted symbol at time instant m, $\mathbf{h} = [h_1, \ldots, h_N]^{\mathrm{T}}$ is the channel vector and $\boldsymbol{\nu}_m$ is zero-mean complex Gaussian noise with variance σ^2.

Maximum Likelihood (ML) coherent detection of a transmitted sequence consists of finding the sequence such that the probability that this sequence was transmitted, conditioned

on the knowledge of the channel, is maximized. Coherent detectors therefore require channel estimation to be performed first.

Since the noise is assumed to be independently identically distributed (i.i.d.) and detection is symbol-by-symbol, we focus on one particular symbol interval, neglecting the time instant without loss of generality. The problem becomes

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmax}} \ \Pr(\mathbf{y} \mid \mathbf{S}, \mathbf{h}) = \underset{\mathbf{S}}{\operatorname{argmax}} \ \frac{1}{(\pi N_0)^N} \exp\left[-\frac{\|\mathbf{y} - \mathbf{S}\mathbf{h}\|^2}{N_0}\right] \tag{7.7}$$

$$= \underset{\mathbf{S}}{\operatorname{argmin}} \left(\sum_{n=1}^{N} |\mathbf{y}_n - \mathbf{S}_{n,n} h_n|^2\right) \tag{7.8}$$

Equation (7.8) is a minimum squared Euclidean distance rule where the symbols $\mathbf{S}_{n,n}$ are weighted by the channel coefficient h_n.

When channel knowledge is not available or cannot be exploited, non-coherent detection can be used. The ML detection problem then maximizes the following conditional probability:

$$\begin{aligned}\hat{\mathbf{S}} &= \underset{\mathbf{S}}{\operatorname{argmax}} \ \Pr(\mathbf{y} \mid \mathbf{S}) \\ &= \underset{\mathbf{S}}{\operatorname{argmax}} \ \mathbb{E}_{\mathbf{h}}[\Pr(\mathbf{y} \mid \mathbf{S}, \mathbf{h})] = \int_{\mathbf{h}} \Pr(\mathbf{y} \mid \mathbf{S}, \mathbf{h}) \Pr(\mathbf{h}) \, \mathrm{d}\mathbf{h} \\ &= \int_{\mathbf{h}} \frac{1}{(\pi \, N_0)^N} \exp\left[-\frac{\|\mathbf{y}_m - \mathbf{S}_m \mathbf{h}\|^2}{N_0}\right] \Pr(\mathbf{h}) \, \mathrm{d}\mathbf{h}\end{aligned} \tag{7.9}$$

By considering the Probability Density Function (PDF) of an AWGN channel, it can be shown [9] that the ML non-coherent detector yields

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmax}} \ \{\mathbf{y}^{\mathrm{H}} \mathbf{S} (I + N_0 R_{\mathbf{h}}^{-1})^{-1} \mathbf{S}^{\mathrm{H}} \mathbf{y}\} \tag{7.10}$$

where the maximization is done over the input symbols $\mathbf{S}$, so all terms which do not depend on $\mathbf{S}$ can be discarded. Depending on the form of $R_{\mathbf{h}}$, the ML non-coherent detector can be implemented in different ways. For example, in the case of a frequency non-selective channel, the channel correlation matrix can be written as $R_{\mathbf{h}} = \sigma_{\mathbf{h}}^2 \mathbf{V}$ where $\mathbf{V}$ is the all-ones matrix.

Under this assumption, the non-coherent ML detector is thus obtained by the maximization of

$$\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmax}} \left\{ \left| \sum_{i=1}^{N} \mathbf{S}[i, i] \mathbf{y}[i] \right|^2 \right\} \tag{7.11}$$

References[10]

[1] J. D. C. Chu, 'Polyphase Codes with Good Periodic Correlation Properties'. *IEEE Trans. on Information Theory*, Vol. 18, pp. 531–532, July 1972.

[10] All web sites confirmed 1st March 2011.

[2] R. Frank, S. Zadoff and R. Heimiller, 'Phase Shift Pulse Codes With Good Periodic Correlation Properties'. *IEEE Trans. on Information Theory*, Vol. 8, pp. 381–382, October 1962.

[3] B. M. Popovic, 'Generalized Chirp-Like Polyphase Sequences with Optimum Correlation Properties'. *IEEE Trans. on Information Theory*, Vol. 38, pp. 1406–1409, July 1992.

[4] 3GPP Technical Specification 25.213, 'Spreading and modulation (FDD) ', www.3gpp.org.

[5] 3GPP Technical Specification 36.211, 'Physical Channels and Modulation ', www.3gpp.org.

[6] M. Cohn and A. Lempel, 'On Fast M-Sequence Transforms'. *IEEE Trans. on Information Theory*, Vol. 23, pp. 135–137, January 1977.

[7] Texas Instruments, NXP, Motorola, Ericsson, and Nokia, 'R4-072215: Simulation Assumptions for Intra-frequency Cell Identification', www.3gpp.org 3GPP TSG RAN WG4, meeting 45, Jeju, Korea, November 2007.

[8] NXP, 'R4-080691: LTE Cell Identification Performance in Multi-cell Environment', www.3gpp.org 3GPP TSG RAN WG4, meeting 46bis, Shenzen, China, February 2008.

[9] D. Reader, 'Blind Maximum Likelihood Sequence Detection over Fast Fading Communication Channels', Dissertation, University of South Australia, Australia, August 1996.

8

Reference Signals and Channel Estimation

Andrea Ancora, Stefania Sesia and Alex Gorokhov

8.1 Introduction

A simple communication system can be generally modelled as in Figure 8.1, where the transmitted signal **x** passes through a radio channel **H** and suffers additive noise before being received. Mobile radio channels usually exhibit multipath fading, which causes Inter-Symbol Interference (ISI) in the received signal. In order to remove ISI, various kinds of equalization and detection algorithms can be utilized, which may or may not exploit knowledge of the Channel Impulse Response (CIR). Orthogonal Frequency Division Multiple Access (OFDMA) is particularly robust against ISI, thanks to its structure and the use of the Cyclic Prefix (CP) which allows the receiver to perform a low-complexity single-tap scalar equalization in the frequency domain, as described in Section 5.2.1.

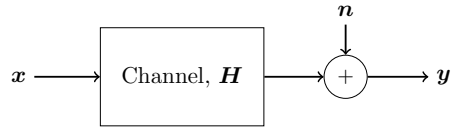


Figure 8.1: A simple transmission model.

As explained in Section 7.3, when the detection method exploits channel knowledge, it is generally said to be 'coherent'; otherwise it is called 'non-coherent'. Coherent detection can make use of both amplitude and phase information carried by the complex signals, and not of only amplitude information as with non-coherent detection. Optimal reception by coherent detection therefore typically requires accurate estimation of the propagation channel.

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

The main advantage of coherent detection is the simplicity of implementation compared to the more complex algorithms required by non-coherent detection for equivalent performance. However, this simplicity comes at a price, namely the overhead needed in order to be able to estimate the channel. A common and simple way to estimate the channel is to exploit known signals which do not carry any data, but which therefore cause a loss in spectral efficiency. In general, it is not an easy task to find the optimal trade-off between minimizing the spectral efficiency loss due to the overhead and providing adequate ability to track variations in the channel.

Other possible techniques for channel estimation include exploiting the correlation properties of the channel or using blind estimation.

Once synchronization between an eNodeB and a UE has been achieved, LTE (in common with earlier systems such as GSM and UMTS) is a coherent communication system, for which purpose known Reference Signals (RSs) are inserted into the transmitted signal structure.

In general, a variety of methods can be used to embed RSs into a transmitted signal. The RSs can be multiplexed with the data symbols (which are unknown at the receiver) in either the frequency, time or code domains (the latter being used in the case of the common pilot channel in the UMTS downlink). A special case of time multiplexing, known as preamble-based training, involves transmitting the RSs at the beginning of each data burst. Multiplexing-based techniques have the advantage of low receiver complexity, as the data symbol detection is decoupled from the channel estimation problem. Alternatively, RSs may be superimposed on top of the unknown data, without the two necessarily being orthogonal. Note that multiplexing RSs in the code domain is a particular type of superposition with a constraint on orthogonality between known RSs and the unknown data. A comprehensive analysis of the optimization of RS design can be found in [1, 2].

Orthogonal RS multiplexing is by far the most common technique. For example, to facilitate channel estimation in the UMTS downlink, two types of orthogonal RS are provided. The first is code-multiplexed, available to all users in a cell, and uses a specific spreading code which is orthogonal to the codes used to spread the users' data. The second type is time-multiplexed dedicated RSs, which may in some situations be inserted into the users' data streams [3].

In the LTE downlink, the OFDM transmission can be described by a two-dimensional lattice in time and frequency, as shown in Figure 6.1 and described in Chapter 6. This structure facilitates the multiplexing of the RSs, which are mapped to specific Resource Elements (REs) of the two-dimensional lattice in Figure 6.1 according to patterns explained in Section 8.2.

In order to estimate the channel as accurately as possible, all correlations between channel coefficients in time, frequency and space should be taken into account. Since RSs are sent only on particular OFDM REs (i.e. on particular OFDM symbols on particular subcarriers), channel estimates for the REs which do not bear RSs have to be computed via interpolation. The optimal interpolating channel estimator in terms of mean-squared error is based on a two-dimensional Wiener filter interpolation [4]. Due to the high complexity of such a filter, a trade-off between complexity and accuracy is achieved by using one-dimensional filters. In Sections 8.4, 8.5 and 8.6, the problem of channel estimation is approached from a theoretical point of view, and some possible solutions are described.

The work done in the field of channel estimation, and the corresponding literature available, is vast. Nevertheless many challenges still remain, and we refer the interested reader to [2, 5] for general surveys of open issues in this area.

8.2 Design of Reference Signals in the LTE Downlink

In the LTE downlink, five different types of RS are provided [6, Section 6.10]:

- Cell-specific RSs (often referred to as 'common' RSs, as they are available to all UEs in a cell and no UE-specific processing is applied to them);
- UE-specific RSs (introduced in Release 8, and extended in Releases 9 and 10), which may be embedded in the data for specific UEs (also known as DeModulation Reference Signals (DM-RSs);
- MBSFN-specific RSs, which are used only for Multimedia Broadcast Single Frequency Network (MBSFN) operation and are discussed further in Section 13.4.1;
- Positioning RSs, which from Release 9 onwards may be embedded in certain 'positioning subframes' for the purpose of UE location measurements; these are discussed in Section 19.3.1;
- Channel State Information (CSI) RSs, which are introduced in Release 10 specifically for the purpose of estimating the downlink channel state and not for data demodulation (see Section 29.1.2).

Each RS pattern is transmitted from an *antenna port* at the eNodeB. An antenna port may in practice be implemented either as a single physical transmit antenna, or as a combination of multiple physical antenna elements. In either case, the signal transmitted from each antenna port is not designed to be further deconstructed by the UE receiver: the transmitted RS corresponding to a given antenna port defines the antenna port from the point of view of the UE, and enables the UE to derive a channel estimate for all data transmitted on that antenna port – regardless of whether it represents a single radio channel from one physical antenna or a composite channel from a multiplicity of physical antenna elements together comprising the antenna port. The designations of the antenna ports available in LTE are summarized below:

- Antenna Ports 0–3: cell-specific RSs – see Section 8.2.1;
- Antenna Port 4: MBSFN – see Section 13.4.1;
- Antenna Port 5: UE-specific RSs for single-layer beamforming – see Section 8.2.2;
- Antenna Port 6: positioning RSs (introduced in Release 9) – see Section 19.3.1;
- Antenna Ports 7–8: UE-specific RSs for dual-layer beamforming (introduced in Release 9) – see Section 8.2.3;
- Antenna Ports 9–14: UE-specific RSs for multi-layer beamforming (introduced in Release 10) – see Section 29.1.1;
- Antenna Ports 15–22: CSI RSs (introduced in Release 10) – see Section 29.1.2.

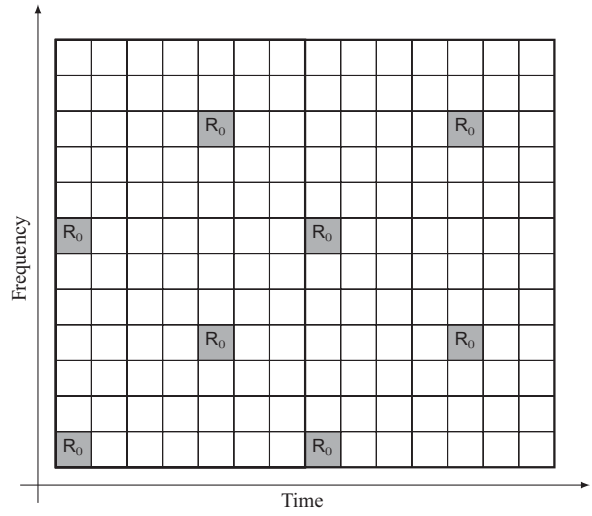


Figure 8.2: Cell-specific reference symbol arrangement in the case of normal CP length for one antenna port. Reproduced by permission of © 3GPP.

Details of the techniques provided for multi-antenna transmission in LTE can be found in Chapter 11 for Releases 8 and 9, and in Chapter 29 for the new techniques introduced in Release 10 for LTE-Advanced.

8.2.1 Cell-Specific Reference Signals

The cell-specific RSs enable the UE to determine the phase reference for demodulating the downlink control channels (see Section 9.3) and the downlink data in most transmission modes[1] of the Physical Downlink Shared Channel (PDSCH – see Section 9.2.2). If UE-specific precoding is applied to the PDSCH data symbols before transmission,[2] downlink control signalling is provided to inform the UE of the corresponding phase adjustment it should apply relative to the phase reference provided by the cell-specific RSs. The cell-specific RSs are also used by the UEs to generate Channel State Information (CSI) feedback (see Sections 10.2.1 and 11.2.2.4).

References [7, 8] show that in an OFDM-based system an equidistant arrangement of reference symbols in the lattice structure achieves the Minimum Mean-Squared Error (MMSE) estimate of the channel. Moreover, in the case of a uniform reference symbol grid, a 'diamond shape' in the time-frequency plane can be shown to be optimal.

In LTE, the arrangement of the REs on which the cell-specific RSs are transmitted follows these principles. Figure 8.2 illustrates the RS arrangement for the normal CP length.[3]

[1] Transmission modes 1 to 6.

[2] In PDSCH transmission modes 3 to 6.

[3] In the case of the extended CP, the arrangement of the reference symbols changes slightly, but the explanations in the rest of this chapter are no less valid. The detailed arrangement of reference symbols for the extended CP can be found in [6].

The LTE system is designed to work under high-mobility assumptions, in contrast to WLAN systems which are generally optimized for pedestrian-level mobility. WLAN systems typically use a preamble-based training sequence, and the degree of mobility such systems can support depends on how often the preamble is transmitted.

The required spacing in time between the reference symbols can be determined by considering the maximum Doppler spread (highest speed) to be supported, which for LTE corresponds to 500 km/h [9]. The Doppler shift is $f_d = (f_c\, v/c)$ where f_c is the carrier frequency, v is the UE speed in metres per second, and c is the speed of light ($3 \cdot 10^8$ m/s). Considering $f_c = 2$ GHz and $v = 500$ km/h, then the Doppler shift is $f_d \simeq 950$ Hz. According to Nyquist's sampling theorem, the minimum sampling frequency needed in order to reconstruct the channel is therefore given by $T_c = 1/(2f_d) \simeq 0.5$ ms under the above assumptions. This implies that two reference symbols per slot are needed in the time domain in order to estimate the channel correctly.

In the frequency direction, there is one reference symbol every six subcarriers on each OFDM symbol that includes reference symbols, but the reference symbols are staggered so that within each Resource Block (RB) there is one reference symbol every threesubcarriers, as shown in Figure 8.2. This spacing is related to the expected coherence bandwidth of the channel, which is in turn related to the channel delay spread. In particular the 90% and 50% coherence bandwidths[4] are given by $B_{c,90\%} = 1/50\sigma_\tau$ and $B_{c,50\%} = 1/5\sigma_\tau$ respectively, where σ_τ is the r.m.s delay spread. In [10], the maximum r.m.s channel delay spread considered is 991 ns, corresponding to $B_{c,90\%} = 20$ kHz and $B_{c,50\%} = 200$ kHz. In LTE, the spacing between two reference symbols in frequency, in one RB, is 45 kHz, thus allowing the expected frequency-domain variations of the channel to be resolved.

Up to four cell-specific antenna ports, numbered 0 to 3, may be used by an LTE eNodeB, thus requiring the UE to derive up to four separate channel estimates.[5] For each antenna port, a different RS pattern has been designed, with particular attention having been given to the minimization of the intra-cell interference between the multiple transmit antenna ports. In Figure 8.3, R_p indicates that the RE is used for the transmission of an RS on antenna port p. When an RE is used to transmit an RS on one antenna port, the corresponding RE on the other antenna ports is set to zero to limit the interference.

From Figure 8.3, it can be noticed that the density of reference symbols for the third and fourth antenna ports is half that of the first two; this is to reduce the overhead in the system. Frequent reference symbols are useful for high-speed conditions as explained above. In cells with a high prevalence of high-speed users, the use of four antenna ports is unlikely, hence for these conditions reference symbols with lower density can provide sufficient channel estimation accuracy.

All the RSs (cell-specific, UE-specific or MBSFN-specific) are QPSK modulated – a constant modulus modulation. This property ensures that the Peak-to-Average Power Ratio (PAPR) of the transmitted waveform is kept low. The signal can be written as

$$r_{l,n_s}(m) = \frac{1}{\sqrt{2}}[1 - 2c(2m)] + j\frac{1}{\sqrt{2}}[1 - 2c(2m+1)] \tag{8.1}$$

[4] $B_{c,x\%}$ is the bandwidth where the autocorrelation of the channel in the frequency domain is equal to $x\%$ of the peak.

[5] Any MBSFN and UE-specific RSs, if transmitted, constitute additional independent antenna ports in the LTE specifications.

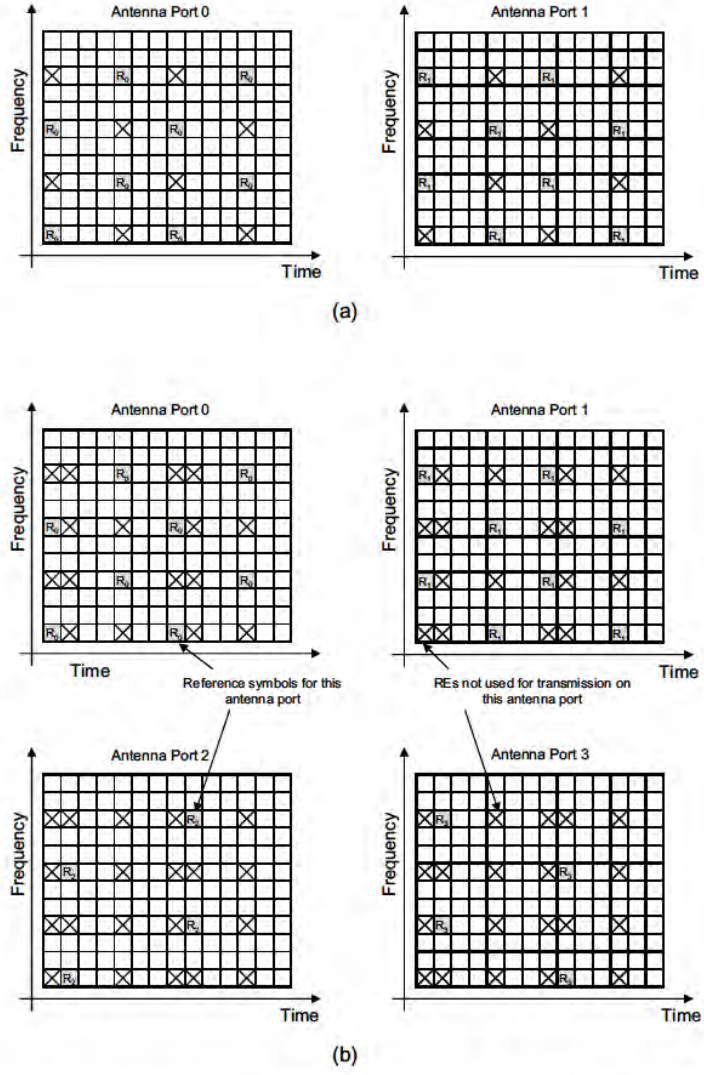


Figure 8.3: Cell-specific RS arrangement in the case of normal CP length for (a) two antenna ports, (b) four antenna ports. Reproduced by permission of © 3GPP.

where m is the index of the RS, n_s is the slot number within the radio frame and 'l' is the symbol number within the time slot. The pseudo-random sequence $c(i)$ is comprised of a length-31 Gold sequence, already introduced in Chapter 6, with different initialization values depending on the type of RSs. For the cell-specific RSs, the sequence is reinitialized at the

start of each OFDM symbol, with a value that depends on the cell identity, $N_{\mathrm{ID}}^{\mathrm{cell}}$. The cell-specific RS sequence therefore carries unambiguously one of the 504 different cell identities.

A cell-specific frequency shift is applied to the patterns of reference symbols shown in Figures 8.2 and 8.3, given by $N_{\mathrm{ID}}^{\mathrm{cell}} \mathrm{mod} 6$.[6] This shift helps to avoid time-frequency collisions between cell-specific RSs from up to six adjacent cells. Avoidance of collisions is particularly relevant in cases when the transmission power of the RS is boosted, as is possible in LTE up to a maximum of 6 dB relative to the surrounding data symbols. RS power-boosting is designed to improve channel estimation in the cell, but if adjacent cells transmit high-power RSs on the same REs, the resulting inter-cell interference will prevent the benefit from being realized.

8.2.2 UE-Specific Reference Signals in Release 8

In Release 8 of LTE, UE-specific RSs may be transmitted in addition to the cell-specific RSs described above if the UE is configured (by higher-layer RRC signalling) to receive its downlink PDSCH data in transmission mode 7 (see Section 9.2.2.1). The UE-specific RSs are embedded only in the RBs to which the PDSCH is mapped for those UEs. If UE-specific RSs are transmitted, the UE is expected to use them to derive the channel estimate for demodulating the data in the corresponding PDSCH RBs. The same precoding is applied to the UE-specific RSs as to the PDSCH data symbols, and therefore there is no need for signalling to inform the UE of the precoding applied. Thus the UE-specific RSs are treated as being transmitted using a distinct antenna port (number 5), with its own channel response from the eNodeB to the UE.

A typical usage of the UE-specific RSs is to enable beamforming of the data transmissions to specific UEs. For example, rather than using the physical antennas used for transmission of the other (cell-specific) antenna ports, the eNodeB may use a correlated array of physical antenna elements to generate a narrow beam in the direction of a particular UE. Such a beam will experience a different channel response between the eNodeB and UE, thus requiring the use of UE-specific RSs to enable the UE to demodulate the beamformed data coherently. The use of UE-specific beamforming is discussed in more detail in Section 11.2.2.3.

As identified in [11], the structure shown in Figure 8.4 (for the normal CP) has been chosen because there is no collision with the cell specific RSs, and hence the presence of UE-specific RSs does not affect features related to the cell-specific RSs. The UE-specific RSs have a similar pattern to that of the cell-specific RSs, which allows a UE to re-use similar channel estimation algorithms. The density is half that of the cell-specific RS, hence minimizing the overhead. Unlike the cell-specific RSs, the sequence for the UE-specific RSs is only reinitialized at the start of each subframe, as the number of REs to which the sequence is mapped in one OFDM symbol may be very small (in the event of a small number of RBs being transmitted to a UE). The initialization depends on the UE's identity.

The corresponding pattern for use in case of the extended CP being configured in a cell can be found in [6, Section 6.10.3.2].

8.2.3 UE-Specific Reference Signals in Release 9

A new design for UE-specific RSs is defined in Release 9 of the LTE specifications in order to extend UE-specific RS support to dual layer transmission. This includes transmission of

[6] The mod6 operation is used because reference symbols are spaced apart by six subcarriers in the lattice grid.

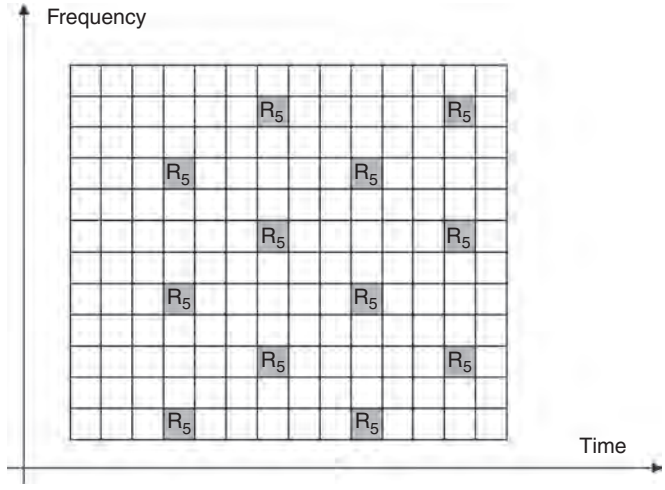


Figure 8.4: Release 8 UE-specific RS arrangement for PDSCH transmission mode 7 with normal CP. Reproduced by permission of © 3GPP.

two spatial layers to one UE, or single-layer transmission to each of two UEs as a Multi-User Multiple Input Multiple Output (MU-MIMO) transmission (see Chapter 11).

Dual layer beamforming in Release 9 is a precursor to higher-rank single- and multi-user transmissions in Release 10, and therefore scalability towards LTE-Advanced (see Section 29.1) was a key factor in the choice of design for the new RSs. This includes enabling efficient inter-cell coordination, which can be facilitated by choosing the set of REs to be non-cell-specific (i.e. without a frequency shift that depends on the cell identity).

A further consideration was backward compatibility with Release 8; this requires coexistence with the 'legacy' physical channels, and therefore the REs carrying the new UE-specific RSs have to avoid the cell-specific RSs and downlink control channels.

A final design requirement was to keep equal power spectral density of the downlink transmission in one RB. This favours designs where the UE-specific RSs for both layers are present in the same OFDM symbol.

Since UE-specific RSs are designed for time-frequency channel estimation within a given RB, the eigen-structure of the time and frequency channel covariance matrix provides insights into the optimal pattern of REs for the RSs for MMSE channel estimation. As an example, Figure 8.5 shows the eigenvalues and three dominant eigenvectors of the time-domain channel covariance matrix for a pair of RBs in a subframe for various mobile speeds assuming spatially uniform (Jakes) scattering. The same behaviour is observed in the frequency domain.

It can be seen that the three principal eigenvectors can be accurately approximated by constant, linear and quadratic functions respectively. Such a channel eigen-structure stems from the time-frequency variations across a pair of RBs being limited. Most of the energy is captured by the constant and linear components, and hence these should be the focus for optimization. While the reference symbol locations do not affect estimation accuracy for the

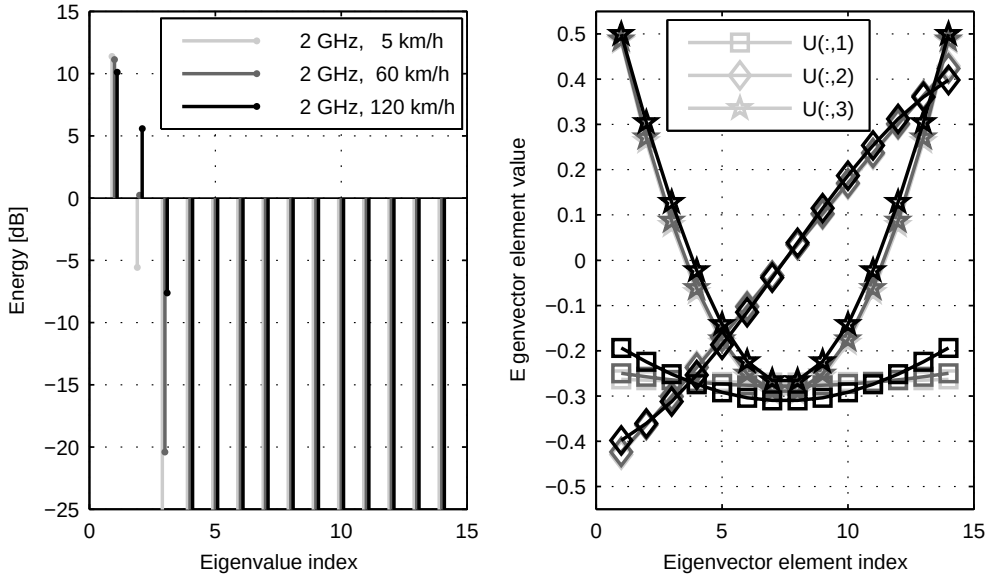


Figure 8.5: Time-domain channel covariance over a pair of RBs:
eigenvalues (left) and eigenvectors (right), assuming spatially uniform (Jakes) scattering.
The eigenvalue energies are normalized relative to the per-RE energy.

constant component, the best strategy for estimation of the linear component is to allocate RS energy at the edges of the measured region. Finally, the quadratic component is small but non-negligible, warranting three disjoint RS positions across the bandwidth of an RB.

Based on these observations, Figure 8.6 shows the pattern adopted for the new UE-specific RS pattern in Release 9.[7] The reference symbols are positioned in the earliest and latest available pairs of OFDM symbols that avoid collisions with the cell-specific RSs. Pairs of REs are used so that the UE-specific RSs for the two layers can be code-multiplexed. The UE-specific RSs for the two layers using this pattern are termed antenna ports 7 and 8. A UE configured to use the Release 9 dual-layer UE-specific RSs is configured in PDSCH transmission mode 8 (see Section 9.2.2.1).

Length-2 orthogonal Walsh codes are used for the code-multiplexing of the two RS ports.[8] Compared to the use of frequency-multiplexed RS ports, code multiplexing can improve the accuracy of interference estimation and potentially simplifies the implementation by maintaining the same set of RS REs regardless of the number of layers transmitted (thus facilitating dynamic switching between one and two layers).

Finally, the Release 9 UE-specific RS sequence is initialized using only the cell identity (without the UE identity used for the Release 8 UE-specific RSs), in order to enable the two orthogonally code-multiplexed UE-specific RS ports to be used for MU-MIMO (with the two RS ports assigned to different UEs). In addition, two different RS sequence initializations are

[7] Note that distributed RB mapping (see Section 9.2.2.1) is not supported in conjunction with UE-specific RSs.

[8] The mapping of the second Walsh code to each pair of REs reverses from one subcarrier to the next; this is to minimize the Peak-to-Average Power Ratio (PAPR) of the RS transmissions [12].

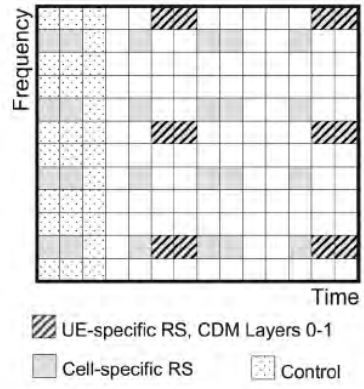


Figure 8.6: Dual-layer UE-specific RS arrangement in Release 9.

provided, enabling non-orthogonal UE-specific RS multiplexing to be used, for example for MU-MIMO with up to dual layer transmission to each UE. An example of a possible usage is shown in Figure 8.7.

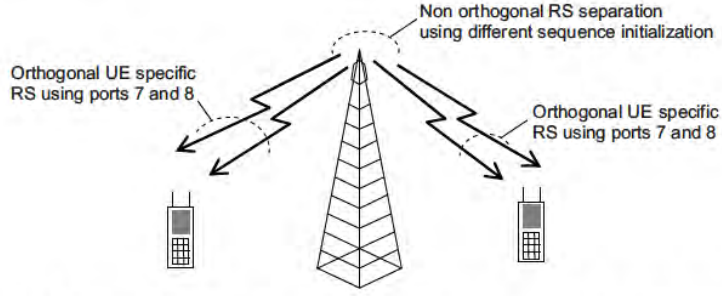


Figure 8.7: Example of usage of orthogonal and non-orthogonal UE-specific RSs in Release 9, for dual-layer transmission to each of two UEs with MU-MIMO.

8.3 RS-Aided Channel Modelling and Estimation

The channel estimation problem is related to the channel model, itself determined by the physical propagation characteristics, including the number of transmit and receive antennas, transmission bandwidth, carrier frequency, cell configuration and relative speed between eNodeB and UE receivers. In general,

- The carrier frequencies and system bandwidth mainly determine the scattering nature of the channel.

- The cell deployment governs its multipath, delay spread and spatial correlation characteristics.
- The relative speed sets the time-varying properties of the channel.

The propagation conditions characterize the channel correlation function in a three-dimensional space comprising frequency, time and spatial domains. In the general case, each MIMO multipath channel component can experience different but related spatial scattering conditions leading to a full three-dimensional correlation function across the three domains. Nevertheless, for the sake of simplicity, assuming that the multipath components of each spatial channel experience the same scattering conditions, the spatial correlation can be assumed to be independent from the other two domains and can be handled separately.

This framework might be suboptimal in general, but is nevertheless useful in mitigating the complexity of channel estimation as it reduces the general three-dimensional joint estimation problem into independent estimation problems.

For a comprehensive survey of MIMO channel estimation, the interested reader is referred to [13]. The following two subsections define the channel model and the corresponding correlation properties which are then used as the basis for an overview of channel estimation techniques.

8.3.1 Time-Frequency-Domain Correlation: The WSSUS Channel Model

The Wide-Sense Stationary Uncorrelated Scattering (WSSUS) channel model is commonly employed for the multipath channels experienced in mobile communications.

Neglecting the spatial dimension for the sake of simplicity, let $h(\tau; t)$ denote the time-varying complex baseband impulse response of a multipath channel realization at time instant t and delay τ.

Considering the channel as a random process in the time direction t, the channel is said to be delay Uncorrelated-Scattered (US) if

$$\mathbb{E}[h(\tau_a; t_1)^* h(\tau_b; t_2)] = \phi_h(\tau_a; t_1, t_2)\delta(\tau_b - \tau_a) \tag{8.2}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. According to the US assumption, two CIR components a and b at relative delays τ_a and τ_b are uncorrelated if $\tau_a \neq \tau_b$.

The channel is Wide-Sense Stationary (WSS) uncorrelated if

$$\phi_h(\tau; t_1, t_2) = \phi_h(\tau; t_2 - t_1) \tag{8.3}$$

which means that the correlation of each delay component of the CIR is only a function of the *difference* in time between each realization.

Hence, the second-order statistics of this model are completely described by its delay cross-power density $\phi_h(\tau; \Delta t)$ or by its Fourier transform, the scattering function defined as

$$S_h(\tau; f) = \int \phi_h(\tau; \Delta t) e^{-j2\pi f \Delta t} \, \mathrm{d}\Delta t \tag{8.4}$$

with f being the Doppler frequency. Other related functions of interest include the *Power Delay Profile* (PDP)

$$\psi_h(\tau) = \phi_h(\tau; 0) = \int S_h(\tau; f) \, \mathrm{d}f$$

the *time-correlation function*

$$\overline{\phi}_h(\Delta t) = \int \phi_h(\tau; \Delta t)\, \mathrm{d}\tau$$

and the *Doppler power spectrum*

$$\overline{S}_h(f) = \int S_h(\tau; f)\, \mathrm{d}\tau.$$

A more general exposition of WSSUS models is given in [14]. Classical results were derived by Clarke [15] and Jakes [16] for the case of a mobile terminal communicating with a stationary base station in a two-dimensional propagation geometry.

These well-known results state that

$$\overline{S}_h(f) = \frac{1}{\sqrt{f_\mathrm{d}^2 - f^2}} \tag{8.5}$$

for $|f| \leq f_\mathrm{d}$ with f_d being the maximum Doppler shift and

$$\bar{\phi}_h(\Delta t) = J_0(2\pi f_\mathrm{d} \Delta t) \tag{8.6}$$

where $J_0(\cdot)$ is the zero$^{\text{th}}$-order Bessel function. Figure 8.8 shows the PSD of the classical Doppler spectrum described by Clarke and Jakes [15, 16]. The Clarke and Jakes derivations are based on the assumption that the physical scattering environment is chaotic and therefore the angle of arrival of the electromagnetic wave at the receiver is a uniformly distributed random variable in the angular domain. As a consequence, the time-correlation function is strictly real-valued, the Doppler spectrum is symmetric and interestingly there is a *delay-temporal separability* property in the general bi-dimensional scattering function $S_h(\tau, \Delta t)$. In other words,

$$S_h(\tau; f) = \psi_h(\tau)\overline{S}_h(f) \tag{8.7}$$

or equivalently

$$\phi_h(\tau; \Delta t) = \psi_h(\tau)\overline{\phi}_h(\Delta t) \tag{8.8}$$

If the time and frequency aspects can be assumed to be separable, the complexity of channel estimation can be significantly reduced, as the problem is reduced to two one-dimensional operations.

The continuous-time correlation properties of the channel $h(\tau, t)$ discussed above apply equivalently to the corresponding low-pass sampled discrete-time CIR, i.e. $h[l, k]$ [17], for every l^{th} delay sampled at the kT^{th} instant where T is the sampling period. Moreover, these properties are maintained if we assume $h[l, k]$ to be well-approximated by a Finite-Impulse Response (FIR) vector $\mathbf{h}[k] = [h[0, k], \cdots, h[L-1, k]]^{\mathrm{T}}$ with a maximum delay spread of L samples. For the sake of notational simplicity and without loss of generality, the index k will be dropped in the following sections.

8.3.2 Spatial-Domain Correlation: The Kronecker Model

While the frequency and time correlations addressed in the previous section are induced by the channel delays and Doppler spread, the spatial correlation arises from the spatial scattering conditions.

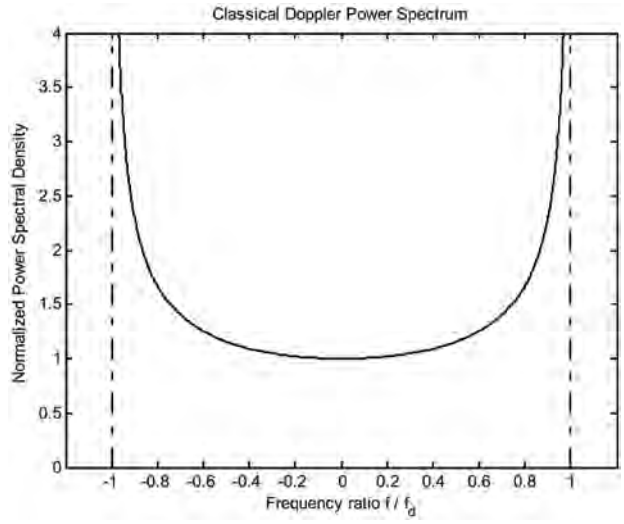


Figure 8.8: Normalized PSD for Clarke's model.

Among the possible spatial correlation models, for performance evaluation of LTE the Kronecker model is generally used [10]. Despite its simplicity, this correlation-based analytical model is widely used for the theoretical analysis of MIMO systems and yields experimentally verifiable results when limited to two or three antennas at each end of the radio link.

Let us assume the narrowband MIMO channel $N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}$ matrix for a system with N_{Tx} transmitting antennas and N_{Rx} receiving antennas to be

$$\mathbf{H} = \begin{bmatrix} h_{0,0} & \cdots & h_{0,N_{\mathrm{Tx}}-1} \\ \vdots & & \vdots \\ h_{N_{\mathrm{Rx}}-1,N_{\mathrm{Tx}}-1} & \cdots & h_{N_{\mathrm{Rx}}-1,N_{\mathrm{Tx}}-1} \end{bmatrix} \tag{8.9}$$

The narrowband assumption particularly suits OFDM systems (see Section 5.2.1), where each MIMO channel component $h_{n,m}$ can be seen as the complex channel coefficient of each spatial link at a given subcarrier index.

The matrix $\mathbf{H}$ is rearranged into a vector by means of the operator $\mathrm{vec}(\mathbf{H}) = [\mathbf{h}_0^{\mathrm{T}}, \mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_{N_{\mathrm{Tx}}-1}^{\mathrm{T}}]^{\mathrm{T}}$ where $\mathbf{h}_i$ is the i^{th} column of $\mathbf{H}$ and $\{\cdot\}^{\mathrm{T}}$ is the transpose operation. Hence, the correlation matrix can be defined as

$$\mathbf{C}_{\mathrm{S}} = \mathbb{E}[\mathrm{vec}(\mathbf{H})\mathrm{vec}(\mathbf{H})^{\mathrm{H}}] \tag{8.10}$$

where $\{\cdot\}^{\mathrm{H}}$ is the Hermitian operation. The matrix in (8.10) is the full correlation matrix of the MIMO channel.

The Kronecker model assumes that the full correlation matrix results from separate spatial correlations at the transmitter and receiver, which is equivalent to writing the full correlation matrix as a Kronecker product ($\otimes$) of the transmitter and receiver correlation matrices:

$$\mathbf{C}_{\mathrm{S}} = \mathbf{C}_{\mathrm{Tx}} \otimes \mathbf{C}_{\mathrm{Rx}} \tag{8.11}$$

with $\mathbf{C}_{\mathrm{Tx}} = \mathbb{E}[\mathbf{H}^{\mathrm{H}}\mathbf{H}]$ and $\mathbf{C}_{\mathrm{Rx}} = \mathbb{E}[\mathbf{H}\mathbf{H}^{\mathrm{H}}]$.

Typical values assumed for these correlations according to the Kronecker model are discussed in Section 20.3.5.

8.4 Frequency-Domain Channel Estimation

In this section, we address the channel estimation problem over one OFDMA symbol (specifically a symbol containing reference symbols) to exploit the frequency-domain characteristics.

In the LTE context, as for any OFDM system with uniformly distributed reference symbols [18], the Channel Transfer Function (CTF) can be estimated using a maximum likelihood approach in the frequency domain at the REs containing the RSs by de-correlating the constant modulus RS. Using a matrix notation, the CTF estimate $\widehat{\mathbf{z}}_p$ on reference symbol p can be written as

$$\widehat{\mathbf{z}}_p = \mathbf{z}_p + \widetilde{\mathbf{z}}_p = \mathbf{F}_p\mathbf{h} + \widetilde{\mathbf{z}}_p \tag{8.12}$$

for $p \in (0, \ldots, P)$ where P is the number of available reference symbols and $\mathbf{h}$ is the $L \times 1$ CIR vector. $\mathbf{F}_p$ is the $P \times L$ matrix obtained by selecting the rows corresponding to the reference symbol positions and the first L columns of the $N \times N$ Discrete Fourier Transform (DFT) matrix where N is the FFT order. $\widetilde{\mathbf{z}}_p$ is a $P \times 1$ zero-mean complex circular white noise vector whose $P \times P$ covariance matrix is given by $\mathbf{C}_{\widetilde{\mathbf{z}}_p}$. The effective channel length $L \leq L_{\mathrm{CP}}$ is assumed to be known.

8.4.1 Channel Estimate Interpolation

8.4.1.1 Linear Interpolation Estimator

The natural approach to estimate the whole CTF is to interpolate its estimate between the reference symbol positions. In the general case, let $\mathbf{A}$ be a generic interpolation filter; then the interpolated CTF estimate at subcarrier index i can be written as

$$\widehat{\mathbf{z}}_i = \mathbf{A}\widehat{\mathbf{z}}_p \tag{8.13}$$

Substituting Equation (8.12) into (8.13), the error of the interpolated CTF estimate is

$$\widetilde{\mathbf{z}}_i = \mathbf{z} - \widehat{\mathbf{z}}_i = (\mathbf{F}_L - \mathbf{A}\mathbf{F}_p)\mathbf{h} - \mathbf{A}\widetilde{\mathbf{z}}_p \tag{8.14}$$

where $\mathbf{F}_L$ is the $N \times L$ matrix obtained by taking the first L columns of the DFT matrix and $\mathbf{z} = \mathbf{F}_L\mathbf{h}$. In Equation (8.14), it can be seen that the channel estimation error is constituted of a bias term (itself dependent on the channel) and an error term.

The error covariance matrix is

$$\mathbf{C}_{\widetilde{\mathbf{z}}_i} = (\mathbf{F}_L - \mathbf{A}\mathbf{F}_p)\mathbf{C}_{\mathbf{h}}(\mathbf{F}_L - \mathbf{A}\mathbf{F}_p)^{\mathrm{H}} + \sigma^2_{\widetilde{\mathbf{z}}_p}\mathbf{A}\mathbf{A}^{\mathrm{H}} \tag{8.15}$$

where $\mathbf{C}_h = \mathbb{E}[\mathbf{h}\mathbf{h}^{\mathrm{H}}]$ is the channel covariance matrix.

Recalling Equation (8.13), linear interpolation would be the intuitive choice. Such an estimator is deterministically biased, but unbiased from the Bayesian viewpoint regardless of the structure of $\mathbf{A}$.

8.4.1.2 IFFT Estimator

As a second straightforward approach, the CTF estimate over all subcarriers can be obtained by IFFT interpolation. In this case, the matrix $\mathbf{A}$ from Equation (8.13) becomes:

$$\mathbf{A}_{\text{IFFT}} = \frac{1}{P}\mathbf{F}_L\mathbf{F}_p^{\text{H}} \tag{8.16}$$

The error of the IFFT-interpolated CTF estimate and its covariance matrix can be obtained by substituting Equation (8.16) into Equations (8.14) and (8.15).

With the approximation of $\mathbf{I}_L \approx (1/P)\mathbf{F}_p^{\text{H}}\mathbf{F}_p$, where $\mathbf{I}_L$ is the $L \times L$ identity matrix, it can immediately be seen that the bias term in Equation (8.14) would disappear, providing for better performance.

Given the LTE system parameters and the patterns of REs used for RSs, in practice $(1/P)\mathbf{F}_p^{\text{H}}\mathbf{F}_p$ is far from being a multiple of an identity matrix. The approximation becomes an equality when $K = N$, $N/W > L$ and N/W is an integer,[9] i.e. the system would have to be dimensioned without guard-bands and the RS would have to be positioned with a spacing which is an exact factor of the FFT order N, namely a power of two.

In view of the other factors affecting the design of the RS RE patterns outlined above, such constraints are impractical.

8.4.2 General Approach to Linear Channel Estimation

Compared to the simplistic approaches presented in the previous section, more elaborate linear estimators derived from both deterministic and statistical viewpoints are proposed in [19–21]. Such approaches include Least Squares (LS), Regularized LS, Minimum Mean-Squared Error (MMSE) and Mismatched MMSE. These can all be expressed under the following general formulation:

$$\mathbf{A}_{\text{gen}} = \mathbf{B}(\mathbf{G}^{\text{H}}\mathbf{G} + \mathbf{R})^{-1}\mathbf{G}^{\text{H}} \tag{8.17}$$

where $\mathbf{B}$, $\mathbf{G}$ and $\mathbf{R}$ are matrices that vary according to each estimator as expressed in Table 8.1, where $\mathbf{0}_L$ is the all-zeros $L \times L$ matrix.

The LS estimator discussed in [19] is theoretically unbiased. However, as shown in [21], it is not possible to apply the LS estimator to LTE directly, because the expression $(\mathbf{F}_p\mathbf{F}_p^{\text{H}})^{-1}$ is ill-conditioned due to the unused portion of the spectrum corresponding to the unmodulated subcarriers.

To counter this problem, the classical robust regularized LS estimator can be used instead, so as to yield a better conditioning of the matrix to be inverted. A regularization matrix $\alpha\mathbf{I}_L$ is introduced [22] where α is a constant (computed off-line) chosen to optimize the performance of the estimator in a given Signal-to-Noise Ratio (SNR) working range.

The MMSE estimator belongs to the class of statistical estimators. Unlike deterministic LS and its derivations, statistical estimators need knowledge of the second-order statistics (PDP and noise variance) of the channel in order to perform the estimation process, normally with much better performance compared to deterministic estimators. However, second-order statistics vary as the propagation conditions change and therefore need appropriate re-estimation regularly. For this reason statistical estimators are, in general, more complex due

[9] W is the spacing (in terms of number of subcarriers) between reference symbols.

Table 8.1: Linear estimators.

Interpolation method	Components of interpolation filter (see Equation (8.17)) **B**	**G**	**R**
Simple interpolator	$\mathbf{A}$	$\mathbf{I}_P$	$\mathbf{0}_L$
FFT	$\frac{1}{P}\mathbf{F}_L\mathbf{F}_p^{\mathrm{H}}$	$\mathbf{I}_P$	$\mathbf{0}_L$
LS	$\mathbf{F}_L$	$\mathbf{F}_p$	$\mathbf{0}_L$
Regularized LS	$\mathbf{F}_L$	$\mathbf{F}_p$	$\alpha\mathbf{I}_L$
MMSE	$\mathbf{F}_L$	$\mathbf{F}_p$	$\sigma^2_{\tilde{\mathbf{z}}_p}\,\mathbf{C}_{\mathbf{h}}{}^{-1}$
Mismatched MMSE	$\mathbf{F}_L$	$\mathbf{F}_p$	$\sigma^2_{\tilde{\mathbf{z}}_p}/\sigma^2_{\mathbf{h}}\cdot\mathbf{I}_L$

to the additional burden of estimating the second-order statistics and computing the filter coefficients.

A mismatched MMSE estimator avoids the estimation of the second-order channel statistics and the consequent on-line inversion of an $L \times L$ matrix (as required in the straightforward application of MMSE) by assuming that the channel PDP is uniform[10] [20]. This estimator is, in practice, equivalent to the regularized LS estimator where the only difference lies in the fact that $\alpha = \sigma^2_{\tilde{\mathbf{z}}_p}/\sigma^2_{\mathbf{h}}$ is estimated and therefore adapted. The mismatched MMSE formulation offers the advantage that the filter coefficients can be computed to be real numbers because the uniform PDP is symmetric. Moreover, since the length of the CIR is small compared to the FFT size, the matrix A_{gen} can be considered to be 'low-density', so storing only the significant coefficients can reduce the complexity.

However, LTE does not use an exactly uniform pattern of reference symbols; for the cell-specific RSs this is the case around the d.c. subcarrier which is not transmitted. This implies that a larger number of coefficients needs to be stored.

8.4.3 Performance Comparison

For the sake of comparison between the performance of the different classes of estimator, we introduce the Truncated Normalized Mean Squared Error (TNMSE).

For each estimator, the TNMSE is computed from its covariance matrix $\mathbf{C}_{\tilde{\mathbf{z}}}$ and the true CTF $\mathbf{z}$ as follows:

$$\mathrm{TNMSE}_{\widehat{\mathbf{z}}} = \frac{\mathrm{Ttr}(\mathbf{C}_{\tilde{\mathbf{z}}})}{\mathrm{Ttr}(\mathbf{F}_L\mathbf{C}_{\mathrm{h}}\mathbf{F}_L^{\mathrm{H}})} \tag{8.18}$$

where $\mathrm{Ttr}\{\cdot\}$ denotes the truncated trace operator, where the truncation is such that only the K used subcarriers are included.

Figure 8.9 shows the performance of an LTE FDD downlink with 10 MHz transmission bandwidth ($N = 1024$) and Spatial Channel Model-A (SCM-A – see Section 20.3.4).

It can be seen that the IFFT and linear interpolation methods yield the lowest performance. The regularized LS and the mismatched MMSE perform equally and the curve of the latter is therefore omitted. As expected, the optimal MMSE estimator outperforms any other estimator.

[10] This results in the second-order statistics of the channel having the structure of an identity matrix.

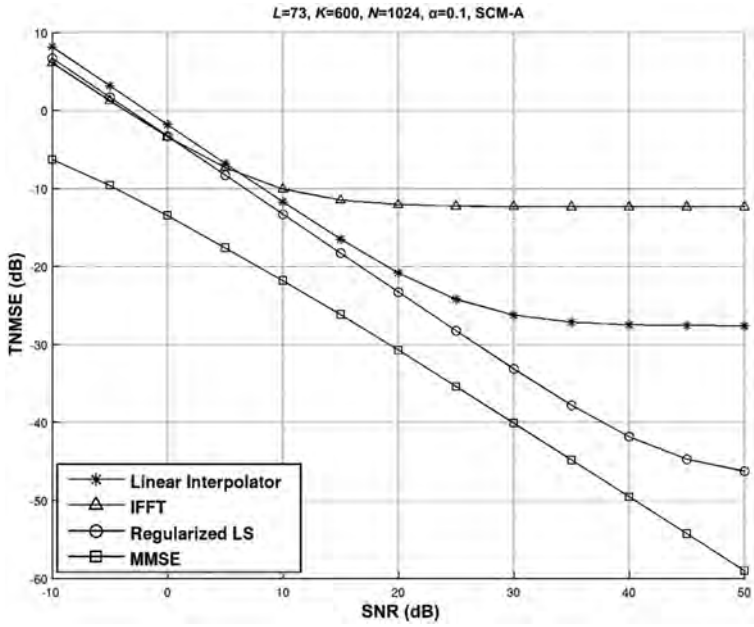


Figure 8.9: Frequency-domain channel estimation performance.

The TNMSE computed over all subcarriers actually hides the behaviour of each estimator in relation to a well-known problem of frequency-domain channel estimation techniques: the band-edge effect. This can be represented by the Gibbs [23] phenomenon in a finite-length Fourier series approximation; following this approach, Figure 8.10 shows that MMSE-based channel estimation suffers the least band-edge degradation, while all the other methods presented are highly adversely affected.

8.5 Time-Domain Channel Estimation

The main benefit of time-domain channel estimation is the possibility to enhance the channel estimate of one OFDM symbol containing reference symbols by exploiting its time correlation with the channel at previous OFDM symbols containing reference symbols. This requires sufficient memory for buffering soft values of data over several OFDM symbols while the channel estimation is carried out.

However, the correlation between consecutive symbols decreases as the UE speed increases, as expressed by Equation (8.6), and this sets a limit on the possibilities for time-domain filtering in high-mobility conditions.

Time-domain filtering is applied to the CIR estimate, rather than to the CTF estimate in the frequency domain. The use of a number of parallel scalar filters equal to the channel length L does not imply a loss of optimality, because of the WSSUS assumption.

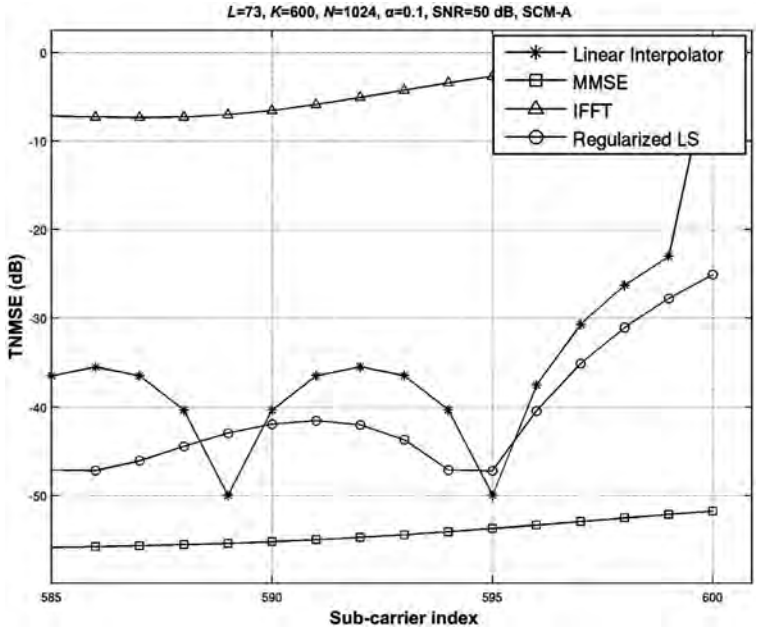


Figure 8.10: Band-edge behaviour of frequency-domain channel estimation.

8.5.1 Finite and Infinite Length MMSE

The statistical time-domain filter which is optimal in terms of Mean Squared Error (MSE) can be approximated in the form of a finite impulse response filter [24]. The channel at the l^{th} tap position and at time instant n is estimated as

$$\hat{\hat{h}}_{l,n} = \mathbf{w}_l^H \hat{\mathbf{h}}_{l,n}^M \tag{8.19}$$

where $\hat{\hat{h}}_l$ is the smoothed CIR l^{th} tap estimate which exploits the vector $\hat{\mathbf{h}}_{l,n}^M = [\hat{h}_{l,n}, \ldots, \hat{h}_{l,n-M+1}]^{\text{T}}$ of length M of the channel tap h_l across estimates at M time instants.[11] This is obtained by inverse Fourier transformation of, for example, any frequency-domain technique illustrated in Section 8.4 or even a raw estimate obtained by RS decorrelation.

The $M \times 1$ vector of Finite Impulse Response (FIR) filter coefficients $\mathbf{w}_l$ is given by

$$\mathbf{w}_l = (\mathbf{R}_h + \sigma_n^2 \mathbf{I})^{-1} \mathbf{r}_h \tag{8.20}$$

where $\mathbf{R}_h = \mathbb{E}[\mathbf{h}_l^M (\mathbf{h}_l^M)^{\text{H}}]$ is the l^{th} channel tap $M \times M$ correlation matrix, σ_n^2 the additive noise variance and $\mathbf{r}_h = \mathbb{E}[\mathbf{h}_l^M h_{l,n}^*]$ (the $M \times 1$ correlation vector between the l^{th} tap of the current channel realization and M previous realizations including the current one).

[11] $h_{l,k}$ is the l^{th} component of the channel vector $\mathbf{h}_k$ at time instant k. $\hat{h}_{l,k}$ is its estimate. Note that for the frequency-domain treatment, the time index was dropped.

In practical cases, the FIR filter length M is dimensioned according to a performance-complexity trade-off as a function of UE speed.

By setting M infinite, the upper bound on performance is obtained.

The MSE performance of the finite-length estimator of a channel of length L can be analytically computed as

$$\epsilon^{(M)} = 1 - \frac{1}{\sigma_h^2} \sum_{l=0}^{L-1} \mathbf{w}_l^{\mathrm{H}} \mathbf{r}_h \tag{8.21}$$

The upper bound given by an infinite-length estimator is therefore given by

$$\epsilon^{(\infty)} = \lim_{M \to \infty} \epsilon^{(M)} \tag{8.22}$$

From Equation (8.20) it can be observed that, unlike frequency-domain MMSE filtering, the size of the matrix to be inverted for a finite-length time-domain MMSE estimator is independent of the channel length L but dependent on the chosen FIR order M. Similarly to the frequency-domain counterpart, the time-domain MMSE estimator requires knowledge of the PDP, the UE speed and the noise variance.

Figure 8.11 shows the performance of time-domain MMSE channel estimation as a function of filter length M (Equation (8.21)) for a single-tap channel with a classical Doppler spectrum for low UE speed. The performance bounds derived for an infinite-length filter in each case are also indicated.

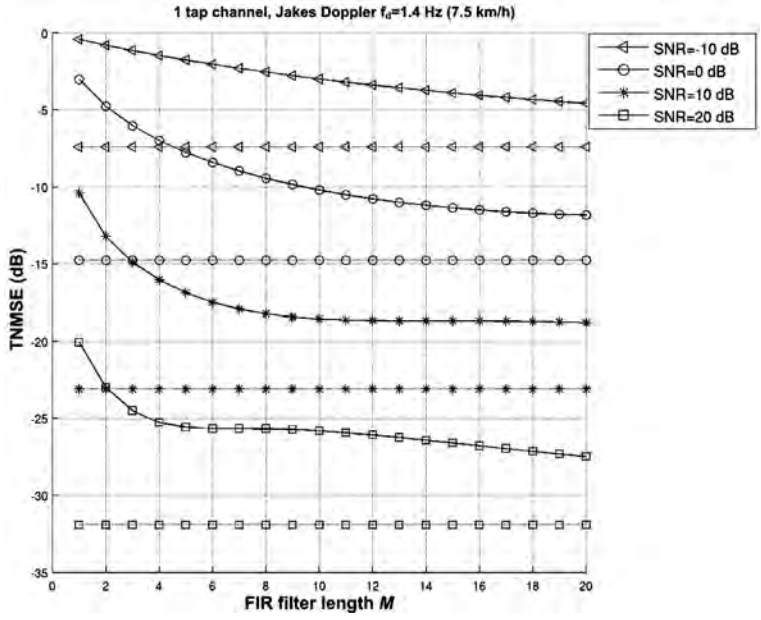


Figure 8.11: Time-domain channel estimation performance.

8.5.2 Normalized Least-Mean-Square

As an alternative to time-domain MMSE channel estimation, an adaptive estimation approach can be considered which does not require knowledge of second-order statistics of both channel and noise. A feasible solution is the Normalized Least-Mean-Square (NLMS) estimator.

It can be expressed exactly as in Equation (8.19) but with the $M \times 1$ vector of filter coefficients $\mathbf{w}$ updated according to

$$\mathbf{w}_{l,n} = \mathbf{w}_{l,n-1} + \mathbf{k}_{l,n-1} e_{l,n} \tag{8.23}$$

where M here denotes the NLMS filter order. The $M \times 1$ updated gain vector is computed according to the well-known NLMS adaptation:

$$\mathbf{k}_{l,n} = \frac{\mu}{||\hat{\mathbf{h}}_{l,n}||^2} \hat{\mathbf{h}}_{l,n}^{M} \tag{8.24}$$

where μ is an appropriately chosen step adaptation, $\hat{\mathbf{h}}_{l,n}^{M}$ is defined as for Equation (8.19) and

$$e_{l,n} = \hat{h}_{l,n} - \hat{h}_{l,n-1} \tag{8.25}$$

It can be observed that the time-domain NLMS estimator requires much lower complexity compared to time-domain MMSE because no matrix inversion or a priori statistical knowledge is required.

Other adaptive approaches could also be considered, such as Recursive Least Squares (RLS) and Kalman-based filtering. Although more complex than NLMS, the Kalman filter is a valuable candidate and is reviewed in detail in [25].

8.6 Spatial-Domain Channel Estimation

It is assumed that an LTE UE has multiple receiving antennas. Consequently, whenever the channel is correlated in the spatial domain, the correlation can be exploited to provide a further means for enhancing the channel estimate.

If it is desired to exploit spatial correlation, a natural approach is again offered by spatial domain MMSE filtering [26].

We consider here the case of a MIMO OFDM communication system with N subcarriers, N_{Tx} transmitting antennas and N_{Rx} receiving antennas. The $N_{\mathrm{Rx}} \times 1$ received signal vector at subcarrier k containing the RS sequence can be written as

$$\mathbf{r}(k) = \frac{1}{\sqrt{N_{\mathrm{Tx}}}} \mathbf{W}(k)\mathbf{s}(k) + \mathbf{n}(k) \tag{8.26}$$

where $\mathbf{W}(k)$ is the $N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}$ channel frequency response matrix at RS subcarrier k, $\mathbf{s}(k)$ is the $N_{\mathrm{Tx}} \times 1$ known zero-mean and unit-variance transmitted RS sequence at subcarrier k, and $\mathbf{n}(k)$ is the $N_{\mathrm{Rx}} \times 1$ complex additive white Gaussian noise vector with zero mean and variance σ_n^2.

The CTF at subcarrier k, $\mathbf{W}(k)$, is obtained by DFT from the CIR matrix at the l^{th} tap $\mathbf{H}_l$ as

$$\mathbf{W}(k) = \sum_{l=0}^{L-1} \mathbf{H}_l \exp\left[-j2\pi \frac{lk}{N}\right] \tag{8.27}$$

The vector $\mathbf{h}$ is obtained by rearranging the elements of all $\mathbf{H}_l$ channel tap matrices as follows:

$$\mathbf{h} = [\text{vec}(\mathbf{H}_0)^{\text{T}}, \text{vec}(\mathbf{H}_1)^{\text{T}}, \ldots, \text{vec}(\mathbf{H}_{L-1})^{\text{T}}]^{\text{T}} \tag{8.28}$$

The correlation matrix of $\mathbf{h}$ is given by

$$\mathbf{C}_h = \mathbb{E}[\mathbf{h}\mathbf{h}^{\text{H}}] \tag{8.29}$$

Using Equations (8.27) and (8.28), Equation (8.26) can now be rewritten as

$$\mathbf{r}(k) = \frac{1}{\sqrt{N_{\text{Tx}}}}\mathbf{G}(k)\mathbf{h} + \mathbf{n}(k) \tag{8.30}$$

where $\mathbf{G}(k) = [\mathbf{D}_0(k), \mathbf{D}_1(k), \ldots, \mathbf{D}_{L-1}(k)]$ and $\mathbf{D}_l(k) = \exp[-j2\pi(lk/N)]\mathbf{s}^{\text{T}}(k) \otimes \mathbf{I}_{N_{\text{Rx}}}$.

Rearranging the received signal matrix $\mathbf{G}$ and the noise matrix at all N subcarriers into a vector as follows,

$$\mathbf{r} = [\mathbf{r}^{\text{T}}(0), \mathbf{r}^{\text{T}}(1), \ldots, \mathbf{r}^{\text{T}}(N-1)]^{\text{T}} \tag{8.31}$$

Equation (8.30) can be rewritten more compactly as

$$\mathbf{r} = \frac{1}{\sqrt{N_{\text{Tx}}}}\mathbf{G}\mathbf{h} + \mathbf{n} \tag{8.32}$$

Hence, the spatial domain MMSE estimation of the rearranged channel impulse vector $\mathbf{h}$ can be simply obtained by

$$\hat{\mathbf{h}} = \mathbf{Q}\mathbf{r} \tag{8.33}$$

with

$$\mathbf{Q} = \frac{1}{\sqrt{N_{\text{Tx}}}}\mathbf{C}_h\mathbf{G}^{\text{H}}\left(\frac{1}{N_{\text{Tx}}}\mathbf{G}\mathbf{C}_h\mathbf{G}^{\text{H}} + \sigma_n^2\mathbf{I}_{N\cdot N_{\text{Rx}}}\right)^{-1} \tag{8.34}$$

Therefore the NMSE can be computed as

$$\text{NMSE}_{\text{SD-MMSE}} = \frac{\text{tr}(\mathbf{C}_{\text{SD-MMSE}})}{\text{tr}(\mathbf{C}_h)} \tag{8.35}$$

where $\mathbf{C}_{\text{SD-MMSE}}$ is the error covariance matrix.

Figure 8.12 shows the spatial domain MMSE performance given by Equation (8.35) compared to the performance obtained by the ML channel estimation on the subcarriers which carry RS.

8.7 Advanced Techniques

The LTE specifications do not mandate any specific channel estimation technique, and there is therefore complete freedom in implementation provided that the performance requirements are met and the complexity is affordable.

Particular aspects, such as channel estimation based on the UE-specific RSs, band-edge effect reduction or bursty reception, might require further improvements which go beyond the techniques described here.

Blind and semi-blind techniques are promising for some such aspects, as they try to exploit not only the a priori knowledge of the RSs but also the unknown data structure. A comprehensive analysis is available in [27] and references therein.

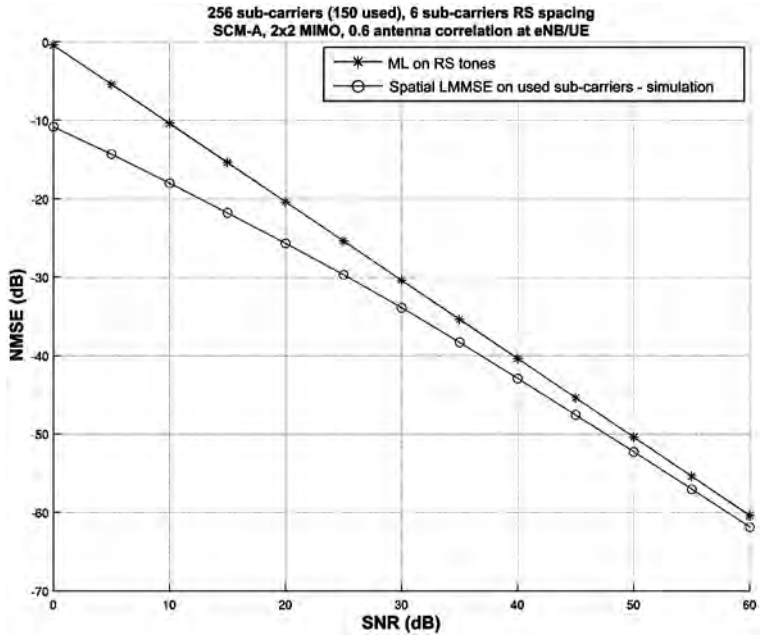


Figure 8.12: Spatial-domain channel estimation performance: CIR NMSE versus SNR.

References[12]

[1] A. Vosoughi and A. Scaglione, 'The Best Training Depends on the Receiver Architecture' in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*, (Montreal, Canada), May 2004.

[2] A. Vosoughi and A. Scaglione, 'Everything You Always Wanted to Know about Training: Guidelines Derived using the Affine Precoding Framework and the CRB'. *IEEE Trans. on Signal Processing*, Vol. 54, pp. 940–954, March 2006.

[3] 3GPP Technical Specification 25.211, 'Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD)', www.3gpp.org.

[4] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: John Wiley & Sons, Ltd, 1949.

[5] D. T. M. Slock, 'Signal Processing Challenges for Wireless Communications' in *Proc. International Symposium on Control, Communications and Signal Processing (ISCCSP' 2004),* (Hammamet, Tunisia), March 2004.

[6] 3GPP Technical Specification 36.211, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation', www.3gpp.org.

[7] R. Negi and J. Cioffi, 'Pilot Tone Selection for Channel Estimation in a Mobile OFDM System'. *IEEE Trans. on Consumer Electronics*, Vol. 44, pp. 1122–1128, August 1998.

[12] All web sites confirmed 1st March 2011.

[8] I. Barhumi, G. Leus and M. Moonen, 'Optimal Training Design for MIMO OFDM Systems in Mobile Wireless Channels'. *IEEE Trans. on Signal Processing*, Vol. 51, pp. 1615–1624, June 2003.

[9] 3GPP Technical Report 25.913, 'Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN) (Release 7)', www.3gpp.org.

[10] 3GPP Technical Specification 36.101, 'User Equipment (UE) Radio Transmission and Reception', www.3gpp.org.

[11] Motorola, Nortel, Broadcomm, Nokia, NSN, NTT DoCoMo, NEC, Mitsubishi, Alcatel-Lucent, CATT, Huawei, Sharp, Texas Instrument, ZTE, Panasonic, Philips and Toshiba, 'R1-081108: Way Forward on Dedicated Reference Signal Design for LTE Downlink with Normal CP', www.3gpp.org, 3GPP TSG RAN WG1, meeting 52, Sorrento, Italy, February 2008.

[12] NTT DoCoMo and NEC, 'R1-094337: Remaining Issues for Rel. 9 Downlink DM-RS Design', www.3gpp.org, 3GPP TSG RAN WG1, meeting 58bis, Miyazaki, Japan, Octboer 2009.

[13] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösty, D. Laurenson, G. Matz, F. Molisch, C. Oestges, and H. Ožcelik, 'Survey of Channel and Radio Propagation Models for Wireless MIMO Systems'. *EURASIP Journal on Wireless Communications and Networking*, pp. 957–1000, July 2007.

[14] J. G. Proakis, *Digital Communications*. New York: McGraw Hill, 1995.

[15] R. H. Clarke, 'A Statistical Theory of Mobile-radio Reception'. *Bell Syst. Tech. J.*, pp. 957–1000, July 1968.

[16] W. C. Jakes, *Microwave Mobile Communications*. New York: John Wiley & Sons, Ltd/Inc., 1974.

[17] P. Hoeher,'A statistical discrete-time model for the WSSUS multipath channel' in *IEEE Transactions on Vehicular Technology*, Vol 41, No. 4, pp. 461–468, 1992.

[18] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Börjesson, 'On Channel Estimation in OFDM Systems' in *Proc. VTC'1995, Vehicular Technology Conference (VTC 1995),* Chicago, USA, July 1995.

[19] M. Morelli and U. Mengali, 'A Comparison of Pilot-aided Channel Estimation Methods for OFDM Systems'. *IEEE Trans. on Signal Processing*, Vol. 49, pp. 3065–3073, December 2001.

[20] P. Hoher, S. Kaiser and P. Robertson, 'Pilot-Symbol-aided Channel Estimation in Time and Frequency' in *Proc. Communication Theory Mini-Conf. (CTMC) within IEEE Global Telecommun. Conf. (Globecom 1997)* (Phoenix, USA), July 1997.

[21] A. Ancora, C. Bona and D. T. M. Slock, 'Down-Sampled Impulse Response Least-Squares Channel Estimation for LTE OFDMA' in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007),* Honolulu, Hawaii, April 2007.

[22] P. C. Hansen, 'Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion', Society for Industrial Mathematics, 1998.

[23] J. W. Gibbs, 'Fourier Series'. *Nature*, Vol. 59, p. 200, 1898.

[24] D. Schafhuber, G. Matz and F. Hlawatsch, 'Adaptive Wiener Filters for Time-varying Channel Estimation in Wireless OFDM Systems' in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003),* Hong Kong, April 2003.

[25] J. Cai, X. Shen and J. W. Mark, 'Robust Channel Estimation for OFDM Wireless Communication Systems. An H∞ Approach'. *IEEE Trans. Wireless Communications*, Vol. 3, November 2004.

[26] H. Zhang, Y. Li, A. Reid and J. Terry, 'Channel Estimation for MIMO OFDM in Correlated Fading Channels' in *Proc. 2005 IEEE International Conference on Communications (ICC 2005),* South Korea, May 2005.

[27] H. Bölcskei, D. Gesbert and C. Papadias, *Space-time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge: Cambridge University Press, 2006.

9

Downlink Physical Data and Control Channels

Matthew Baker and Tim Moulsley

9.1 Introduction

Chapters 7 and 8 have described the signals which enable User Equipment (UEs) to synchronize with the network and estimate the downlink radio channel in order to be able to demodulate data. This chapter first reviews the downlink physical channels which transport the data and then explains the control-signalling channels; the latter support the data channels by indicating the particular time-frequency transmission resources to which the data is mapped and the format in which the data itself is transmitted.

9.2 Downlink Data-Transporting Channels

9.2.1 Physical Broadcast Channel (PBCH)

In cellular systems, the basic System Information (SI) which allows the other channels in the cell to be configured and operated is usually carried by a Broadcast CHannel (BCH). Therefore the achievable coverage for reception of the BCH is crucial to the successful operation of such cellular communication systems; LTE is no exception. As already noted in Chapter 3, the broadcast (SI) is divided into two categories:

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

- The 'Master Information Block' (MIB), which consists of a limited number of the most frequently transmitted parameters essential for initial access to the cell,[1] and is carried on the Physical Broadcast CHannel (PBCH).
- The other System Information Blocks (SIBs) which, at the physical layer, are multiplexed with unicast data transmitted on the Physical Downlink Shared CHannel (PDSCH) as discussed in Section 9.2.2.2.

This section focuses in particular on the PBCH, the design of which reflects some specific requirements:

- Detectability without prior knowledge of the system bandwidth;
- Low system overhead;
- Reliable reception right to the edge of the LTE cells;
- Decodability with low latency and low impact on UE battery life.

The resulting overall PBCH structure is shown in Figure 9.1.

Detectability without the UE having prior knowledge of the system bandwidth is achieved by mapping the PBCH only to the central 72 subcarriers of the OFDM[2] signal (which corresponds to the minimum possible LTE system bandwidth of 6 Resource Blocks (RBs)), regardless of the actual system bandwidth. The UE will have first identified the system centre-frequency from the synchronization signals as described in Chapter 7.

Low system overhead for the PBCH is achieved by deliberately keeping the amount of information carried on the PBCH to a minimum, since achieving stringent coverage requirements for a large quantity of data would result in a high system overhead. The size of the MIB is therefore just 14 bits, and, since it is repeated every 40 ms, this corresponds to a data rate on the PBCH of just 350 bps.

The main mechanisms employed to facilitate reliable reception of the PBCH in LTE are time diversity, Forward Error Correction (FEC) coding and antenna diversity.

Time diversity is exploited by spreading out the transmission of each MIB on the PBCH over a period of 40 ms. This significantly reduces the likelihood of a whole MIB being lost in a fade in the radio propagation channel, even when the mobile terminal is moving at pedestrian speeds.

The FEC coding for the PBCH uses a convolutional coder, as the number of information bits to be coded is small; the details of the convolutional coder are explained in Section 10.3.3. The basic code rate is $1/3$, after which a high degree of repetition of the systematic (i.e. information) bits and parity bits is used, such that each MIB is coded at a very low code rate ($1/48$ over a 40 ms period) to give strong error protection.

Antenna diversity may be utilized at both the eNodeB and the UE. The UE performance requirements specified for LTE assume that all UEs can achieve a level of decoding performance commensurate with dual-antenna receive diversity (although it is recognized that in low-frequency deployments, such as below 1 GHz, the advantage obtained from receive antenna diversity is reduced due to the correspondingly higher correlation between

[1] The MIB information consists of the downlink system bandwidth, the PHICH size (Physical Hybrid ARQ Indicator CHannel, see Section 9.3.4), and the most-significant eight bits of the System Frame Number (SFN) – the remaining two bits being gleaned from the 40 ms periodicity of the PBCH.

[2] Orthogonal Frequency Division Multiplexing.

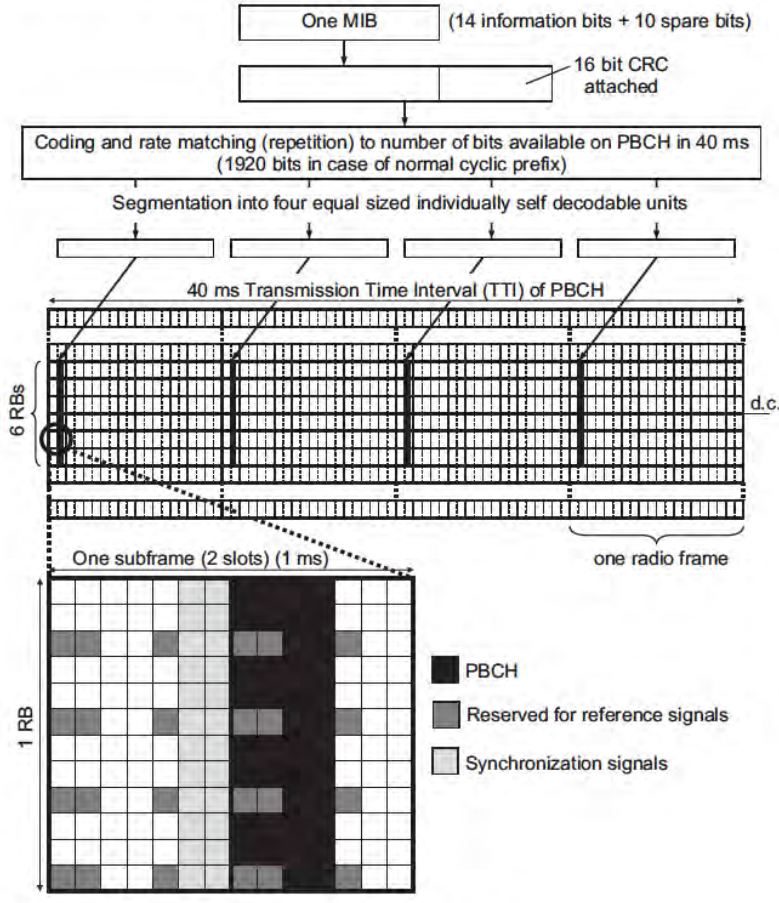


Figure 9.1: PBCH structure.

the antennas); this enables LTE system planners to rely on this level of performance being common to all UEs, thereby enabling wider cell coverage to be achieved with fewer cell sites than would otherwise be possible. Transmit antenna diversity may be also employed at the eNodeB to further improve coverage, depending on the capability of the eNodeB; eNodeBs with two or four transmit antenna ports transmit the PBCH using a Space-Frequency Block Code (SFBC), details of which are explained in Section 11.2.2.1.

The precise set of Resource Elements (REs) used by the PBCH is independent of the number of transmit antenna ports used by the eNodeB; any REs which may be used for Reference Signal (RS) transmission are avoided by the PBCH, irrespective of the actual number of transmit antenna ports deployed at the eNodeB. The number of transmit antenna ports used by the eNodeB must be ascertained blindly by the UE, by performing the decoding

for each SFBC scheme corresponding to the different possible numbers of transmit antenna ports (namely one, two or four). This discovery of the number of transmit antenna ports is further facilitated by the fact that the Cyclic Redundancy Check (CRC) on each MIB is masked with a codeword representing the number of transmit antenna ports.

Finally, achieving low latency and a low impact on UE battery life is also facilitated by the design of the coding outlined above: the low code rate with repetition enables the full set of coded bits to be divided into four subsets, each of which is self-decodable in its own right. Each of these subsets of the coded bits is then transmitted in a different one of the four radio frames during the 40 ms transmission period, as shown in Figure 9.1. This means that if the Signal to Interference Ratio (SIR) of the radio channel is sufficiently good to allow the UE to decode the MIB correctly from the transmission in less than four radio frames, then the UE does not need to receive the other parts of the PBCH transmission in the remainder of the 40 ms period; on the other hand, if the SIR is low, the UE can receive further parts of the MIB transmission, soft-combining each part with those received already, until successful decoding is achieved.

The timing of the 40 ms transmission interval for each MIB on the PBCH is not indicated explicitly to the UE; it is ascertained implicitly from the scrambling and bit positions, which are re-initialized every 40 ms. The UE can therefore initially determine the 40 ms timing by performing four separate decodings of the PBCH using each of the four possible phases of the PBCH scrambling code, checking the CRC for each decoding.

When a UE initially attempts to access a cell by reading the PBCH, a variety of approaches may be taken to carry out the necessary blind decodings. A simple approach is always to perform the decoding using a soft combination of the PBCH over four radio frames, advancing a 40 ms sliding window one radio frame at a time until the window aligns with the 40 ms period of the PBCH and the decoding succeeds. However, this would result in a 40–70 ms delay before the PBCH can be decoded. A faster approach would be to attempt to decode the PBCH from the first single radio frame, which should be possible provided the SIR is sufficiently high; if the decoding fails for all four possible scrambling code phases, the PBCH from the first frame could be soft-combined with the PBCH bits received in the next frame – there is a 3-in-4 chance that the two frames contain data from the same transport block. If decoding still fails, a third radio frame could be combined, and failing that a fourth. It is evident that the latter approach may be much faster (potentially taking only 10 ms), but on the other hand requires slightly more complex logic.

9.2.2 Physical Downlink Shared CHannel (PDSCH)

The Physical Downlink Shared CHannel (PDSCH) is the main data-bearing downlink channel in LTE. It is used for all user data, as well as for broadcast system information which is not carried on the PBCH, and for paging messages – there is no specific physical layer paging channel in LTE. The use of the PDSCH for user data is explained in Section 9.2.2.1; the use of the PDSCH for system information and paging is covered in Section 9.2.2.2.

Data is transmitted on the PDSCH in units known as *Transport Blocks* (TBs), each of which corresponds to a Medium Access Control (MAC) layer Protocol Data Unit (PDU) as described in Section 4.4. Transport blocks may be passed down from the MAC layer to the physical layer once per Transmission Time Interval (TTI), where a TTI is 1 ms, corresponding to the subframe duration.

9.2.2.1 General Use of the PDSCH

When employed for user data, one or, at most, two TBs can be transmitted per UE per subframe, depending on the transmission mode selected for the PDSCH for each UE. The transmission mode configures the multi-antenna transmission scheme usually applied:[3]

Transmission Mode 1: Transmission from a single eNodeB antenna port;

Transmission Mode 2: Transmit diversity (see Section 11.2.2.1);

Transmission Mode 3: Open-loop spatial multiplexing (see Section 11.2.2.2);

Transmission Mode 4: Closed-loop spatial multiplexing (see Section 11.2.2.2);

Transmission Mode 5: Multi-User Multiple-Input Multiple-Output (MU-MIMO) (see Section 11.2.3);

Transmission Mode 6: Closed-loop rank-1 precoding (see Section 11.2.2.2);

Transmission Mode 7: Transmission using UE-specific RSs with a single spatial layer (see Sections 8.2 and 11.2.2.3);

Transmission Mode 8: Introduced in Release 9, transmission using UE-specific RSs with up to two spatial layers (see Sections 8.2.3 and 11.2.2.3);

Transmission Mode 9: Introduced in Release 10, transmission using UE-specific RSs with up to eight spatial layers (see Sections 29.1 and 29.3).

With the exception of transmission modes 7, 8 and 9, the phase reference for demodulating the PDSCH is given by the cell-specific Reference Signals (RSs) described in Section 8.2.1, and the number of eNodeB antenna ports used for transmission of the PDSCH is the same as the number of antenna ports used in the cell for the PBCH. In transmission modes 7, 8 and 9, UE-specific RSs (see Sections 8.2.2, 8.2.3 and 29.1.1 respectively) provide the phase reference for the PDSCH. The configured transmission mode also controls the formats of the associated downlink control signalling messages, as described in Section 9.3.5.1, and the modes of channel quality feedback from the UE (see Section 10.2.1).

After channel coding (see Section 10.3.2) and mapping to spatial layers according to the selected transmission mode, the coded PDSCH data bits are mapped to modulation symbols depending on the modulation scheme selected for the current radio channel conditions and required data rate.

The modulation order may be selected between two bits per symbol (using QPSK (Quadrature Phase Shift Keying)), four bits per symbol (using 16QAM (Quadrature Amplitude Modulation)) and six bits per symbol (using 64QAM); constellation diagrams for these modulation schemes are illustrated in Figure 9.2. Support for reception of 64QAM modulation is mandatory for all classes of LTE UE.

The REs used for the PDSCH can be any which are not reserved for other purposes (i.e. RSs, synchronization signals, PBCH and control signalling). Thus when the control

[3] In addition to the transmission schemes listed here for each mode, transmission modes 3 to 9 also support the use of transmit diversity as a 'fallback' technique; this is useful, for example, when radio conditions are temporarily inappropriate for the usual scheme, or to ensure that a common scheme is available during reconfiguration of the transmission mode.

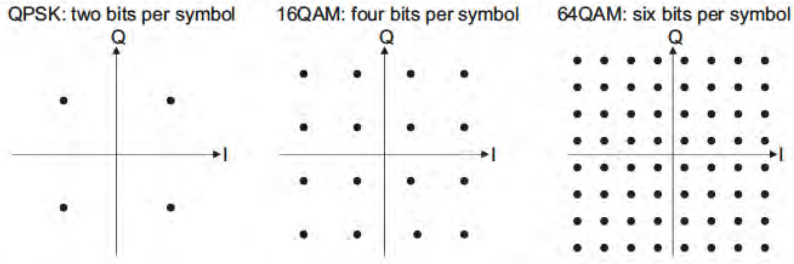


Figure 9.2: Constellations of modulation schemes applicable to PDSCH transmission.

signalling informs a UE that a particular pair of RBs[4] in a subframe are allocated to that UE, it is only the *available* REs within those RBs which actually carry PDSCH data.

Normally the allocation of pairs of RBs to PDSCH transmission for a particular UE is signalled to the UE by means of dynamic control signalling transmitted at the start of the relevant subframe using the Physical Downlink Control Channel (PDCCH), as described in Section 9.3.

The mapping of data to physical RBs can be carried out in one of two ways: *localized mapping* and *distributed mapping*.[5]

Localized resource mapping entails allocating all the available REs in a pair of RBs to the same UE. This is suitable for most scenarios, including the use of dynamic channel-dependent scheduling according to frequency-specific channel quality information reported by the UE (see Sections 10.2.1 and 12.4).

Distributed resource mapping entails separating in frequency the two physical RBs comprising each pair, with a frequency-hop occurring at the slot boundary in the middle of the subframe, as shown in Figure 9.3. This is a useful means of obtaining frequency diversity for small amounts of data which would otherwise be constrained to a narrow part of the downlink bandwidth and would therefore be more susceptible to narrow-band fading. An example of a typical use for this transmission mode could be a Voice-over-IP (VoIP) service, where, in order to minimize overhead, certain frequency resources may be 'semi-persistently scheduled' (see Section 4.4.2.1) – in other words, certain RBs in the frequency domain are allocated on a periodic basis to a specific UE by Radio Resource Control (RRC) signalling rather than by dynamic PDCCH signalling. This means that the transmissions are not able to benefit from dynamic channel-dependent scheduling, and therefore the frequency diversity which is achieved through distributed mapping is a useful tool to improve performance. Moreover, as the amount of data to be transmitted per UE for a VoIP service is small (typically sufficient to occupy only one or two pairs of RBs in a given subframe), the degree of frequency diversity obtainable via localized scheduling is very limited.

[4]The term 'pair of RBs' here means a pair of resource blocks which occupy the same set of 12 subcarriers and are contiguous in time, thus having a duration of one subframe.

[5]Distributed mapping is not supported in conjunction with UE-specific RSs in transmission modes 8 and 9.

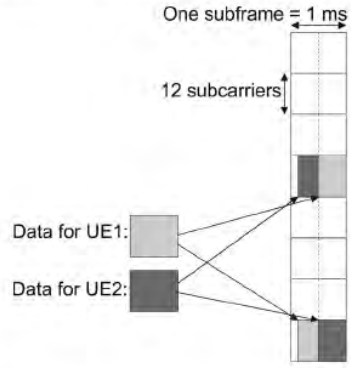
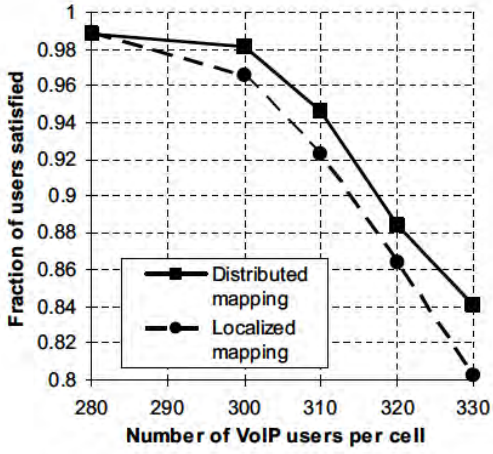


Figure 9.3: Frequency-distributed data mapping in LTE downlink.

The potential increase in the number of VoIP users which can be accommodated in a cell as a result of using distributed resource mapping as opposed to localized resource mapping is illustrated by way of example in Figure 9.4.



Parameter	Value
Carrier frequency	2 GHz
Bandwidth	5 MHz
Channel model	Urban micro, 3 km/h
eNB transmit power	43 dBm
VoIP codec	12.2 kbps
Voice activity factor	0.43
Modulation scheme	QPSK
Code rate	2/3
Satisfaction criterion	98% packets within 50 ms

Figure 9.4: Example of increase in VoIP capacity arising from frequency-distributed resource mapping.

9.2.2.2 Special Uses of the PDSCH

As noted above, the PDSCH is used for some special purposes in addition to normal user data transmission.

One such use is for the broadcast system information (i.e. SIBs) that is not carried on the PBCH. The RBs used for broadcast data of this sort are indicated by signalling messages on the PDCCH in the same way as for other PDSCH data, except that the identity indicated on the PDCCH is not the identity of a specific UE but is, rather, a designated broadcast identity known as the System Information Radio Network Temporary Identifier (SI-RNTI), which is fixed in the specifications (see Section 7.1 of [1]) and therefore known a priori to all UEs. Some constraints exist as to which subframes may be used for SI messages on the PDSCH; these are explained in Section 3.2.2.

Another special use of the PDSCH is paging, as no separate physical channel is provided in LTE for this purpose. In previous systems such as WCDMA,[6] a special 'Paging Indicator Channel' was provided, which was specially designed to enable the UE to wake up its receiver periodically for a very short period of time, in order to minimize the impact on battery life; on detecting a paging indicator (typically for a group of UEs), the UE would then keep its receiver switched on to receive a longer message indicating the exact identity of the UE being paged. By contrast, in LTE the PDCCH signalling is already very short in duration, and therefore the impact on UE battery life of monitoring the PDCCH from time to time is low. Therefore the normal PDCCH signalling can be used to carry the equivalent of a paging indicator, with the detailed paging information being carried on the PDSCH in RBs indicated by the PDCCH. In a similar way to broadcast data, paging indicators on the PDCCH use a single fixed identifier called the Paging RNTI (P-RNTI). Rather than providing different paging identifiers for different groups of UEs, different UEs monitor different subframes for their paging messages, as described in Section 3.4. Paging messages may be received in subframes 0, 4, 5 or 9 in each radio frame.

9.2.3 Physical Multicast Channel (PMCH)

In LTE Release 9, the use of a third data-transporting channel became available, namely the Physical Multicast CHannel (PMCH), designed to carry data for Multimedia Broadcast and Multicast Services (MBMS). The PMCH can only be transmitted in certain specific subframes known as MBSFN (Multimedia Broadcast Single Frequency Network) subframes, indicated in the system information carried on the PDSCH. A Release 8 UE must be aware of the possible existence of MBSFN subframes, but is not required to decode the PMCH. The details of the PMCH are explained in Section 13.4.1.

9.3 Downlink Control Channels

9.3.1 Requirements for Control Channel Design

The control channels in LTE are provided to support efficient data transmission. In common with other wireless systems, the control channels convey physical layer messages which cannot be carried sufficiently efficiently, quickly or conveniently by higher layer protocols. The design of the control channels in the LTE downlink aims to balance a number of somewhat conflicting requirements, the most important of which are discussed below.

[6] Wideband Code Division Multiple Access.

9.3.1.1 Physical Layer Signalling to Support the MAC Layer

The general requirement to support MAC operation is very similar to that in WCDMA, but there are a number of differences of detail, mainly arising from the frequency domain resource allocation supported in the LTE multiple access schemes.

The use of the uplink transmission resources on the Physical Uplink Shared Channel (PUSCH) is determined dynamically by an uplink scheduling process in the eNodeB, and therefore physical layer signalling must be provided to indicate to UEs which time/frequency resources they have been granted permission to use.

The eNodeB also schedules downlink transmissions on the PDSCH, and therefore similar physical layer messages from the eNodeB are needed to indicate which resources in the frequency domain contain the downlink data transmissions intended for particular UEs, together with parameters such as the modulation scheme and code rate used for the data. Explicit signalling of this kind avoids the considerable additional complexity which would arise if UEs had to search for their data among all the possible combinations of data packet size, format and resource allocation.

In order to facilitate efficient operation of HARQ,[7] further physical layer signals are needed to convey acknowledgements of uplink data packets received by the eNodeB, and power control commands are needed to ensure that uplink transmissions are made at appropriate power levels (as explained in Section 18.3).

9.3.1.2 Flexibility, Overhead and Complexity

The LTE physical layer specification is intended to allow operation in any system bandwidth from six resource blocks (1.08 MHz) to 110 resource blocks (19.8 MHz). It is also designed to support a range of scenarios including, for example, just a few users in a cell each demanding high data rates, or very many users with low data rates. Considering the possibility that both uplink resource grants and downlink resource allocations could be required for every UE in each subframe, the number of control channel messages carrying resource information could be as many as a couple of hundred if every resource allocation were as small as one resource block. Since every additional control channel message implies additional overhead which consumes downlink resources, it is desirable that the control channel is designed to minimize unnecessary overhead for any given signalling load, whatever the system bandwidth.

Similar considerations apply to the signalling of HARQ acknowledgements for each uplink packet transmission.

Furthermore, as in any mobile communication system, the complexity and power consumption of the terminals are important considerations for LTE. Therefore, the control signalling must be designed so that the necessary scalability and flexibility is achieved without undue decoding complexity.

9.3.1.3 Coverage and Robustness

In order to achieve good coverage it must be possible to configure the system so that the control channels can be received with sufficient reliability over a substantial part of every cell. As an example, if a message indicating resource allocation is not received correctly, then the corresponding data transmission will also fail, with a direct and proportionate impact

[7] Hybrid Automatic Repeat reQuest – see Sections 4.4 and 10.3.2.5.

on throughput efficiency. Techniques such as channel coding and frequency diversity can be used to make the control channels more robust. However, in order to make good use of system resources, it is desirable to be able to adapt the transmission parameters of the control signalling for different UEs or groups of UEs, so that lower code rates and higher power levels are only applied for those UEs for which it is necessary (e.g. near the cell border, where signal levels are likely to be low and interference from other cells high).

Also, it is desirable to avoid unintended reception of control channels from other cells, by applying cell-specific randomization.

9.3.1.4 System-Related Design Aspects

Since the different parts of LTE are intended to provide a complete system, some aspects of control channel design cannot be considered in isolation.

A basic design decision in LTE is that a control channel message is intended to be transmitted to a particular UE (or, in some cases, a group of UEs). Therefore, in order to reach multiple UEs in a cell within a subframe, it must be possible to transmit multiple control channels within the duration of a single subframe. However, in cases where the control channel messages are intended for reception by more than one UE (for example, when relating to the transmission of a SIB on the PDSCH), it is more efficient to arrange for all the UEs to receive a single transmission rather than to transmit the same information to each UE individually. This requires that both common and dedicated control channel messages be supported.

Finally, some scenarios may be characterized by the data arriving at regular intervals, as is typical for VoIP traffic. It is then possible to predict in advance when resources will need to be assigned in the downlink or granted in the uplink, and the number of control channel messages which need to be sent can be reduced by means of 'Semi-Persistent Scheduling' (SPS) as discussed in Section 4.4.2.1.

9.3.2 Control Channel Structure

Three downlink physical control channels are provided in LTE: the Physical Control Format Indicator CHannel (PCFICH), the Physical HARQ Indicator CHannel (PHICH) and the Physical Downlink Control CHannel (PDCCH). In general, the downlink control channels can be configured to occupy the first 1, 2 or 3 OFDM symbols in a subframe, extending over the entire system bandwidth as shown in Figure 9.5.

This flexibility allows the control channel overhead to be adjusted according to the particular system configuration, traffic scenario and channel conditions. There are two special cases: in subframes containing MBSFN transmissions (see Sections 9.2.3 and 13.4.1) there may be 0, 1 or 2 OFDM symbols for control signalling, while for narrow system bandwidths (less than 10 RBs) the number of control symbols is increased and may be 2, 3 or 4 to ensure sufficient coverage at the cell border.

9.3.3 Physical Control Format Indicator CHannel (PCFICH)

The PCFICH carries a Control Format Indicator (CFI) which indicates the number of OFDM symbols (i.e. normally 1, 2 or 3) used for transmission of control channel information in each subframe. In principle, the UE could deduce the value of the CFI without a channel such as

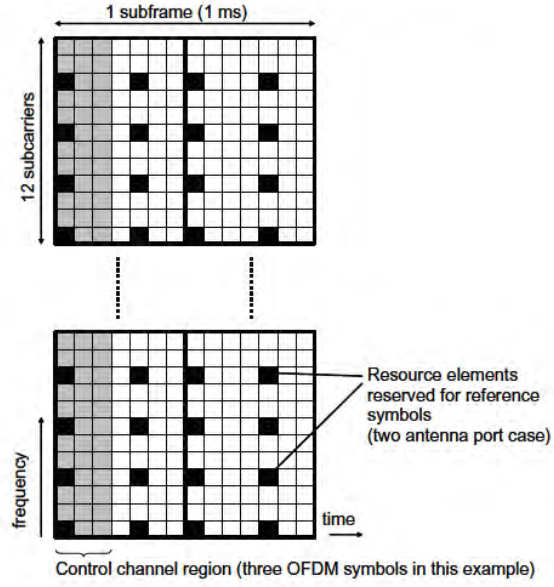


Figure 9.5: The time-frequency region used for downlink control signalling.

the PCFICH, for example by multiple attempts to decode the control channels assuming each possible number of symbols, but this would result in significant additional processing load. Three different CFI values are used in LTE. In order to make the CFI sufficiently robust, each codeword is 32 bits in length, mapped to 16 REs using QPSK modulation. These 16 REs are arranged in groups of 4, known as Resource Element Groups (REGs). The REs occupied by RSs are not included within the REGs, which means that the total number of REGs in a given OFDM symbol depends on whether or not cell-specific RSs are present. The concept of REGs (i.e. mapping in groups of four REs) is also used for the other downlink control channels (the PHICH and PDCCH).

The PCFICH is transmitted on the same set of antenna ports as the PBCH, with transmit diversity being applied if more than one antenna port is used.

In order to achieve frequency diversity, the 4 REGs carrying the PCFICH are distributed across the frequency domain. This is done according to a predefined pattern in the first OFDM symbol in each downlink subframe (see Figure 9.6), so that the UEs can always locate the PCFICH information, which is a prerequisite to being able to decode the rest of the control signalling.

To minimize the possibility of confusion with PCFICH information from a neighbouring cell, a cell-specific frequency offset is applied to the positions of the PCFICH REs; this offset depends on the Physical Cell ID (PCI), which is deduced from the Primary and Secondary Synchronization Signals (PSS and SSS) as explained in Section 7.2. In addition, a cell-specific scrambling sequence (again a function of the PCI) is applied to the CFI codewords, so that the UE can preferentially receive the PCFICH from the desired cell.

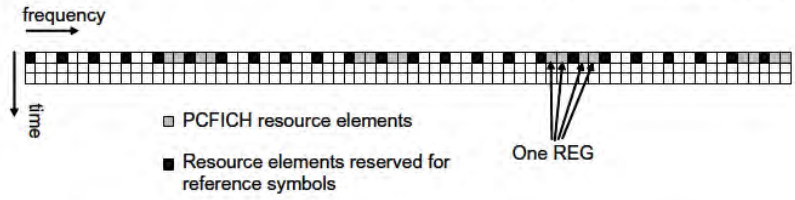


Figure 9.6: PCFICH mapping to Resource Element Groups (REGs).

9.3.4 Physical Hybrid ARQ Indicator Channel (PHICH)

The PHICH carries the HARQ ACK/NACK, which indicates whether the eNodeB has correctly received a transmission on the PUSCH. The HARQ indicator is set to 0 for a positive ACKnowledgement (ACK) and 1 for a Negative ACKnowledgement (NACK). This information is repeated in each of three BPSK[8] symbols.

Multiple PHICHs are mapped to the same set of REs. These constitute a PHICH group, where different PHICHs within the same PHICH group are separated through different complex orthogonal Walsh sequences. Each PHICH is uniquely identified by a PHICH index, which indicates both the group and the sequence. The sequence length is four for the normal cyclic prefix (or two in the case of the extended cyclic prefix). As the sequences are complex, the number of PHICHs in a group (i.e. the number of UEs receiving their acknowledgements on the same set of downlink REs) can be up to twice the sequence length. A cell-specific scrambling sequence is applied.

Factor-3 repetition coding is applied for robustness, resulting in three instances of the orthogonal Walsh code being transmitted for each ACK or NACK. The error rate on the PHICH is intended to be of the order of 10^{-2} for ACKs and as low as 10^{-4} for NACKs. The resulting PHICH construction, including repetition and orthogonal spreading, is shown in Figure 9.7.

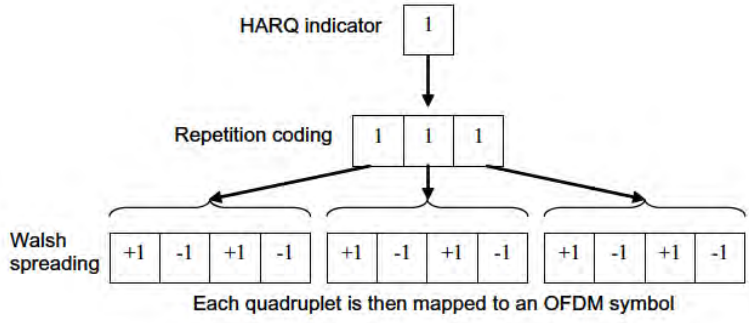


Figure 9.7: An example of PHICH signal construction.

[8]Binary Phase Shift Keying.

The PHICH duration, in terms of the number of OFDM symbols used in the time domain, is configurable (by an indication transmitted on the PBCH), normally to either one or three OFDM symbols.[9] As the PHICH cannot extend into the PDSCH transmission region, the duration configured for the PHICH puts a lower limit on the size of the control channel region at the start of each subframe (as signalled by the PCFICH).

Finally, each of the three instances of the orthogonal code of a PHICH transmission is mapped to an REG on one of the first three OFDM symbols of each subframe,[10] in such a way that each PHICH is partly transmitted on each of the available OFDM symbols. This mapping is illustrated in Figure 9.8 for each possible PHICH duration.

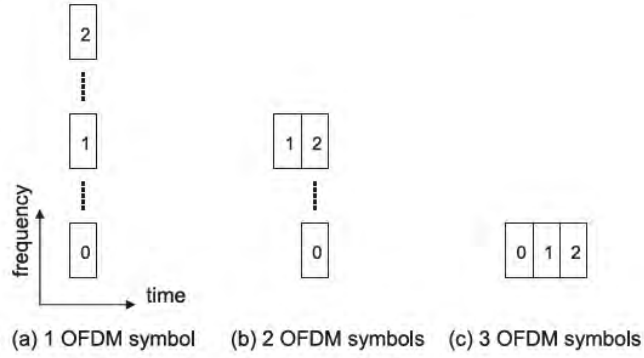


Figure 9.8: Examples of the mapping of the three instances of a PHICH orthogonal code to OFDM symbols, depending on the configured PHICH duration.

The PBCH also signals the number of PHICH groups configured in the cell, which enables the UEs to deduce to which remaining REs in the control region the PDCCHs are mapped.[11]

In order to obviate the need for additional signalling to indicate which PHICH carries the ACK/NACK response for each PUSCH transmission, the PHICH index is implicitly associated with the index of the lowest uplink RB used for the corresponding PUSCH transmission. This relationship is such that adjacent PUSCH RBs are associated with PHICHs in different PHICH groups, to enable some degree of load balancing. However, this mechanism alone is not sufficient to enable multiple UEs to be allocated the same RBs for a PUSCH transmission, as occurs in the case of uplink multi-user MIMO (see Section 16.6); in this case, different cyclic shifts of the uplink demodulation RSs are configured for the different UEs which are allocated the same time-frequency PUSCH resources, and the same

[9] In some special cases, the three-OFDM-symbol duration is reduced to two OFDM symbols; these cases are (i) MBSFN subframes on mixed carriers supporting MBSFN and unicast data, and (ii) the second and seventh subframes in case of frame structure type 2 for Time Division Duplex (TDD) operation.

[10] The mapping avoids REs used for reference symbols or PCFICH.

[11] For Frequency Division Duplex (FDD) operation with Frame Structure Type 1 (see Section 6.2), the configured number of PHICH groups is the same in all subframes; for TDD operation with Frame Structure Type 2, the number of PHICH groups is 0, 1 or 2 times the number signalled by the PBCH, according to the correspondence with uplink subframes.

cyclic shift index is then used to shift the PHICH allocations in the downlink so that each UE receives its ACK or NACK on a different PHICH. This mapping of the PHICH allocations is illustrated in Figure 9.9.

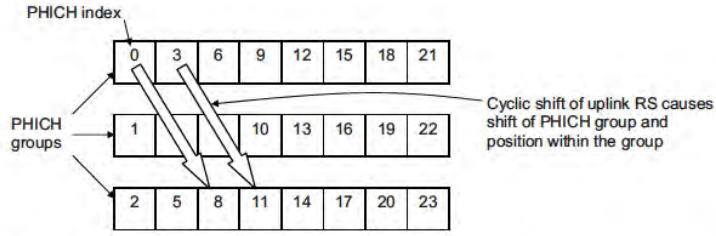


Figure 9.9: Indexing of PHICHs within PHICH groups, and shifting in the case of cyclic shifting of the uplink demodulation reference signals.

The PHICH indexing for the case of uplink MIMO in Release 10 is explained in Section 29.4.1. The use of the PHICH in the case of aggregation of multiple carriers in Release 10 is explained in Section 28.3.1.3.

The PHICHs are transmitted on the same set of antenna ports as the PBCH, and transmit diversity is applied if more than one antenna port is used.

9.3.5 Physical Downlink Control CHannel (PDCCH)

Each PDCCH carries a message known as Downlink Control Information (DCI), which includes resource assignments and other control information for a UE or group of UEs. In general, several PDCCHs can be transmitted in a subframe.

Each PDCCH is transmitted using one or more *Control Channel Elements* (CCEs), where each CCE corresponds to nine REGs. Four QPSK symbols are mapped to each REG.

Four PDCCH formats are supported, as listed in Table 9.1.

Table 9.1: PDCCH formats.

PDCCH format	Number of CCEs (n)	Number of REGs	Number of PDCCH bits
0	1	9	72
1	2	18	144
2	4	36	288
3	8	72	576

CCEs are numbered and used consecutively, and, to simplify the decoding process, a PDCCH with a format consisting of n CCEs may only start with a CCE with a number equal to a multiple of n.

The number of CCEs aggregated for transmission of a particular PDCCH is known as the 'aggregation level' and is determined by the eNodeB according to the channel conditions. For example, if the PDCCH is intended for a UE with a good downlink channel (e.g. close to the eNodeB), then one CCE is likely to be sufficient. However, for a UE with a poor channel (e.g. near the cell border) then eight CCEs may be required in order to achieve sufficient robustness. In addition, the power level of a PDCCH may be adjusted to match the channel conditions.

9.3.5.1 Formats for Downlink Control Information (DCI)

The required content of the control channel messages depends on the system deployment and UE configuration. For example, if the infrastructure does not support MIMO, or if a UE is configured in a transmission mode which does not involve MIMO, there is no need to signal the parameters that are only required for MIMO transmissions. In order to minimize the signalling overhead, it is therefore desirable that several different message formats are available, each containing the minimum payload required for a particular scenario. On the other hand, to avoid too much complexity in implementation and testing, it is desirable not to specify too many formats. The set of DCI message formats in Table 9.2 is specified in LTE; Format 2B was added in Release 9, and Formats 2C and 4 were added in Release 10. Additional formats may be defined in future.

Table 9.2: Supported DCI formats.

DCI format	Purpose	Applicable PDSCH transmission mode(s)
0	PUSCH grants	All
1	PDSCH assignments with a single codeword	1,2,7
1A	PDSCH assignments using a compact format	All
1B	PDSCH assignments for rank-1 transmission	6
1C	PDSCH assignments using a very compact format	n/a
1D	PDSCH assignments for multi-user MIMO	5
2	PDSCH assignments for closed-loop MIMO operation	4
2A	PDSCH assignments for open-loop MIMO operation	3
2B	PDSCH assignments for dual-layer beamforming	8
2C	PDSCH assignments for up to 8-layer spatial multiplexing	9
3	Transmit Power Control (TPC) commands for multiple users for PUCCH and PUSCH with 2-bit power adjustments	n/a
3A	Transmit Power Control (TPC) commands for multiple users for PUCCH and PUSCH with 1-bit power adjustments	n/a
4	PUSCH grants for up to 4-layer spatial multiplexing	All (if configured for PUSCH transmission mode 2)

The information content of the different DCI message formats is listed below for Frequency Division Duplex (FDD) operation. Some small differences exist for Time Division Duplex (TDD), and these are outlined afterwards.

Format 0. DCI Format 0 is used for the transmission of resource grants for the PUSCH. The following information is transmitted:

- Flag to differentiate between Format 0 and Format 1A;
- Resource assignment and frequency hopping flag;
- Modulation and Coding Scheme (MCS);
- New Data Indicator (NDI);
- HARQ information and Redundancy Version (RV);
- Power control command for scheduled PUSCH;
- Cyclic shift for uplink Demodulation RS;
- Request for transmission of an aperiodic CQI report (see Sections 10.2.1 and 28.3.2.3).

Format 1. DCI Format 1 is used for the transmission of resource assignments for single codeword PDSCH transmissions (transmission modes 1, 2 and 7 (see Section 9.2.2.1)). The following information is transmitted:

- Resource allocation type (see Section 9.3.5.4);
- RB assignment;
- MCS;
- HARQ information and RV;
- Power control command for Physical Uplink Control CHannel (PUCCH).

Format 1A. DCI Format 1A is used for compact signalling of resource assignments for single codeword PDSCH transmissions for any PDSCH transmission mode. It is also used to allocate a dedicated preamble signature to a UE to trigger contention-free random access (see Section 17.3.2); in this case the PDCCH message is known as a *PDCCH order*. The following information is transmitted:

- Flag to differentiate between Format 0 and Format 1A;
- Flag to indicate that the distributed mapping mode (see Section 9.2.2.1) is used for the PDSCH transmission (otherwise the allocation is a contiguous set of physical RBs);
- RB assignment;
- MCS;
- HARQ information and RV;
- Power control command for PUCCH.

Format 1B. DCI Format 1B is used for compact signalling of resource assignments for PDSCH transmissions using closed-loop precoding with rank-1 transmission (transmission mode 6). The information transmitted is the same as in Format 1A, but with the addition of an indicator of the precoding vector applied for the PDSCH transmission.

Format 1C. DCI Format 1C is used for very compact transmission of PDSCH assignments. When format 1C is used, the PDSCH transmission is constrained to using QPSK modulation. This is used, for example, for signalling paging messages and some broadcast system information messages (see Section 9.2.2.2), and for notifying UEs of a change of MBMS control information on the Multicast Control Channel (MCCH – see Section 13.6.3.2). The following information is transmitted:

- RB assignment;
- Coding scheme.

The RV is not signalled explicitly, but is deduced from the SFN (see [1, Section 5.3.1]).

Format 1D. DCI Format 1D is used for compact signalling of resource assignments for PDSCH transmissions using multi-user MIMO (transmission mode 5). The information transmitted is the same as in Format 1B, but, instead of one of the bits of the precoding vector indicators, there is a single bit to indicate whether a power offset is applied to the data symbols. This is needed to show whether the transmission power is shared between two UEs.

Format 2. DCI Format 2 is used for the transmission of resource assignments for PDSCH for closed-loop MIMO operation (transmission mode 4). The following information is transmitted:

- Resource allocation type (see Section 9.3.5.4);
- RB assignment;
- Power control command for PUCCH;
- HARQ information and RV for each codeword;
- MCS for each codeword;
- A flag to indicate if the mapping from transport blocks to codewords is reversed;
- Number of spatial layers;
- Precoding information and indication of whether one or two codewords are transmitted on the PDSCH.

Format 2A. DCI Format 2A is used for the transmission of resource assignments for PDSCH for open-loop MIMO operation (transmission mode 3). The information transmitted is the same as for Format 2, except that if the eNodeB has two transmit antenna ports, there is no precoding information, and, for four antenna ports, two bits are used to indicate the transmission rank.

Format 2B. DCI Format 2B is introduced in Release 9 and is used for the transmission of resource assignments for PDSCH for dual-layer beamforming (transmission mode 8). The information transmitted is similar to Format 2A, except that no precoding information is included and the bit in Format 2A for indicating reversal of the transport block to codeword mapping is replaced in Format 2B by a bit indicating the scrambling code applied to the UE-specific RSs for the corresponding PDSCH transmission (see Section 8.2.3).

Format 2C. DCI Format 2C is introduced in Release 10 and is used for the transmission of resource assignments for PDSCH for closed-loop single-user or multi-user MIMO operation with up to 8 layers (transmission mode 9). The information transmitted is similar to Format 2B; full details are given in Section 29.3.2.

Formats 3 and 3A. DCI Formats 3 and 3A are used for the transmission of power control commands for PUCCH and PUSCH, with 2-bit or 1-bit power adjustments respectively. These DCI formats contain individual power control commands for a group of UEs.

Format 4. DCI Format 4 is introduced in Release 10 and is used for the transmission of resource grants for the PUSCH when the UE is configured in PUSCH transmission mode 2 for uplink single-user MIMO. The information transmitted is similar to Format 0, with the addition of MCS and NDI information for a second transport block, and precoding information; full details are given in Section 29.4.

DCI Formats for TDD. In TDD operation, the DCI formats contain the same information as for FDD, but with some additions (see Section 23.4.3 for an explanation of the usage of these additions):

- Uplink index (in DCI Formats 0 and 4, uplink-downlink configuration 0 only);
- Downlink Assignment Index (DAI) (in DCI Formats 0, 1, 1A, 1B, 1D, 2, 2A 2B, 2C and 4, uplink-downlink configurations 1–6 only); see Section 23.4.3 for details of DAI usage.

DCI Format modifications in Release 10. In the case of aggregation of multiple carriers in Release 10, DCI Formats 0, 1, 1A, 1B, 1D, 2, 2A, 2B, 2C and 4 can be configured to include a carrier indicator for cross-carrier scheduling; this is explained in detail in Section 28.3.1.1. In DCI Formats 0 and 4, additional fields are included to request transmission of an aperiodic Sounding Reference Signal (SRS) (see Section 29.2.2) and to indicate whether the uplink PRB allocation is contiguous or multi-clustered (see Section 28.3.6.2 for details). In TDD operation, DCI Formats 2B and 2C may also be configured to include an additional field to request transmission of an aperiodic SRS.

9.3.5.2 PDCCH CRC Attachment

In order that the UE can identify whether it has received a PDCCH transmission correctly, error detection is provided by means of a 16-bit CRC appended to each PDCCH. Furthermore, it is necessary that the UE can identify which PDCCH(s) are intended for it. This could in theory be achieved by adding an identifier to the PDCCH payload; however, it turns out to be more efficient to scramble the CRC with the 'UE identity', which saves the additional payload but at the cost of a small increase in the probability of falsely detecting a PDCCH intended for another UE.

In addition, for UEs which support antenna selection for uplink transmissions (see Section 16.6), the requested antenna may be indicated using Format 0 by applying an antenna-specific mask to the CRC. This has the advantage that the same size of DCI message can be used, irrespective of whether antenna selection is used.

9.3.5.3 PDCCH Construction

In general, the number of bits required for resource assignment depends on the system bandwidth, and therefore the message sizes also vary with the system bandwidth. The numbers of payload bits for each DCI format (including information bits and CRC) are summarized in Table 9.3, for each of the supported values of system bandwidth. In addition, padding bits are added if necessary in the following cases:

- To ensure that Formats 0 and 1A are the same size, even in the case of different uplink and downlink bandwidths, in order to avoid additional complexity at the UE receiver;
- To ensure that Formats 3 and 3A are the same size as Formats 0 and 1A, likewise to avoid additional complexity at the UE receiver;
- To avoid potential ambiguity in identifying the correct PDCCH location as described in Section 9.3.5.5;
- To ensure that Format 1 has a different size from Formats 0/1A, so that these formats can be easily distinguished at the UE receiver;
- To ensure that Format 4 has a different size from Formats 1/2/2A/2B/2C, again so that this format can be easily distinguished.

In Release 10, some optional additional information bits may be configured in some of the DCI formats; these are not included in Table 9.3, but are explained below the table. Because their presence may affect the number of padding bits, any such additional bits do not necessarily increase the transmitted size of the DCI format by the same amount.

In order to provide robustness against transmission errors, the PDCCH information bits are coded as described in Section 10.3.3. The set of coded and rate-matched bits for each PDCCH are then scrambled with a cell-specific scrambling sequence; this reduces the possibility of confusion with PDCCH transmissions from neighbouring cells. The scrambled bits are mapped to blocks of four QPSK symbols (REGs). Interleaving is applied to these symbol blocks, to provide frequency diversity, followed by mapping to the available physical REs on the set of OFDM symbols indicated by the PCFICH. This mapping process excludes the REs reserved for RSs and the other control channels (PCFICH and PHICH).

The PDCCHs are transmitted on the same set of antenna ports as the PBCH, and transmit diversity is applied if more than one antenna port is used.

9.3.5.4 Resource Allocation

Conveying indications of physical layer resource allocation is one of the major functions of the PDCCHs. While the exact use of the PDCCHs depends on the algorithms implemented in the eNodeB, it is nevertheless possible to outline some general principles of typical operation.

In each subframe, PDCCHs indicate the frequency-domain resource allocations. As discussed in Section 9.2.2.1, resource allocations are normally localized, meaning that a Physical RB (PRB) in the first half of a subframe is paired with the PRB at the same frequency in the second half of the subframe. For simplicity, the explanation here is in terms of the first half subframe only.

The main design challenge for the signalling of frequency-domain resource allocations (in terms of a set of RBs) is to find a good compromise between flexibility and signalling overhead. The most flexible, and arguably the simplest, approach is to send each UE a

Table 9.3: DCI format payload sizes (in bits), without padding, for different FDD system bandwidths.

	Bandwidth (PRBs)					
	6	15	25	50	75	100
Format 0	35	37	39	41	42	43
Format 1	35	38	43	47	49	55
Format 1A	36	38	40	42	43	44
Format 1B/1D (2 transmit antenna ports)	38	40	42	44	45	46
Format 1C	24	26	28	29	30	31
Format 2 (2 transmit antenna ports)	47	50	55	59	61	67
Format 2A (2 transmit antenna ports)	44	47	52	56	58	64
Format 2B (2 or 4 transmit antenna ports)	44	47	52	56	58	64
Format 2C	46	49	54	58	60	66
Format 4 (2 UE transmit antennas)	46	47	50	52	53	54
Format 1B/1D (4 transmit antenna ports)	40	42	44	46	47	48
Format 2 (4 transmit antenna ports)	50	53	58	62	64	70
Format 2A (4 transmit antenna ports)	46	49	54	58	60	66
Format 4 (4 UE transmit antennas)	49	50	53	55	56	57

Note that for Release 10 UEs:

- DCI Format 0 is extended by 1 bit for a multi-cluster resource allocation flag and may be further extended by 1 or 2 bits for aperiodic CQI request (see Section 28.3.2.3) and aperiodic SRS request (see Section 29.2.2), depending on configuration.
- DCI Format 1A may be extended by 1 bit for aperiodic SRS request, depending on configuration.
- In TDD operation, DCI Formats 2B and 2C may be extended by 1 bit for aperiodic SRS request, depending on configuration.
- DCI Format 4 always includes 2 bits for requesting aperiodic SRS (see Section 29.2.2), and 1 bit for aperiodic CQI request, but may be extended by an additional 1 bit for aperiodic CQI request in case of carrier aggregation (see Section 28.3.2.3).
- DCI Formats 0, 1, 1A, 1B, 1D, 2, 2A, 2B, 2C and 4 can be configured to be extended by 3 bits for a carrier indicator field (see Section 28.3.1.1).

bitmap in which each bit indicates a particular PRB. This would work well for small system bandwidths, but for large system bandwidths (i.e. up to 110 PRBs) the bitmap would need 110 bits, which would be a prohibitive overhead – particularly for small packets, where the PDCCH message could be larger than the data packet! One possible solution would be to send a combined resource allocation message to all UEs, but this was rejected on the grounds of the high power needed to reach all UEs reliably, including those at the cell edges. The approaches adopted in LTE Releases 8 and 9 are listed in Table 9.4, and further details are given below.

Resource allocation Type 0. In resource allocations of Type 0, a bitmap indicates the Resource Block Groups (RBGs) which are allocated to the scheduled UE, where an RBG is a set of consecutive PRBs. The RBG size (P) is a function of the system bandwidth as shown in Table 9.5. The total number of RBGs (N_{RBG}) for a downlink system bandwidth of

Table 9.4: Methods for indicating Resource Block (RB) allocation.

Method	UL/DL	Description	Number of bits required (see text for definitions)
Direct bitmap	DL	The bitmap comprises 1 bit per RB. This method is the only one applicable when the bandwidth is less than 10 resource blocks.	$N_{\mathrm{RB}}^{\mathrm{DL}}$
Bitmap: 'Type 0'	DL	The bitmap addresses Resource Block Groups (RBGs), where the group size (2, 3 or 4) depends on the system bandwidth.	$\lceil N_{\mathrm{RB}}^{\mathrm{DL}}/P \rceil$
Bitmap: 'Type 1'	DL	The bitmap addresses individual RBs in a subset of RBGs. The number of subsets (2, 3, or 4) depends on the system bandwidth. The number of bits is arranged to be the same as for Type 0, so the same DCI format can carry either type of allocation.	$\lceil N_{\mathrm{RB}}^{\mathrm{DL}}/P \rceil$
Contiguous allocations: 'Type 2'	DL or UL	Any possible arrangement of contiguous RB allocations can be signalled in terms of a starting position and number of RBs.	$\lceil \log_2(N_{\mathrm{RB}}^{\mathrm{DL}}(N_{\mathrm{RB}}^{\mathrm{DL}}+1)) \rceil$ or $\lceil \log_2(N_{\mathrm{RB}}^{\mathrm{UL}}(N_{\mathrm{RB}}^{\mathrm{UL}}+1)) \rceil$
Distributed allocations	DL	In the downlink, a limited set of resource allocations can be signalled where the RBs are scattered across the frequency domain and shared between two UEs. The number of bits is arranged to be the same as for contiguous allocations Type 2, so the same DCI format can carry either type of allocation.	$\lceil \log_2(N_{\mathrm{RB}}^{\mathrm{DL}}(N_{\mathrm{RB}}^{\mathrm{DL}}+1)) \rceil$

$N_{\mathrm{RB}}^{\mathrm{DL}}$ PRBs is given by $N_{\mathrm{RBG}} = \lceil N_{\mathrm{RB}}^{\mathrm{DL}}/P \rceil$. An example for the case of $N_{\mathrm{RB}}^{\mathrm{DL}} = 25$, $N_{\mathrm{RBG}} = 13$ and $P = 2$ is shown in Figure 9.10, where each bit in the bitmap indicates a pair of PRBs (i.e. two PRBs which are adjacent in frequency).

Table 9.5: RBG size for Type 0 resource allocation.

Downlink bandwidth $N_{\mathrm{RB}}^{\mathrm{DL}}$	RBG size (P)
$0 \leq 10$	1
11–26	2
27–63	3
64–110	4

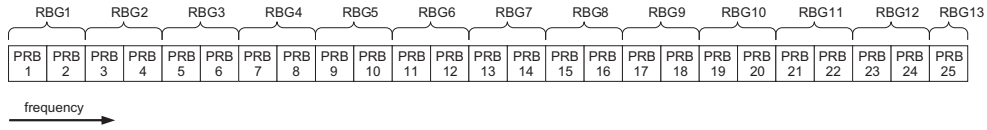


Figure 9.10: PRB addressed by a bitmap Type 0, each bit addressing a complete RBG.

Resource allocation Type 1. In resource allocations of Type 1, individual PRBs can be addressed (but only within a subset of the PRBs available). The bitmap used is slightly smaller than for Type 0, since some bits are used to indicate which subset of the RBG is addressed, and a shift in the position of the bitmap. The total number of bits (including these additional flags) is the same as for Type 0. An example for the case of $N_{\mathrm{RB}}^{\mathrm{DL}} = 25$, $N_{\mathrm{RBG}} = 11$ and $P = 2$ is shown in Figure 9.11.

The motivation for providing this method of resource allocation is flexibility in spreading the resources across the frequency domain to exploit frequency diversity.

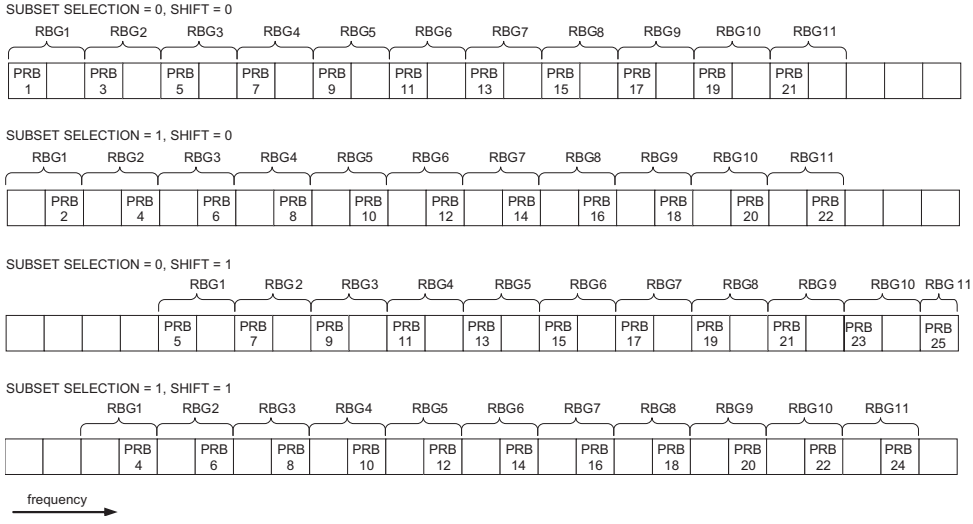


Figure 9.11: PRBs addressed by a bitmap Type 1, each bit addressing a subset of an RBG, depending on a subset selection and shift value.

Resource allocation Type 2. In resource allocations of Type 2, the resource allocation information indicates a contiguous set of PRBs, using either localized or distributed mapping (see Section 9.2.2.1) as indicated by a 1-bit flag in the resource allocation message. PRB allocations may vary from a single PRB up to a maximum number of PRBs spanning the system bandwidth. A Type 2 resource allocation field consists of a Resource Indication

Value (RIV) corresponding to a starting RB (RB_{START}) and a length in terms of contiguously-allocated RBs (L_{CRBs}). The resource indication value is defined by

$$\text{if} \quad (L_{CRBs} - 1) \leq \lfloor N_{RB}^{DL}/2 \rfloor \quad \text{then} \quad RIV = N_{RB}^{DL}(L_{CRBs} - 1) + RB_{START}$$
$$\text{else} \quad RIV = N_{RB}^{DL}(N_{RB}^{DL} - L_{CRBs} + 1) + (N_{RB}^{DL} - 1 - RB_{START})$$

An example of a method for reversing the mapping to derive the resource allocation from the RIV can be found in [2].

Resource allocation in Release 10. In addition to the above methods, Release 10 supports a non-contiguous resource allocation method for the uplink, allowing two separate contiguous sets of PRBs to be assigned, as explained in detail in Section 28.3.6.2.

9.3.5.5 PDCCH Transmission and Blind Decoding

The previous discussion has covered the structure and possible contents of an individual PDCCH message, and transmission by an eNodeB of multiple PDCCHs in a subframe. This section addresses the question of how these transmissions are organized so that a UE can locate the PDCCHs intended for it, while at the same time making efficient use of the resources allocated for PDCCH transmission.

A simple approach, at least for the eNodeB, would be to allow the eNodeB to place any PDCCH anywhere in the PDCCH resources (or CCEs) indicated by the PCFICH. In this case, the UE would need to check all possible PDCCH locations, PDCCH formats and DCI formats, and act on the messages with correct CRCs (taking into account that the CRC is scrambled with a UE identity). Carrying out such a 'blind decoding' of all the possible combinations would require the UE to make many PDCCH decoding attempts in every subframe. For small system bandwidths, the computational load would be reasonable, but for large system bandwidths, with a large number of possible PDCCH locations, it would become a significant burden, leading to excessive power consumption in the UE receiver. For example, blind decoding of 100 possible CCE locations for PDCCH Format 0 would be equivalent to continuously receiving a data rate of around 4 Mbps.

The alternative approach adopted for LTE is to define for each UE a limited set of CCE locations where a PDCCH may be placed. Such a constraint may lead to some limitations as to which UEs can be sent PDCCHs within the same subframe, which thus restricts the UEs to which the eNodeB can grant resources. Therefore it is important for good system performance that the set of possible PDCCH locations available for each UE is not too small.

The set of CCE locations in which the UE may find its PDCCHs can be considered as a 'search space'. In LTE, the search space is a different size for each PDCCH format. Moreover, separate *UE-specific* and *common* search spaces are defined; a UE-specific search space is configured for each UE individually, whereas all UEs are aware of the extent of the common search space. Note that the UE-specific and common search spaces may overlap for a given UE. The sizes of the common and UE-specific search spaces in Releases 8 and 9 are listed in Table 9.6.

With such small search spaces, it is quite possible in a given subframe that the eNodeB cannot find CCE resources to send PDCCHs to all the UEs that it would like to, because having assigned some CCE locations, the remaining ones are not in the search space of a

Table 9.6: Search spaces for PDCCH formats in Releases 8 and 9.

PDCCH format	Number of CCEs (n)	Number of candidates in common search space	Number of candidates in UE-specific search space
0	1	—	6
1	2	—	6
2	4	4	2
3	8	2	2

Note: The search space sizes in Release 10 are discussed in Section 28.3.1.1.

particular UE. To minimize the possibility of such blocking persisting into the next subframe, a UE-specific hopping sequence (derived from the UE identity) is applied to the starting positions of the UE-specific search spaces from subframe to subframe.

In order to keep under control the computational load arising from the total number of blind decoding attempts, the UE is not required to search for all the defined DCI formats simultaneously. Typically, in the UE-specific search space, the UE will always search for Formats 0 and 1A, which are the same size and are distinguished by a flag in the message. In addition, a UE may be required to receive a further format (i.e. 1, 1B, 1D, 2, 2A or 2B, depending on the PDSCH transmission mode configured by the eNodeB).

In the common search space, the UE will typically search for Formats 1A and 1C. In addition, the UE may be configured to search for Formats 3 or 3A, which have the same size as Formats 0 and 1A, and may be distinguished by having the CRC scrambled by a different (common) identity, rather than a UE-specific one.

Considering the above, a Release 8/9 UE is required to carry out a maximum of 44 blind decodings in any subframe (12 in the common search space and 32 in the UE-specific search space). This does not include checking the same message with different CRC values, which requires only a small additional computational complexity. The number of blind decodings required for a Release 10 UE is discussed in Section 28.3.1.1.

Finally, the PDCCH structure is also adapted to avoid cases where a PDCCH CRC 'pass' might occur for multiple positions in the configured search-spaces due to repetition in the channel coding (for example, if a PDCCH was mapped to a high number of CCEs with a low code rate, then the CRC could pass for an overlapping smaller set of CCEs as well, if the channel coding repetition was thus aligned). Such cases are avoided by adding a padding bit to any PDCCH messages having a size which could result in this problem occurring.

9.3.6 PDCCH Scheduling Process

To summarize the arrangement of the PDCCH transmissions in a given subframe, a typical sequence of steps carried out by the eNodeB is depicted in Figure 9.12.

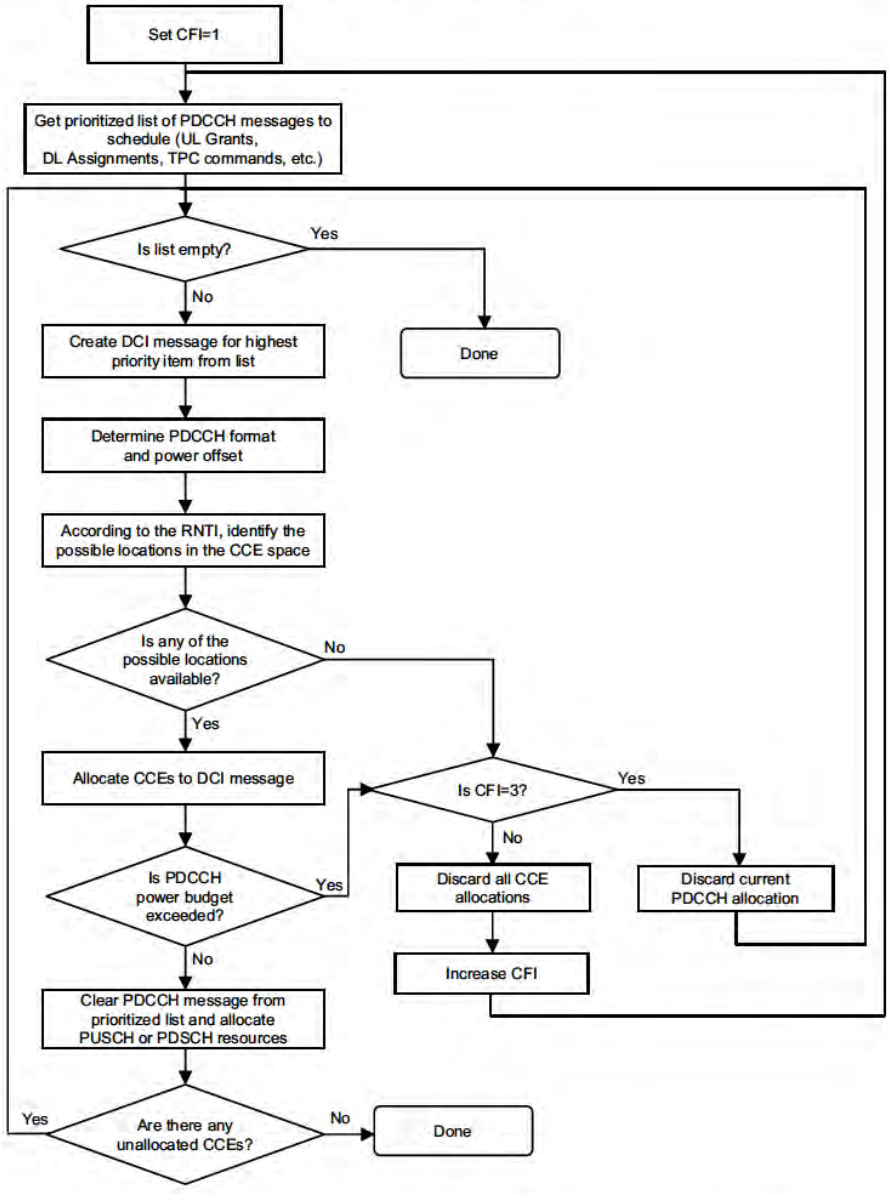


Figure 9.12: A typical sequence of PDCCH scheduling operations in a subframe.

References[12]

[1] 3GPP Technical Specification 36.321, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification', www.3gpp.org.

[2] NEC, 'R1-072119: DL Unicast Resource Allocation Signalling', www.3gpp.org, 3GPP TSG RAN WG1, meeting 49, Kobe, Japan, May 2007.

[12] All web sites confirmed 1st March 2011.

10

Link Adaptation and Channel Coding

Brian Classon, Ajit Nimbalker, Stefania Sesia and Issam Toufik

10.1 Introduction

The principle of link adaptation is fundamental to the design of a radio interface which is efficient for packet-switched data traffic. Unlike the early versions of UMTS (Universal Mobile Telecommunication System), which used fast closed-loop power control to support circuit-switched services with a roughly constant data rate, link adaptation in HSPA (High Speed Packet Access) and LTE adjusts the transmitted information data rate (modulation scheme and channel coding rate) dynamically to match the prevailing radio channel capacity for each user. Link adaptation is therefore very closely related to the design of the channel coding scheme used for forward error correction.

For the downlink data transmissions in LTE, the eNodeB typically selects the modulation scheme and code rate depending on a prediction of the downlink channel conditions. An important input to this selection process is the Channel Quality Indicator (CQI) feedback transmitted by the User Equipment (UE) in the uplink. CQI feedback is an indication of the data rate which can be supported by the channel, taking into account the Signal-to-Interference-plus-Noise Ratio (SINR) and the characteristics of the UE's receiver. Section 10.2 explains the principles of link adaptation as applied in LTE; it also shows how the eNodeB can select different CQI feedback modes to trade off the improved downlink link adaptation enabled by CQI against the uplink overhead caused by the CQI itself.

The LTE specifications are designed to provide the signalling necessary for interoperability between the eNodeB and the UEs so that the eNodeB can optimize the link adaptation,

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

but the exact methods used by the eNodeB to exploit the information that is available are left to the manufacturer's choice of implementation.

In general, in response to the CQI feedback the eNodeB can select between QPSK, 16QAM and 64QAM[1] schemes and a wide range of code rates. As discussed further in Section 10.2.1, the optimal switching points between the different combinations of modulation order and code rate depend on a number of factors, including the required quality of service and cell throughput.

The channel coding scheme for forward error correction, on which the code rate adaptation is based, was the subject of extensive study during the standardization of LTE. The chapter therefore continues with a review of the key theoretical aspects of the types of channel coding studied for LTE: convolutional codes, turbo codes with iterative decoding, and a brief introduction of Low-Density Parity Check (LDPC) codes. The theory of channel coding has seen intense activity in recent decades, especially since the discovery of turbo codes offering near-Shannon limit performance, and the development of iterative processing techniques in general. 3GPP was an early adopter of these advanced channel coding techniques, with the turbo code being standardized in the first version of the UMTS as early as 1999. Later releases of UMTS (in HSPA) added more advanced channel coding features with the introduction of link adaptation, including Hybrid Automatic Repeat reQuest (HARQ), a combination of ARQ and channel coding which provides more robustness against fading; these schemes include incremental redundancy, whereby the code rate is progressively reduced by transmitting additional parity information with each retransmission. However, the underlying turbo code from the first version of UMTS remained untouched. Meanwhile, the academic and research communities were generating new insights into code design, iterative decoding and the implementation of decoders. Section 10.3.2 explains how these developments impacted the design of the channel coding for LTE, and in particular the decision to enhance the turbo code from UMTS by means of a new contention-free interleaver, rather than to adopt a new LDPC code.

For the LTE uplink transmissions, the link adaptation process is similar to that for the downlink, with the selection of modulation and coding schemes also being under the control of the eNodeB. An identical channel coding structure is used for the uplink, while the modulation scheme may be selected between QPSK, 16QAM and, for the highest category of UE, also 64QAM. The main difference from the downlink is that instead of basing the link adaptation on CQI feedback, the eNodeB can directly make its own estimate of the supportable uplink data rate by channel sounding, for example using the Sounding Reference Signals (SRSs) which are described separately in Section 15.6.

A final important aspect of link adaptation is its use in conjunction with multi-user scheduling in time and frequency, which enables the radio transmission resources to be shared efficiently between users as the channel capacity to individual users varies. The CQI can therefore be used not only to adapt the modulation and coding rate to the channel conditions, but also for the optimization of time/frequency selective scheduling and for inter-cell interference management as discussed in Chapter 12.

[1] Quadrature Phase Shift Keying (QPSK) and Quadrature Amplitude Modulation (QAM).

10.2 Link Adaptation and CQI Feedback

In cellular communication systems, the quality of the signal received by a UE depends on the channel quality from the serving cell, the level of interference from other cells, and the noise level. To optimize system capacity and coverage for a given transmission power, the transmitter should try to match the information data rate for each user to the variations in received signal quality (see, for example, [1, 2] and references therein). This is commonly referred to as link adaptation and is typically based on Adaptive Modulation and Coding (AMC).

The degrees of freedom for the AMC consist of the modulation and coding schemes:

- **Modulation Scheme.** Low-order modulation (i.e. few data bits per modulated symbol, e.g. QPSK) is more robust and can tolerate higher levels of interference but provides a lower transmission bit rate. High-order modulation (i.e. more bits per modulated symbol, e.g. 64QAM) offers a higher bit rate but is more prone to errors due to its higher sensitivity to interference, noise and channel estimation errors; it is therefore useful only when the SINR is sufficiently high.

- **Code rate.** For a given modulation, the code rate can be chosen depending on the radio link conditions: a lower code rate can be used in poor channel conditions and a higher code rate in the case of high SINR. The adaptation of the code rate is achieved by applying puncturing (to increase the code rate) or repetition (to reduce the code rate) to the output of a mother code, as explained in Section 10.3.2.4.

A key issue in the design of the AMC scheme for LTE was whether all Resource Blocks (RBs) allocated to one user in a subframe should use the same Modulation and Coding Scheme (MCS) (see, for example, [3–6]) or whether the MCS should be frequency-dependent within each subframe. It was shown that, in general, only a small throughput improvement arises from a frequency-dependent MCS compared to an RB-common MCS in the absence of transmission power control, and therefore the additional control signalling overhead associated with frequency-dependent MCS is not justified. Therefore in LTE the modulation and channel coding rates are constant over the allocated frequency resources for a given user, and time-domain channel-dependent scheduling and AMC are supported instead. In addition, when multiple transport blocks are transmitted to one user in a given subframe using multistream Multiple-Input Multiple-Output (MIMO) (as discussed in Chapter 11), each transport block can use an independent MCS.

In LTE, the UE can be configured to report CQIs to assist the eNodeB in selecting an appropriate MCS to use for the downlink transmissions. The CQI reports are derived from the downlink received signal quality, typically based on measurements of the downlink reference signals (see Section 8.2). It is important to note that, like HSDPA, the reported CQI is not a direct indication of SINR. Instead, the UE reports the highest MCS that it can decode with a BLER (BLock Error Rate, computed on the transport blocks) probability not exceeding 10%. Thus the information received by the eNodeB takes into account the characteristics of the UE's receiver, and not just the prevailing radio channel quality. Hence a UE that is designed with advanced signal processing algorithms (for example, using interference cancellation techniques) can report a higher channel quality and, depending on the characteristics of the eNodeB's scheduler, can receive a higher data rate.

A simple method by which a UE can choose an appropriate CQI value could be based on a set of BLER thresholds, as illustrated in Figure 10.1. The UE would report the CQI value corresponding to the MCS that ensures BLER $\leq 10^{-1}$ based on the measured received signal quality.

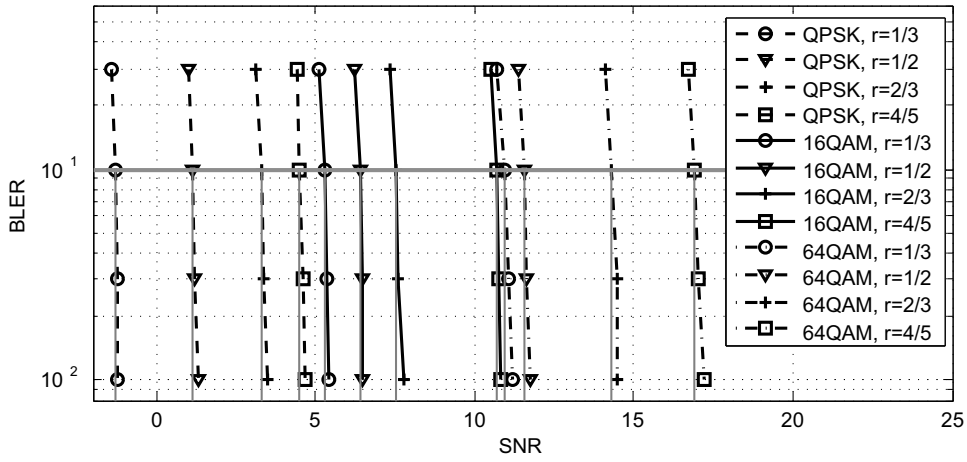


Figure 10.1: Typical BLER versus Signal-to-Noise Ratio (SNR) for different MCSs. From left to right, the curves in this example correspond to QPSK, 16QAM and 64QAM, rates 1/3, 1/2, 2/3 and 4/5.

The list of modulation schemes and code rates which can be signalled by means of a CQI value is shown in Table 10.1.

AMC can exploit the UE feedback by assuming that the channel fading is sufficiently slow. This requires the channel coherence time to be at least as long as the time between the UE's measurement of the downlink reference signals and the subframe containing the correspondingly adapted downlink transmission on the Physical Downlink Shared CHannel (PDSCH). This time is typically 7–8 ms (equivalent to a UE speed of ∼16 km/h at 1.9 GHz).

However, a trade-off exists between the amount of CQI information reported by the UEs and the accuracy with which the AMC can match the prevailing conditions. Frequent reporting of the CQI in the time domain allows better matching to the channel and interference variations, while fine resolution in the frequency domain allows better exploitation of frequency-domain scheduling. However, both lead to increased feedback overhead in the uplink. Therefore, the eNodeB can configure both the time-domain update rate and the frequency-domain resolution of the CQI, as discussed in the following section.

10.2.1 CQI Feedback in LTE

The periodicity and frequency resolution to be used by a UE to report CQI are both controlled by the eNodeB. In the time domain, both periodic and aperiodic CQI reporting are supported.

Table 10.1: CQI values. Reproduced by permission of © 3GPP.

CQI index	Modulation	Approximate code rate	Efficiency (information bits per symbol)
0	'Out-of-range'	—	—
1	QPSK	0.076	0.1523
2	QPSK	0.12	0.2344
3	QPSK	0.19	0.3770
4	QPSK	0.3	0.6016
5	QPSK	0.44	0.8770
6	QPSK	0.59	1.1758
7	16QAM	0.37	1.4766
8	16QAM	0.48	1.9141
9	16QAM	0.6	2.4063
10	64QAM	0.45	2.7305
11	64QAM	0.55	3.3223
12	64QAM	0.65	3.9023
13	64QAM	0.75	4.5234
14	64QAM	0.85	5.1152
15	64QAM	0.93	5.5547

The Physical Uplink Control CHannel (PUCCH, see Section 16.3.1) is used for periodic CQI reporting only while the Physical Uplink Shared CHannel (PUSCH, see Section 16.2) is used for aperiodic reporting of the CQI, whereby the eNodeB specifically instructs the UE to send an individual CQI report embedded into a resource which is scheduled for uplink data transmission.

The frequency granularity of the CQI reporting is determined by defining a number of sub-bands (N), each comprised of k contiguous Physical Resource Blocks (PRBs).[2] The value of k depends on the type of CQI report considered and is a function of the system bandwidth. In each case, the number of sub-bands spans the whole system bandwidth and is given by $N = \lceil N_{\mathrm{RB}}^{\mathrm{DL}}/k \rceil$, where $N_{\mathrm{RB}}^{\mathrm{DL}}$ is the number of RBs across the system bandwidth. The CQI reporting modes can be Wideband CQI, eNodeB-configured sub-band feedback, or UE-selected sub-band feedback. These are explained in detail in the following sections. In addition, in the case of multiple transmit antennas at the eNodeB, CQI value(s) may be reported for a second codeword.

For some downlink transmission modes, additional feedback signalling consisting of Precoding Matrix Indicators (PMI) and Rank Indications (RI) is also transmitted by the UE, as explained in Section 11.2.2.4.

10.2.1.1 Aperiodic CQI Reporting

Aperiodic CQI reporting on the PUSCH is scheduled by the eNodeB by setting a CQI request bit in an uplink resource grant sent on the Physical Downlink Control CHannel (PDCCH).

[2]Note that the last sub-band may have less than k contiguous PRBs.

The type of CQI report is configured by the eNodeB by RRC signalling. Table 10.2 summarizes the relationship between the configured downlink transmission mode (see Section 9.2.2.1) and the possible CQI reporting type.

Table 10.2: Aperiodic CQI feedback types on PUSCH for each PDSCH transmission mode.

		CQI type		
PDSCH transmission mode (See Section 9.2.2.1)	PDSCH transmission scheme assumed by UE for deriving CQI	Wideband only	eNodeB-configured sub-bands	UE-selected sub-bands
Mode 1	Single antenna port		X	X
Mode 2	Transmit diversity		X	X
Mode 3	Transmit diversity if RI=1, otherwise large-delay CDD[a]		X	X
Mode 4	Closed-loop spatial multiplexing	X	X	X
Mode 5	Multi-user MIMO		X	
Mode 6	Closed-loop spatial multiplexing using a single transmission layer	X	X	X
Mode 7	Single antenna port if one PBCH antenna port, otherwise transmit diversity		X	X
Mode 8[b] with PMI/RI configured	Closed-loop spatial multiplexing	X	X	X
Mode 8[b] without PMI/RI configured	Single antenna port if one PBCH antenna port, otherwise transmit diversity		X	X
Mode 9[c] with PMI/RI configured and >1 CSI-RS port[d]	Closed-loop spatial multiplexing	X	X	X
Mode 9[c] otherwise	Single antenna port if one PBCH antenna port, otherwise transmit diversity		X	X

[a] Cyclic Delay Diversity. [b] Introduced in Release 9. [c] Introduced in Release 10. [d] See Section 29.1.2.

The CQI reporting types are as follows:

- **Wideband feedback.** The UE reports one wideband CQI value for the whole system bandwidth.

- **eNodeB-configured sub-band feedback.** The UE reports a wideband CQI value for the whole system bandwidth. In addition, the UE reports a CQI value for each sub-band, calculated assuming transmission only in the relevant sub-band. Sub-band CQI reports are encoded differentially with respect to the wideband CQI using 2-bits as follows:

  Sub-band differential CQI offset = Sub-band CQI index − Wideband CQI index

  Possible sub-band differential CQI offsets are $\{\leq -1, 0, +1, \geq +2\}$. The sub-band size k is a function of system bandwidth as summarized in Table 10.3.

Table 10.3: Sub-band size (k) versus system bandwidth for eNodeB-configured aperiodic CQI reports. Reproduced by permission of © 3GPP.

System bandwidth (RBs)	Sub-band size (k RBs)
6–7	(Wideband CQI only)
8–10	4
11–26	4
27–63	6
64–110	8

- **UE-selected sub-band feedback.** The UE selects a set of M preferred sub-bands of size k (where k and M are given in Table 10.4 for each system bandwidth range) within the whole system bandwidth. The UE reports one wideband CQI value and one CQI value reflecting the average quality of the M selected sub-bands. The UE also reports the positions of the M selected sub-bands using a combinatorial index r defined as

$$r = \sum_{i=0}^{M-1} \left\langle {N - s_i \atop M - i} \right\rangle$$

  where the set $\{s_i\}_{i=0}^{M-1}$, $1 \leq s_i \leq N$, $s_i < s_{i+1}$ contains the M sorted sub-band indices and

$$\left\langle {x \atop y} \right\rangle = \begin{cases} \binom{x}{y} & \text{if } x \geq y \\ 0 & \text{if } x < y \end{cases}$$

  is the extended binomial coefficient, resulting in a unique label $r \in \{0, \ldots, \binom{N}{M} - 1\}$. Some possible algorithms for deriving the combinatorial index r in the UE and extracting the information from it in the eNodeB can be found in [7, 8].

  The CQI value for the M selected sub-bands for each codeword is encoded differentially using two bits relative to its respective wideband CQI as defined by

  Differential CQI

  = Index for average of M preferred sub-bands − Wideband CQI index

  Possible differential CQI values are $\{\leq +1, +2, +3, \geq +4\}$.

Table 10.4: Sub-band size k and number of preferred sub-bands (M) versus downlink system bandwidth for aperiodic CQI reports for UE-selected sub-band feedback. Reproduced by permission of © 3GPP.

System bandwidth (RBs)	Sub-band size (k RBs)	Number of preferred sub-bands (M)
6–7	(Wideband CQI only)	(Wideband CQI only)
8–10	2	1
11–26	2	3
27–63	3	5
64–110	4	6

10.2.1.2 Periodic CQI Reporting

If the eNodeB wishes to receive periodic reporting of the CQI, the UE will transmit the reports using the PUCCH.[3]

Only wideband and UE-selected sub-band feedback is possible for periodic CQI reporting, for all downlink (PDSCH) transmission modes. As with aperiodic CQI reporting, the type of periodic reporting is configured by the eNodeB by RRC signalling. For wideband periodic CQI reporting, the period can be configured to[4]):

$$\{2,\ 5,\ 10,\ 16,\ 20,\ 32,\ 40,\ 64,\ 80,\ 128,\ 160\} \text{ ms, or Off.}$$

While the wideband feedback mode is similar to that sent via the PUSCH, the 'UE-selected sub-band' CQI using PUCCH is different. In this case, the total number of sub-bands N is divided into J fractions called *bandwidth parts*. The value of J depends on the system bandwidth as summarized in Table 10.5. In this case, one CQI value is computed and reported for a single selected sub-band from each bandwidth part, along with the corresponding sub-band index.

Table 10.5: Periodic CQI reporting with UE-selected sub-bands: sub-band size (k) and bandwidth parts (J) versus downlink system bandwidth. Reproduced by permission of © 3GPP.

System bandwidth (RBs)	Sub-band size (k RBs)	Number of bandwidth parts (J)
6–7	(Wideband CQI only)	1
8–10	4	1
11–26	4	2
27–63	6	3
64–110	8	4

[3] If PUSCH transmission resources are scheduled for the UE in one of the periodic subframes, the periodic CQI report is sent on the PUSCH instead, as explained in Section 16.4 and Figure 16.15.

[4] These values apply to FDD operation; for TDD, see [9, Section 7.2.2]

10.2.1.3 CQI Reporting for Spatial Multiplexing

If the UE is configured in PDSCH transmission modes 3, 4, 8 or 9,[5] the eNodeB may use spatial multiplexing to transmit two codewords simultaneously to the UE with independently selected MCSs. To support this, the following behaviour is defined for the UE's CQI reports in these modes:

- If the UE is not configured to send RI feedback, or if the reported RI is equal to 1, or in any case in transmission mode 3,[6] the UE feeds back only one CQI report corresponding to a single codeword.
- If RI feedback is configured and the reported RI is greater than 1 in transmission modes 4 or 8:
  - For aperiodic CQI reporting, each CQI report (whether wideband or sub-band) comprises two independent CQI reports for the two codewords.
  - For periodic CQI reporting, one CQI report is fed back for the first codeword, and a second three-bit differential CQI report is fed back for the second codeword (for both wideband and sub-band reporting). The differential CQI report for the second codeword can take the following values relative to the CQI report for the first codeword: $\leq -4, -3, -2, -1, 0, +1, +2, \geq +3$.

10.3 Channel Coding

Channel coding, and in particular the channel decoder, has retained its reputation for being the dominant source of complexity in the implementation of wireless communications, in spite of the relatively recent prevalence of advanced antenna techniques with their associated complexity.

Section 10.3.1 introduces the theory behind the families of channel codes of relevance to LTE. This is followed in Sections 10.3.2 and 10.3.3 by an explanation of the practical design and implementation of the channel codes used in LTE for data and control signalling respectively.

10.3.1 Theoretical Aspects of Channel Coding

This section first explains convolutional codes, as not only do they remain relevant for small data blocks but also an understanding of them is a prerequisite for understanding turbo codes. The turbo-coding principle and the Soft-Input Soft-Output (SISO) decoding algorithms are then discussed. The section concludes with a brief introduction to LDPC codes.

10.3.1.1 From Convolutional Codes to Turbo Codes

A convolutional encoder $C(k, n, m)$ is composed of a shift register with m stages. At each time instant, k information bits enter the shift register and k bits in the last position of

[5] Mode 8 from Release 9 onwards; mode 9 from Release 10 onwards.

[6] If RI feedback is configured in transmission mode 3 and the RI fed back is greater than 1, although only one CQI corresponding to the first codeword is fed back, its value is adapted on the assumption that a second codeword will also be transmitted.

the shift register are dropped. The set of n output bits is a linear combination of the content of the shift register. The *rate* of the code is defined as $R_c = k/n$. Figure 10.2 shows the convolutional encoder used in LTE [10] with $m = 6,\ n = 3,\ k = 1$ and rate $R_c = 1/3$. The linear combinations are defined via n generator sequences $\mathbf{G} = [\mathbf{g}_0, \ldots, \mathbf{g}_{n-1}]$ where $\mathbf{g}_\ell = [g_{\ell,0},\ g_{\ell,1}, \ldots, g_{\ell,m}]$. The generator sequences used in Figure 10.2 are

$$\mathbf{g}_0 = [1\ 0\ 1\ 1\ 0\ 1\ 1] = [133](\text{oct}),$$

$$\mathbf{g}_1 = [1\ 1\ 1\ 1\ 0\ 0\ 1] = [171](\text{oct}),$$

$$\mathbf{g}_2 = [1\ 1\ 1\ 0\ 1\ 0\ 1] = [165](\text{oct}).$$

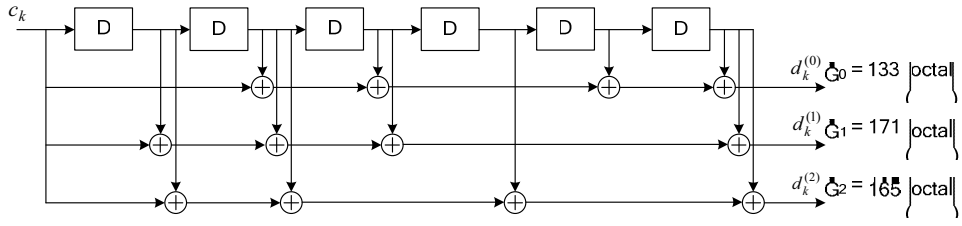


Figure 10.2: Rate $1/3$ convolutional encoder used in LTE with $m = 6,\ n = 3,\ k = 1$ [10]. Reproduced by permission of © 3GPP.

A convolutional encoder can be described by a trellis diagram [11], which is a representation of a finite state machine including the time dimension.

Consider an input block with L bits encoded with a rate $1/n$ (i.e. $k = 1$) convolutional encoder, resulting in a codeword of length $(L + m) \times n$ bits, including m trellis termination bits (or *tail bits*) which are inserted at the end of the information block to drive the shift register contents back to all zeros at the end of the encoding process. Note that using tail bits is just one possible way of terminating an input sequence. Other trellis termination methods include simple truncation (i.e. no tail bits appended) and so-called *tail-biting* [12]. In the tail-biting approach, the initial and final states of the convolutional encoder are required to be identical. Usually, tail-biting for feed-forward convolutional encoders is achieved by initializing the shift register contents with the last m information bits in the input block. Tail-biting encoding facilitates uniform protection of the information bits and suffers no rate-loss owing to the tail bits. Tail-biting convolutional codes can be decoded using, for example, the *Circular Viterbi Algorithm* (CVA) [13, 14].

Let the received sequence $\mathbf{y}$ be expressed as

$$\mathbf{y} = \sqrt{E_b}\mathbf{x} + \mathbf{n} \tag{10.1}$$

where $\mathbf{n} = [n_0,\ n_1, \ldots, n_\ell, \ldots, n_{(L+m)\times(n-1)}]$ and $n_\ell \sim N(0, N_0)$ is the Additive White Gaussian Noise (AWGN) and E_b is the energy per bit. The transmitted codeword is $\mathbf{x} = [\mathbf{x}_0,\ \mathbf{x}_1, \ldots, \mathbf{x}_\ell, \ldots, \mathbf{x}_{L+m-1}]$ where $\mathbf{x}_\ell$ is the convolutional code output sequence at time

instant ℓ for the input information bit i_ℓ, given by $\mathbf{x}_\ell = [x_{\ell,0}, \ldots, x_{\ell,n-1}]$. Equivalently, $\mathbf{y} = [\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_\ell, \ldots, \mathbf{y}_{L+m-1}]$ where $\mathbf{y}_\ell = [y_{\ell,0}, \ldots, y_{\ell,n-1}]$ is the noisy received version of $\mathbf{x}$. $(L + m)$ is the total trellis length.

10.3.1.2 Soft-Input Soft-Output (SISO) Decoders

In order to define the performance of a communication system, the codeword error probability or bit error probability can be considered. The minimization of the bit error probability is in general more complicated and requires the maximization of the a posteriori bit probability (MAP symbol-by-symbol). The minimization of the codeword/sequence error probability is in general easier and is equivalent to the maximization of A Posteriori Probability (APP) for each codeword; this is expressed by the MAP sequence detection rule, whereby the estimate $\hat{\mathbf{x}}$ of the transmitted codeword is given by

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, P(\mathbf{x} \mid \mathbf{y}) \tag{10.2}$$

When all codewords are equiprobable, the MAP criterion is equivalent to the Maximum Likelihood (ML) criterion which selects the codeword that maximizes the probability of the received sequence $\mathbf{y}$ conditioned on the estimated transmitted sequence $\mathbf{x}$, i.e.

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}}\, P(\mathbf{y} \mid \mathbf{x}) \tag{10.3}$$

Maximizing Equation (10.3) is equivalent to maximizing the logarithm of $P(\mathbf{y} \mid \mathbf{x})$, as $\log(\cdot)$ is a monotonically increasing function. This leads to simplified processing.[7] The log-likelihood function for a memoryless channel can be written as

$$\log P(\mathbf{y} \mid \mathbf{x}) = \sum_{i=0}^{L+m-1} \sum_{j=0}^{n-1} \log P(y_{i,j} \mid x_{i,j}) \tag{10.4}$$

For an AWGN channel, the conditional probability in Equation (10.4) is $P(y_{i,j} \mid x_{i,j}) \sim N(\sqrt{E_b} x_{i,j}, N_0)$, hence

$$\log P(\mathbf{y} \mid \mathbf{x}) \propto \|\mathbf{y}_i - \sqrt{E_b}\mathbf{x}_i\|^2 \tag{10.5}$$

The maximization of the metric in Equation (10.5) yields a codeword that is closest to the received sequence in terms of the Euclidean distance [15]. This maximization can be performed in an efficient manner by operating on the trellis.

As an example, Figure 10.3 shows a simple convolutional code with generator polynomials $\mathbf{g}_0 = [1\ 0\ 1]$ and $\mathbf{g}_1 = [1\ 1\ 1]$ and Figure 10.4 represents the corresponding trellis diagram. Each edge in the trellis corresponds to a transition from a state s to a state s', which can be obtained for a particular input information bit. In Figure 10.4, the edges are parametrized with the notation $i_\ell / x_{\ell,0}\ x_{\ell,1}$, i.e. the input/output of the convolutional encoder. The shift registers of the convolutional code are initialized to the all-zero state.

In Figure 10.4, $M(\mathbf{y}_i \mid \mathbf{x}_i) = \sum_{j=0}^{n-1} \log P(y_{i,j} \mid x_{i,j})$ denotes the branch metric at the i^{th} trellis step (i.e. the cost of choosing a branch at trellis step i), given by Equation (10.5). The Viterbi Algorithm (VA) selects the best path through the trellis by computing at each step the best

[7] The processing is simplified because the multiplication operation can be transformed to the simpler addition operation in the logarithmic domain.

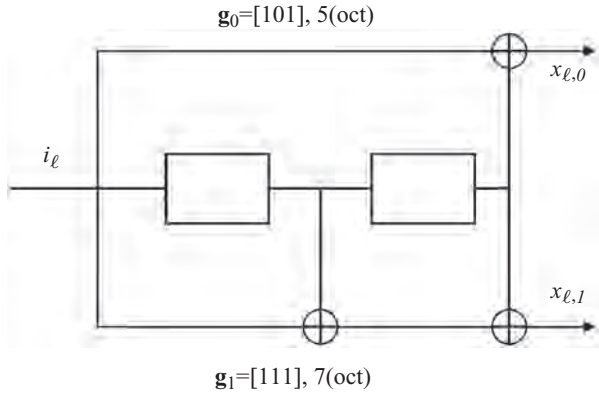


Figure 10.3: Rate $\frac{1}{2}$ convolutional encoder with $m = 2$, $m = 2$, $k = 1$, corresponding to generator polynomials $\mathbf{g}_0 = [1\ 0\ 1]$ and $\mathbf{g}_1 = [1\ 1\ 1]$.

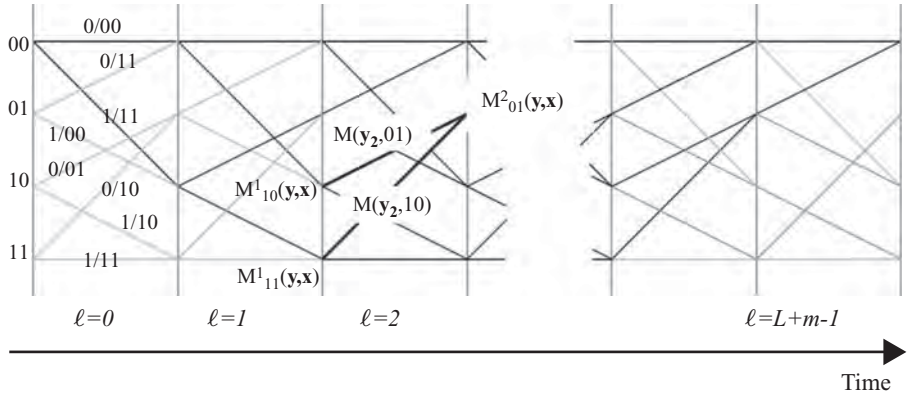


Figure 10.4: Trellis corresponding to convolutional code with generator polynomials $\mathbf{g}_0 = [1\ 0\ 1]$ and $\mathbf{g}_1 = [1\ 1\ 1]$.

partial metric (the accumulated cost of following a specific path until the ℓ^{th} transition) and selecting at the end the best total metric [16, 17]. It then traces back the selected path in the trellis to provide the estimated input sequence.

Although the original VA outputs a hard-decision estimate of the input sequence, the VA can be modified to output soft information[8] along with the hard-decision estimate of the input sequence [18]. The reliability of coded or information bits can be obtained via the computation of the a posteriori probability.

Let i_ℓ be the information bit which enters the shift register of the convolutional code at time ℓ. Assuming BPSK (Binary Phase Shift Keying) modulation ($0 \rightarrow +1$, and $1 \rightarrow -1$), the Log Likelihood Ratio (LLR) of an information symbol (or bit) i_ℓ is

[8] A soft decision gives additional information about the reliability of the decision [15].

$$\Lambda(i_\ell) = \Lambda(x_{\ell,0}) = \log \frac{\text{APP}(i_\ell = 1)}{\text{APP}(i_\ell = -1)} = \log \frac{P(i_\ell = 1 \mid \mathbf{y})}{P(i_\ell = -1 \mid \mathbf{y})} \tag{10.6}$$

The decoder can make a decision by comparing $\Lambda(i_\ell)$ to zero. Thus, the sign of the LLR gives an estimate of the information bit i_ℓ (LLR $\geq 0 \rightarrow 0$, and LLR $< 0 \rightarrow 1$), and the magnitude indicates the reliability of the estimate of the bit.

The a posteriori LLR $\Lambda(i_\ell)$ can be computed via the BCJR algorithm (named after its inventors, Bahl, Cocke, Jelinek and Raviv [19]). The BCJR algorithm is a SISO decoding algorithm that uses two Viterbi-like recursions going forwards and backwards in the trellis to compute Equation (10.6) efficiently. For this reason it is also referred to as a 'forward-backward' algorithm.

In order to explain the BCJR algorithm, it is better to write the APP from Equation (10.6) in terms of the joint probability of a transition in the trellis from the state s_ℓ at time instant ℓ, to the state $s_{\ell+1}$ at time instant $\ell + 1$,

$$\Lambda(i_\ell) = \log\Big(\frac{\sum_{S^+} \text{P}(s_\ell = s', s_{\ell+1} = s, \mathbf{y})}{\sum_{S^-} \text{P}(s_\ell = s', s_{\ell+1} = s, \mathbf{y})}\Big) = \log\Big(\frac{\sum_{S^+} p(s', s, \mathbf{y})}{\sum_{S^-} p(s', s, \mathbf{y})}\Big) \tag{10.7}$$

where $s, s' \in S$ are possible states of the convolutional encoder, S^+ is the set of ordered pairs (s', s) such that a transition from state s' to state s at time ℓ is caused by the input bit $i_\ell = 0$. Similarly, S^- is the set of transitions caused by $i_\ell = 1$. The probability $p(s', s, \mathbf{y})$ can be decomposed as follows:

$$p(s', s, \mathbf{y}) = p(s', \mathbf{y}_{t<\ell}) p(s, \mathbf{y}_\ell \mid s') p(\mathbf{y}_{t>\ell} \mid s) \tag{10.8}$$

where $\mathbf{y}_{t<\ell} \stackrel{\Delta}{=} [\mathbf{y}_0, \ldots, \mathbf{y}_{\ell-1}]$ and $\mathbf{y}_{t>\ell} \stackrel{\Delta}{=} [\mathbf{y}_\ell, \ldots, \mathbf{y}_{L+m-1}]$.

By defining

$$\alpha_\ell(s') \stackrel{\Delta}{=} p(s', \mathbf{y}_{t<\ell}) \tag{10.9}$$

$$\beta_{\ell+1}(s') \stackrel{\Delta}{=} p(\mathbf{y}_{t>\ell} \mid s) \tag{10.10}$$

$$\gamma_\ell(s', s) \stackrel{\Delta}{=} p(s, \mathbf{y}_\ell \mid s') \tag{10.11}$$

the APP in Equation (10.7) takes the usual form [19], i.e.

$$\Lambda(i_\ell) = \log\Big(\frac{\sum_{S^+} \alpha_\ell(s') \gamma_\ell(s', s) \beta_{\ell+1}(s)}{\sum_{S^-} \alpha_\ell(s') \gamma_\ell(s', s) \beta_{\ell+1}(s)}\Big) \tag{10.12}$$

The probability $\alpha_\ell(s')$ is computed iteratively in a forward recursion

$$\alpha_{\ell+1}(s) = \sum_{s' \in S} \alpha_\ell(s') \gamma_\ell(s', s) \tag{10.13}$$

with initial condition $\alpha_0(0) = 1$, $\alpha_0(s) = 0$ for $s \neq 0$, assuming that the initial state of the encoder is 0. Similarly, the probability $\beta_\ell(s')$ is computed via a backward recursion

$$\beta_\ell(s') = \sum_{s \in S} \beta_{\ell+1}(s) \gamma_\ell(s', s) \tag{10.14}$$

with initial condition $\beta_{L+m-1}(0) = 1$, $\beta_{L+m-1}(s) = 0$ for $s \neq 0$ assuming trellis termination is employed using tail bits. The transition probability can be computed as

$$\gamma_\ell(s', s) = \mathrm{P}(\mathbf{y}_\ell \mid \mathbf{x}_\ell) P(i_\ell)$$

The MAP algorithm consists of initializing $\alpha_0(s')$, $\beta_{L+m-1}(s)$, computing the branch metric $\gamma_\ell(s', s)$, and continuing the forward and backward recursion in Equations (10.13) and (10.14) to compute the updates of $\alpha_{\ell+1}(s)$ and $\beta_\ell(s')$.

The complexity of the BCJR is approximately three times that of the Viterbi decoder. In order to reduce the complexity a log-MAP decoder can be considered, where all operations are performed in the logarithmic domain. Thus, the forward and backward recursions $\alpha_{\ell+1}(s)$, $\beta_\ell(s')$ and $\gamma_\ell(s, s')$ are replaced by $\alpha^*_{\ell+1}(s) = \log[\alpha_{\ell+1}(s)]$, $\beta^*_\ell(s') = \log[\beta_\ell(s')]$ and $\gamma^*_\ell(s, s') = \log[\gamma_\ell(s, s')]$. This gives advantages for implementation and can be shown to be numerically more stable. By defining

$$\max^*(z_1, z_2) \stackrel{\Delta}{=} \log(e^{z_1} + e^{z_2}) = \max(z_1, z_2) + \log(1 + e^{-|z_1 - z_2|}) \tag{10.15}$$

it can be shown that the recursion in Equations (10.13) and (10.14) becomes

$$\begin{aligned} \alpha^*_{\ell+1}(s) &= \log\Big(\sum_{s' \in S} e^{\alpha^*_\ell(s') + \gamma^*_\ell(s', s)}\Big) \\ &= \max^*_{s' \in S}\{\alpha^*_\ell(s') + \gamma^*_\ell(s', s)\} \end{aligned} \tag{10.16}$$

with initial condition $\alpha^*_0(0) = 0$, $\alpha^*_0(s) = -\infty$ for $s \neq 0$. Similarly,

$$\beta^*_\ell(s') = \max^*_{s' \in S}\{\beta^*_{\ell+1}(s) + \gamma^*_\ell(s', s)\} \tag{10.17}$$

with initial condition $\beta^*_{L+m}(0) = 0$, $\beta^*_{L+m}(s) = -\infty$ for $s \neq 0$. Figure 10.5 shows a schematic representation of the forward and backward recursion.

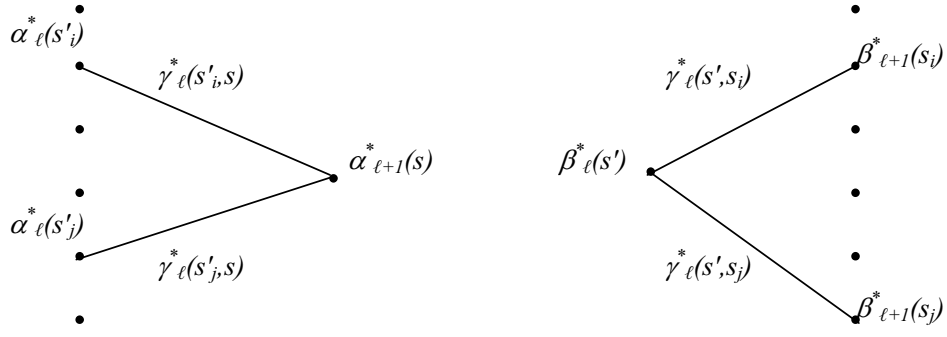


Figure 10.5: Forward and backward recursion for the BCJR decoding algorithm.

The APP in Equation (10.12) becomes

$$\Lambda(i_\ell) = \max^*_{(s',s)\in S^+}\{\alpha^*_\ell(s') + \gamma^*_\ell(s', s) + \beta^*_{\ell+1}(s)\} - \max^*_{(s',s)\in S^-}\{\alpha^*_\ell(s') + \gamma^*_\ell(s', s) + \beta^*_{\ell+1}(s)\} \quad (10.18)$$

In order to reduce the complexity further, the *max-log-MAP* approximation can be used, where the $\max^* = \log(e^x + e^y)$ is approximated by $\max(x, y)$. The advantage is that the algorithm is simpler and faster, with the complexity of the forward and backward passes being equivalent to a Viterbi decoder. The drawback is the loss of accuracy arising from the approximation.

Convolutional codes are the most widely used family of error-correcting codes owing to their reasonably good performance and the possibility of extremely fast decoders based on the VA, as well as their flexibility in supporting variable input codeword sizes. However, it is well known that there remains a significant gap between the performance of convolutional codes and the theoretical limits set by Shannon.[9] In the early 1990s, an encoding and decoding algorithm based on convolutional codes was proposed [20], which exhibited performance within a few tenths of a deciBel (dB) from the Shannon limit – the turbo code family was born. Immediately after turbo codes were discovered, Low-Density Parity Check (LDPC) codes [23–25] that also provided near-Shannon limit performance were rediscovered.

10.3.1.3 Turbo Codes

Berrou, Glavieux and Thitimajshima introduced turbo codes and the concept of iterative decoding to achieve near-Shannon limit performance [20, 26]. Some of the reasoning regarding probabilistic processing can be found in [18, 27], as recognized by Berrou in [28].

A turbo encoder consists of a concatenation of two convolutional encoders linked by an interleaver. For instance, the turbo encoder adopted in UMTS and LTE [10] is schematically represented in Figure 10.6 with two identical convolutional codes with generator polynomial given by $\mathbf{G} = [1, \mathbf{g}_0/\mathbf{g}_1]$ where $\mathbf{g}_0 = [1011]$ and $\mathbf{g}_1 = [1101]$. Thus, a turbo code encodes the input block twice (with and without interleaving) to generate two distinct set of parity bits. Each constituent encoder may be terminated to the all-zero state by using tail bits. The nominal code rate of the turbo code shown in Figure 10.6 is 1/3.

Like convolutional codes, the optimal decoder for turbo codes would ideally be the MAP or ML decoder. However the number of states in the trellis of a turbo code is significantly larger due to the interleaver Π, thus making a true ML or MAP decoder intractable (except for trivial block sizes). Therefore, Berrou *et al.* [26] proposed the principle of iterative decoding. This is based on a suboptimal approach using a separate optimal decoder for each constituent convolutional coder, with the two constituent decoders iteratively exchanging information via a (de)interleaver. The classical decoding structure is shown in Figure 10.7.

The SISO decoder for each constituent convolutional code can be implemented via the BCJR algorithm (or the MAP algorithm), which was briefly described in the previous section. The SISO decoder outputs an a posteriori LLR value for each information bit, and this can be used to obtain a hard decision estimate as well as a reliability estimate. In addition, the

[9] 'Shannon's theorem', otherwise known as the 'noisy-channel coding theorem', states that if and only if the information rate is less than the channel capacity, then there exists a coding-decoding scheme with vanishing error probability when the code's block length tends to infinity [21, 22].

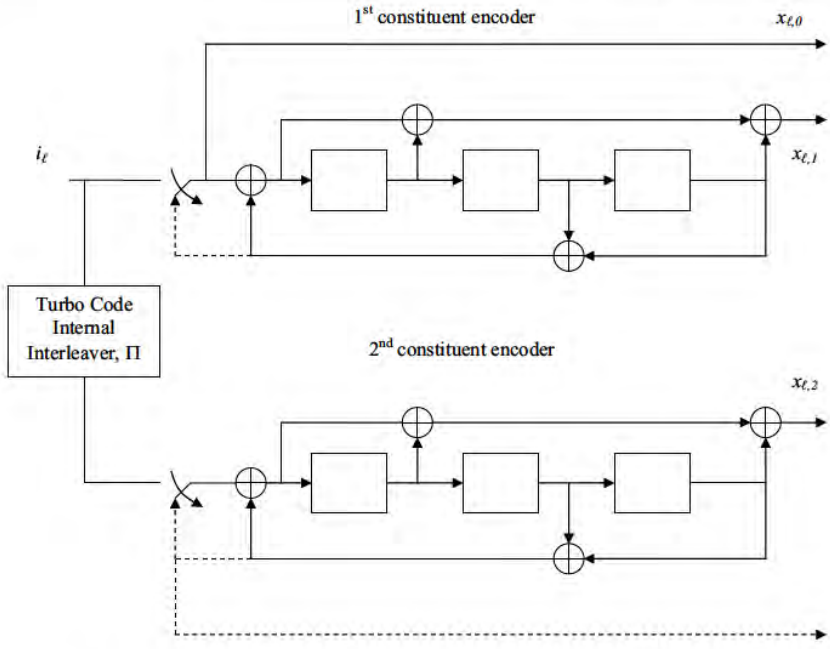


Figure 10.6: Schematic view of parallel turbo code used in LTE and UMTS [10].
Reproduced by permission of © 3GPP.

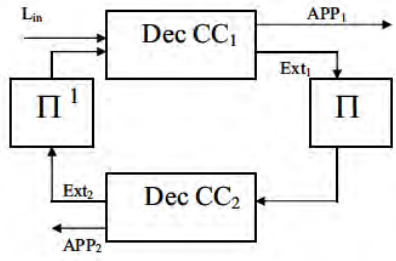


Figure 10.7: Schematic turbo decoder representation.

SISO decoder generates extrinsic LLRs (i.e. the extrinsic information of the bit x_ℓ is the LLR obtained from all the bits of the convolutional code except the bit corresponding to x_ℓ) for each information/coded bit that is utilized by the other decoder as a priori information after suitable (de)interleaving. The soft input to the first SISO decoder is the so-called *channel*

observation or channel likelihood, obtained as

$$\mathrm{L}_{\mathrm{in}}(x_{\ell,k}) = \frac{P(y_{\ell,k} \mid x_{\ell,k} = 1)}{P(y_{\ell,k} \mid x_{\ell,k} = -1)}$$

Thus the two decoders cooperate by iteratively exchanging the extrinsic information via the (de)interleaver. After a certain number of iterations (after convergence), the a posteriori LLR output can be used to obtain final hard decision estimates of the information bits.

10.3.1.4 Low-Density Parity-Check (LDPC) Codes

Low-Density Parity-Check (LDPC) codes were first studied by Gallager in his doctoral thesis [23] and have been extensively analysed by many researchers in recent years. LDPC codes are linear parity-check codes with a parity-check equation given by $H\mathbf{c}^T = 0$, where H is the $(n-k) \times n$ parity-check matrix of the code $C(k, n)$ and $\mathbf{c}$ is a length-n valid codeword belonging to the code C. Similarly to turbo codes, ML decoding for LDPC codes becomes too complex as the block size increases. In order to approximate ML decoding, Gallager in [23] introduced an iterative decoding technique which can be considered as the forerunner of message-passing algorithms [24, 29], which lead to an efficient LDPC decoding algorithm with a complexity which is linear with respect to the block size.

The term 'low-density' refers to the fact that the parity-check matrix entries are mostly zeros – in other words, the density of ones is low. The parity-check matrix of an LDPC code can be represented by a 'Tanner graph', in which two types of node, variable and check, are interconnected. The variable nodes and check nodes correspond to the codeword bits and the parity-check constraints respectively. A variable node v_j is connected to a check node c_i if the corresponding codeword bit participates in the parity-check equation, i.e. if $H(i, j) = 1$. Thus, the Tanner graph is an excellent tool by which to visualize the code constraints and to describe the decoding algorithms. Since an LDPC code has a low density of ones in H, the number of interconnections in the Tanner graph is small (and typically linear with respect to the codeword size). Figure 10.8 shows an example of a Tanner graph of a $(10, 5)$ code together with its parity-check matrix H.[10]

As mentioned above, LDPC codes are decoded using message-passing algorithms, such as Belief Propagation (BP) or the Sum-Product Algorithm (SPA). The idea of BP is to calculate approximate marginal a posteriori LLRs by applying Bayes' rule locally and iteratively at the nodes in the Tanner graph. The variable nodes and check nodes in the Tanner graph exchange LLR messages along their interconnections in an iterative fashion, thus cooperating with each other in the decoding process. Only extrinsic messages are passed along the interconnections to ensure that a node does not receive repeats of information which it already possesses, analogous to the exchange of extrinsic information in turbo decoding.

It is important to note that the number of operations per iteration in the BP algorithm is linear with respect to the number of message nodes. However, the BP decoding of LDPC codes is also suboptimal, like the turbo codes, owing to the cycles in the Tanner graph.[11] Cycles increase the dependencies between the extrinsic messages being received at each node during the iterative process. However, the performance loss can be limited by avoiding

[10] This particular matrix is not actually 'low density', but it is used for the sake of illustration.

[11] A cycle of length P in a Tanner graph is a set of P connected edges that starts and ends at the same node. A cycle verifies the condition that no node (except the initial and final nodes) appears more than once.

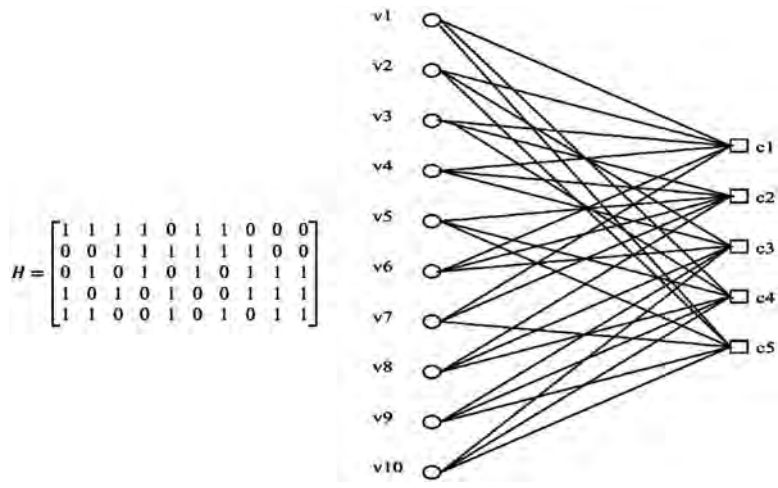


Figure 10.8: A Tanner graph representing the parity-check matrix of a (3, 6)-regular LDPC code $C(10, 5)$ and its associated parity-check matrix H.

excessive occurrences of short cycles. After a certain number of iterations, the a posteriori LLRs for the codeword bits are computed to obtain the final hard decision estimates. Typically, more iterations are required for LDPC codes to converge than are required for turbo codes. In practice, a variation of BP known as the Layered BP (LBP) algorithm is preferred, as it requires half as many iterations as the conventional BP algorithm [30].

From an encoding perspective, a straightforward algorithm based on the parity-check matrix would require a number of operations that is proportional to the square of the codeword size. However, there exist several structured LDPC code designs (including those used in the IEEE 802.16e, IEEE 802.11n and DVB-S2 standards) which have linear encoding complexity and extremely good performance. With structured or vectorized LDPC codes, the parity-check matrices are constructed using submatrices which are all-zero or circulant (shifted identity matrices), and such codes require substantially lower encoding/decoding complexity without sacrificing performance. Numerous references exist for LDPC code design and analysis, and the interested reader is referred, for example, to [23, 24, 29, 31–33] and references therein.

10.3.2 Channel Coding for Data Channels in LTE

Turbo codes found a home in UMTS relatively rapidly after their publication in 1993, with the benefits of their near-Shannon limit performance outweighing the associated costs of memory and processing requirements. Once the turbo code was defined in UMTS there was little incentive to reopen the specification as long as it was functional.

Therefore, although in UMTS Releases 5, 6 and 7 the core turbo code was not changed, it was enhanced by the ability to select different redundancy versions for HARQ

retransmissions. However, for LTE, with data rates of 100 Mbps to 300 Mbps in view, the UMTS turbo code needed to be re-examined, especially in terms of decoding complexity.

Sections 10.3.2.1 and 10.3.2.2 explain the eventual decision not to use LDPC codes for LTE, while replacing the turbo interleaver with a 'contention-free' interleaver. The following sections explain the specific differences between LTE channel coding and UMTS as summarized in Table 10.6.

Table 10.6: Major features of UMTS and LTE channel coding schemes.

Channel coding	UMTS	LTE
Constituent code	Tailed, eight states, $R = 1/3$ mother code	Same
Turbo interleaver	Row/column permutation	Contention-free quadratic permutation polynomial (QPP) interleaver
Rate matching	Performed on concatenated code blocks	Virtual Circular Buffer (CB) rate matching, performed per code block
Hybrid ARQ	Redundancy Versions (RVs) defined, Chase operation allowed	RVs defined on virtual CB, Chase operation allowed
Control channel	256-state tailed convolutional code	64-state tail-biting convolutional code, CB rate matching
Per-code-block operations	Turbo coding only	CRC attachment, turbo coding, rate matching, modulation mapping

10.3.2.1 The Lure of the LDPC Code

The draw of LDPC codes is clear: performance almost up to the Shannon limit, with claims of up to eight times less complexity than turbo codes. LDPC codes had also recently been standardized as an option in IEEE 802.16-2005. The complexity angle is all-important in LTE, provided that the excellent performance of the turbo code is maintained. However, LDPC proposals put forward for LTE claimed widely varying complexity benefits, from no benefit to 2.4 times [35] up to 7.35 times [36]. In fact, it turns out that the complexity benefit is code-rate dependent, with roughly a factor-of-two reduction in the operations count at code rate $1/2$ [37]. This comparison assumes that the LDPC code is decoded with the Layered Belief Propagation (LBP) decoder [30] while the turbo code is decoded with a log-MAP decoder.

On operations count alone, LDPC would be the choice for LTE. However, at least two factors curbed enthusiasm for LDPC. First, LDPC decoders have significant implementation complexity for memory and routing, which makes simple operation counts unrepresentative. Second, turbo codes were already standardized in UMTS, and a similar lengthy standardization process was undesired when perhaps a relatively simple enhancement to the known

turbo code would suffice. It was therefore decided to keep the same turbo code constituent encoders as in the UMTS turbo code, including the tailing method, but to enhance it using a new contention-free interleaver that enables efficient decoding at the high data rates targeted for LTE. Table 10.7 gives a comparison of some of the features of turbo and LDPC codes which are explained in the following sections.

Table 10.7: Comparison of turbo and LDPC codes.

	Turbo codes	LDPC codes
Standards	UMTS, LTE, WiMAX	IEEE 802.16e, IEEE 802.11n DVB-S2
Encoding	Linear time	Linear time with certain designs (802.16e, 802.11n)
Decoding	log-MAP and variants	Scheduling: Layered Belief Propagation, turbo-decoding Node processing: Min-Sum, normalized Min-Sum
Main decoding concern	Computationally intensive (more operations)	Memory intensive (more memory); routing network
Throughput	No inherent parallelism; Parallelism obtained via contention-free interleaver	Inherently parallelizable; Structured LDPC codes for flexible design
Performance	Comparable for information block sizes 10^3–10^4; Slightly better than LDPC at small block sizes; Four to eight iterations on average	Can be slightly better than turbo codes at large block lengths (10^4 or larger), iterative decoding threshold analysis; Very low error floor (e.g. at BLER around 10^{-7}); 10 to 15 iterations on average
HARQ	Simple for both Chase and IR (via mother code puncturing)	Simple for Chase; IR possible using model matrix extension or puncturing from a mother code
Complexity comparison (operations count)	Slightly higher operations count than LDPC at high code rates	Slightly lower operations count than turbo codes at high code rates

10.3.2.2 Contention-Free Turbo Decoding

The key to high data-rate turbo decoding is to parallelize the turbo decoder efficiently while maintaining excellent performance. The classical turbo decoder has two MAP decoders

(usually realized via a single hardware instantiation) that exchange extrinsic information in an iterative fashion between the two component codes, as described in Section 10.3.1. Thus, the first consideration is whether the parallelism is applied internally to the MAP decoder (single codeword parallelization) or externally by employing multiple turbo decoders (multiple codeword parallelizations with no exchange of extrinsic information between codewords).

If external parallelism is adopted, an input block may be split into X pieces, resulting in X codewords, and the increase in processing speed is obtained by operating up to X turbo decoders in parallel on the X codewords. In this case, in addition to a performance loss due to the smaller turbo interleaver size in each codeword, extra cost is incurred for memory and ancillary gate counts for forward-backward decoders. For this reason, one larger codeword with internal MAP parallelization is preferred. With one larger codeword, the MAP algorithm is run in parallel on each of the X pieces and the pieces can exchange extrinsic information, thus benefiting from the coding gains due to the large interleaver size. Connecting the pieces is most easily done with forward and backward state initialization based on the output of the previous iterations, although training within adjacent pieces during the current iteration is also possible [38]. As long as the size of each piece is large enough (e.g. 32 bits or greater), performance is essentially unaffected by the parallel processing.

While it is clear that the MAP algorithm can be parallelized, the MAP is not the entire turbo decoder. Efficient handling of the extrinsic message exchange (i.e. the (de)interleaving operation) is required as well. Since multiple MAP processors operate in parallel, multiple extrinsic values need to be read from or written to the memories concurrently. The memory accesses depend on the interleaver structure.

The existing UMTS interleaver has a problem with memory access contentions. Contentions result in having to read from or write to the same memory at the same time (as shown in Figure 10.9).

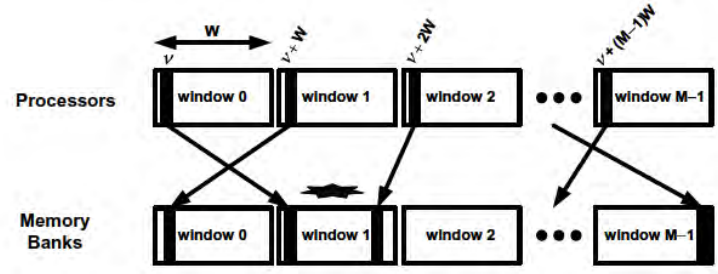


Figure 10.9: Memory access contention in window 1 due to concurrent access by Processors 0 and 2.

Contention resolution is possible (see, e.g., [39, 40]), but at a cost of extra hardware, and the resolution time (cycles) may vary for every supported interleaver size. Complex memory management may also be used as contention resolution for any arbitrary interleaver [41].

A new, Contention-Free (CF) interleaver solves the problem by ensuring that no contentions occur in the first place. An interleaver $\pi(i)$, $0 \leq i < K$, is said to be contention-free for a window size W when it satisfies the following for both interleaver $\psi = \pi$ and deinterleaver

$\psi = \pi^{-1}$ [42]:

$$\left\lfloor \frac{\psi(u_1 W + v)}{W} \right\rfloor \neq \left\lfloor \frac{\psi(u_2 W + v)}{W} \right\rfloor \tag{10.19}$$

where $0 \leq vW$, $u_1 \geq 0$ and $u_2 < M$ for all $u_1 \neq u_2$. The terms on both sides of Equation (10.19) are the memory bank indices that are accessed by the M processors on the v^{th} step. Inequality (10.19) requires that, for any given position v of each window, the memory banks accessed be unique between any two windows, thus eliminating access contentions. For significant hardware savings, instead of using M physically separate memories, it is better to use a single physical memory and fetch or store M values on each cycle from a single address. This requires the CF interleaver also to satisfy a vectorized decoding property where the intra-window permutation is the same for each window:

$$\pi(uW + v) \bmod W = \pi(v) \bmod W \tag{10.20}$$

for all $1 \leq u < M$ and $0 \leq v < W$.

Performance, implementation complexity and flexibility are concerns. However, even a simple CF interleaver composed of look-up tables (for each block size) and a bit-reversal permutation [43] can be shown to have excellent performance. In terms of flexibility, a maximally contention-free interleaver can have a parallelism order (number of windows) that is any factor of the block size. A variety of possible parallelism factors provides freedom for each individual manufacturer to select the degree of parallelism based on the target data rates for different UE categories.

After consideration of performance, available flexible classes of CF interleavers and complexity benefits, a new contention-free interleaver was selected for LTE.

10.3.2.3 The LTE Contention-Free Interleaver

The main choices for the CF interleaver included Almost Regular Permutation (ARP) [44] and Quadratic Permutation Polynomial (QPP) [45] interleavers. The ARP and QPP interleavers share many similarities and they are both suitable for LTE turbo coding, offering flexible parallelism orders for each block size, low-complexity implementation, and excellent performance. A detailed overview of ARP and QPP proposals for LTE (and their comparison with the UMTS Release-99 turbo interleaver) is given in [46]. Of these two closely competing designs, the QPP interleaver was selected for LTE as it provides more parallelism factors M and requires less storage, thus making it better-suited to high data rates.

For an information block size K, a QPP interleaver is defined by the following polynomial:

$$\pi(i) = (f_1 i + f_2 i^2) \bmod K \tag{10.21}$$

where i is the output index ($0 \leq i \leq K - 1$), $\pi(i)$ is the input index and f_1 and f_2 are the coefficients that define the permutation with the following conditions:

- f_1 is relatively prime to block size K;
- all prime factors of K also factor f_2.

The inverses of QPP interleavers can also be described via permutation polynomials but they are not necessarily quadratic, as such a requirement may result in inferior turbo code performance for certain block sizes. Therefore, it is better to select QPP interleavers with

low-degree inverse polynomials (the maximum degree of the inverse polynomial is equal to four for LTE QPP interleavers) without incurring a performance penalty. In general, permutation polynomials (quadratic and non-quadratic) are implementation-friendly as they can be realized using only adders in a recursive fashion. A total of 188 interleavers are defined for LTE, of which 153 have quadratic inverses while the remaining 35 have degree-3 and degree-4 inverses. The performance of the LTE QPP interleavers is shown alongside the UMTS turbo code in Figure 10.10.

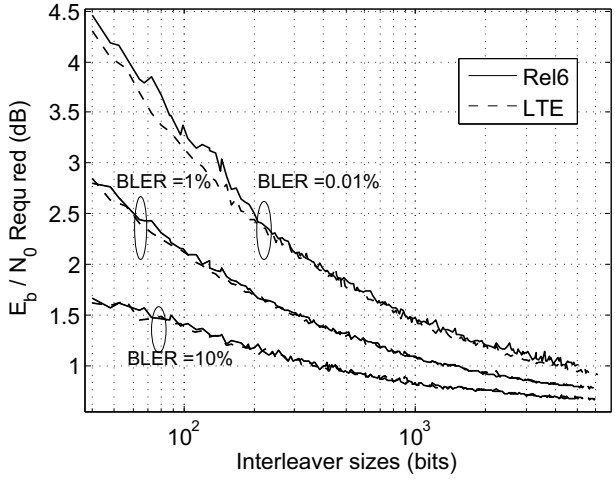


Figure 10.10: Performance of LTE QPP interleaver design versus UMTS turbo interleaver in static AWGN with QPSK modulation and eight iterations of a max-log-MAP decoder.

An attractive (and perhaps the most important) feature of QPP interleavers is that they are 'maximum contention-free', which means that they provide the maximum flexibility in supported parallelism, i.e. every factor of K is a supported parallelism factor. For example, for $K = 1024$, supported parallelism factors include $\{1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024\}$, although factors that result in a window size less than 32 may not be required in practice.

The QPP interleavers also have the 'even-even' property whereby even and odd indices in the input are mapped to even and odd indices respectively in the output; this enables the encoder and decoder to process two information bits per clock cycle, facilitating radix-4 implementations (i.e. all log-MAP decoding operations can process two bits at a time, including forward and backward recursions, extrinsic LLR generation and memory read and write).

One key impact of the decision to use a CF interleaver is that not all input block sizes are natively supported. As the amounts of parallelism depend on the factorization of the block size, certain block sizes (e.g. prime sizes) are not natively supportable by the turbo code. Since the input can be any size, filler bits are used whenever necessary to pad the input to the nearest QPP interleaver size. The QPP sizes are selected such that:

- The number of interleavers is limited (fewer interleavers implies more filler bits).
- The fraction of filler bits is roughly the same as the block size increases (spacing increases as block size increases).
- Multiple parallelism values are available (block sizes are spaced an integer number of bytes apart).

For performance within approximately 0.1 dB of the baseline UMTS interleaver, as few as 45 interleaver sizes [47] are feasible, although the percentage of filler bits may be high (nearly 12%). For LTE, the following 188 byte-aligned interleaver sizes spaced in a semi-log manner were selected with approximately 3% filler bits [10]:

$$K = \begin{cases} 40 + 8t & \text{if } 0 \le t \le 59 \quad \text{(40–512 in steps of 8 bits)} \\ 512 + 16t & \text{if } 0 < t \le 32 \quad \text{(528–1024 in steps of 16 bits)} \\ 1024 + 32t & \text{if } 0 < t \le 32 \quad \text{(1056–2048 in steps of 32 bits)} \\ 2048 + 64t & \text{if } 0 < t \le 64 \quad \text{(2112–6144 in steps of 64 bits)} \end{cases} \tag{10.22}$$

The maximum turbo interleaver size was also increased from 5114 in UMTS to 6144 in LTE, such that a 1500-byte TCP/IP packet would be segmented into only two segments rather than three, thereby minimizing the potential segmentation penalty and (marginally) increasing the turbo interleaver gain.

10.3.2.4 Rate-Matching

The Rate-Matching (RM) algorithm selects bits for transmission from the rate 1/3 turbo coder output via puncturing and/or repetition. Since the number of bits for transmission is determined based on the available physical resources, the RM should be capable of generating puncturing patterns for arbitrary rates. Furthermore, the RM should send as many new bits as possible in retransmissions to maximize the Incremental Redundancy (IR) HARQ gains (see Section 4.4 for further details about the HARQ protocol).

The main contenders for LTE RM were to use the same (or a similar) algorithm as HSPA, or to use Circular Buffer (CB) RM as in CDMA2000 1xEV and WiMAX. The primary advantage of the HSPA RM is that while it appears complex, it has been well studied and is well understood. However, a key drawback is that there are some severe performance degradations at higher code rates, especially near code rates 0.78 and 0.88 [48]. Therefore, circular buffer RM was selected for LTE, as it generates puncturing patterns simply and flexibly for any arbitrary code rate, with excellent performance.

In the CB approach, as shown in Figure 10.11, each of the three output streams of the turbo coder (systematic part, parity0, and parity1) is rearranged with its own interleaver (called a sub-block interleaver). In LTE, the 12 tail bits are distributed equally into the three streams as well, resulting in the sub-block size $K_s = K + 4$, where K is the QPP interleaver size. Then, an output buffer is formed by concatenating the rearranged systematic bits with the interlacing of the two rearranged parity streams. For any desired code rate, the coded bits for transmission are simply read out serially from a certain starting point in the buffer, wrapping around to the beginning of the buffer if the end of the buffer is reached.

A Redundancy Version (RV) specifies a starting point in the circular buffer to start reading out bits. Different RVs are specified by defining different starting points to enable HARQ

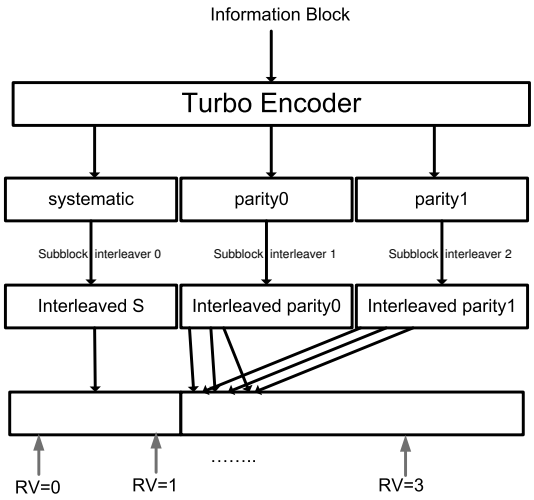


Figure 10.11: Rate-matching algorithm based on the CB. RV = 0 starts at an offset relative to the beginning of the CB to enable systematic bit puncturing on the first transmission.

operation. Usually RV = 0 is selected for the initial transmission to send as many systematic bits as possible. The scheduler can choose different RVs on transmissions of the same packet to support both IR and Chase combining HARQ.

The turbo code tail bits are uniformly distributed into the three streams, with all streams the same size. Each sub-block interleaver is based on the traditional row-column interleaver with 32 columns (for all block size) and a simple length-32 intra-column permutation.

- The bits of each stream are written row-by-row into a matrix with 32 columns (the number of rows is determined by the stream size), with dummy bits padded to the front of each stream to completely fill the matrix.
- A length-32 column permutation is applied and the bits are read out column-by-column to form the output of the sub-block interleaver. This sub-block interleaver has the property that it naturally first puts all the even indices and then all the odd indices into the rearranged sub-block. The permutation order is as follows:

  [0, 16, 8, 24, 4, 20, 12, 28, 2, 18, 10, 26, 6, 22, 14, 30, 1, 17, 9, 25, 5, 21, 13, 29, 3, 19, 11, 27, 7, 23, 15, 31]

Given the even-even property of the QPP permutations and the above property of the sub-block interleaver, the sub-block interleaver of the second parity stream is offset by an odd value $\delta = 1$ to ensure that the odd and even input indices have equal levels of protection. Thus, for index i, if $\pi_{\text{sys}}(i)$ denotes the permutation of the systematic bit sub-block interleaver, then the permutation of the two parity sub-block interleavers are $\pi_{\text{par0}}(i) = \pi_{\text{sys}}(i)$, and $\pi_{\text{par1}}(i) = (\pi_{\text{sys}}(i) + \delta) \bmod K_s$, where K_s is the sub-block size. With the offset $\delta = 1$ used in LTE, the first K_s parity bits in the interlaced parity portion of the circular buffer correspond to the K_s systematic bits, thus ensuring equal protection to all systematic bits [49]. Moreover, the offset enables the systematic bit puncturing feature whereby a small percentage of systematic bits are punctured in an initial transmission to enhance performance at high code rates. With the offset, RV = 0 results in partially systematic codes that are self-decodable at high coding

rates, i.e. avoiding the 'catastrophic' puncturing patterns which have been shown to exist at some code rates in UMTS.

A two-dimensional interpretation of the circular buffer (with a total of 96 columns) is shown in Figure 10.12 (where different shadings indicate bits from different streams and black cells indicate dummy bits). The one-dimensional CB may be formed by reading bits out column-by-column from the two-dimensional CB. Bits are read column-by-column starting from a column top RV location, and the dummy bits are discarded during the output bit generation. Although the dummy bits can be discarded during sub-block interleaving, in LTE the dummy bits are kept to allow a simpler implementation.

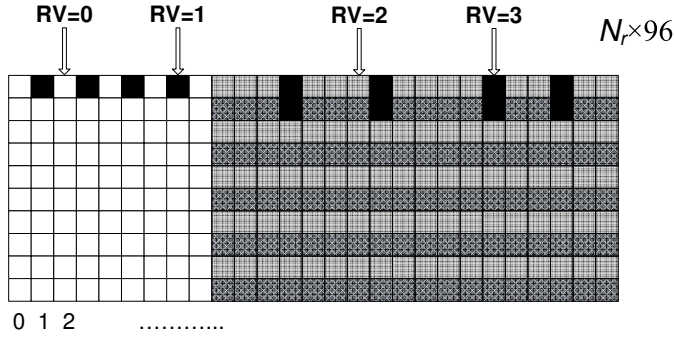


Figure 10.12: Two-dimensional visualization of the VCB. The starting points for the four RVs are the top of the selected columns.

This leads to the foremost advantage of the LTE CB approach, in that it enables efficient HARQ operation, since the CB operation can be performed without requiring an intermediate step of forming any actual physical buffer. In other words, for any combination of the 188 stream sizes and four RV values, the desired codeword bits can be equivalently obtained directly from the output of the turbo encoder using simple addressing based on sub-block permutation.

Therefore the term 'Virtual Circular Buffer' (VCB) is more appropriate in LTE. The LTE VCB operation also allows Systematic Bit Puncturing (SBP) by defining RV = 0 to skip the first two systematic columns of the CB, leading to approximately 6% punctured systematic bits (with no wrap around). Thus, with systematic bit puncturing and uniform spaced RVs, the four RVs start at the top of columns 2, 26, 50 and 74.

10.3.2.5 HARQ in LTE

The physical layer in LTE supports HARQ on the physical downlink and uplink shared channels, with separate control channels to send the associated acknowledgement feedback. In Frequency Division Duplex (FDD) operation, eight Stop-And-Wait (SAW) HARQ processes are available in both downlink and uplink with a typical Round-Trip Time (RTT) of 8 ms (see Figures 10.13 and 10.14 respectively). Each HARQ process requires a separate soft buffer allocation in the receiver for the purpose of combining the retransmissions. In

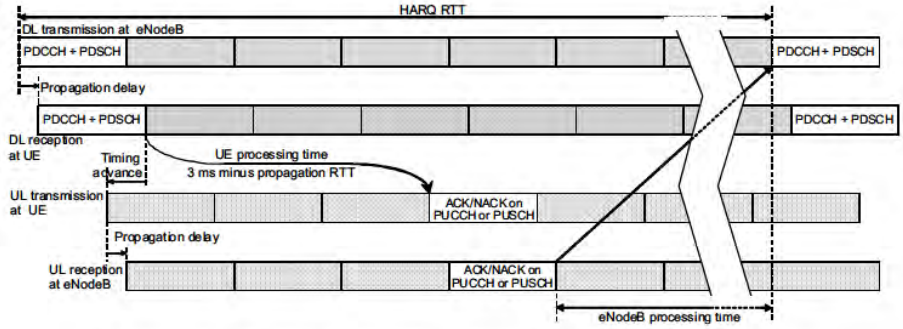


Figure 10.13: Timing diagram of the downlink HARQ (SAW) protocol.

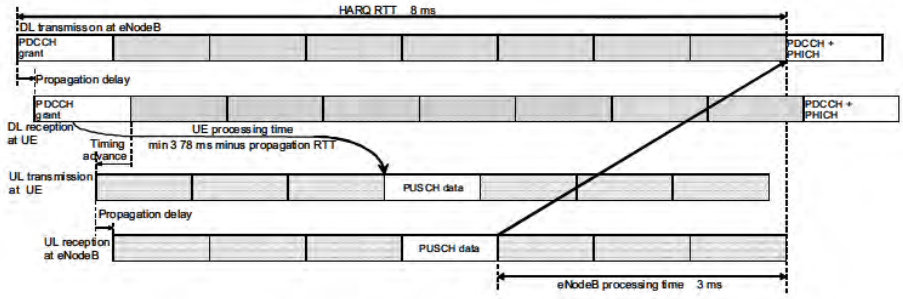


Figure 10.14: Timing diagram of the uplink HARQ (SAW) protocol.

FDD, the HARQ process to which a transport block belongs is identified by a unique three-bit HARQ process IDentifier (HARQ ID). In TDD, the number of HARQ processes depends on the uplink/downlink configuration as explained in Section 23.4.3, and four bits are used to identify the process.

There are several fields in the downlink control information (see Section 9.3.5.1) to aid the HARQ operation:

- New Data Indicator (NDI): toggled whenever a new packet transmission begins;
- Redundancy Version (RV): indicates the RV selected for the transmission or retransmission;
- MCS: Modulation and Coding Scheme.

As explained in Section 4.4.1.1, the LTE downlink HARQ is asynchronous and adaptive, and therefore every downlink transmission is accompanied by explicit signalling of control information. The uplink HARQ is synchronous, and either non-adaptive or adaptive. The uplink non-adaptive HARQ operation requires a predefined RV sequence 0, 2, 3, 1, 0, 2, 3, 1, ... for successive transmissions of a packet due to the absence

of explicit control signalling. For the adaptive HARQ operation, the RV is explicitly signalled. There are also other uplink modes in which the redundancy version (or the modulation) is combined with other control information to minimize control signalling overhead. Aspects of HARQ control signalling specifically related to TDD operation are discussed in Section 23.4.3.

10.3.2.6 Limited Buffer Rate Matching (LBRM)

A major contributor to the UE implementation complexity is the UE HARQ soft buffer size, which is the total memory (over all the HARQ processes) required for LLR storage to support HARQ operation. The aim of Limited Buffer Rate Matching (LBRM) is therefore to reduce the required UE HARQ soft buffer sizes while maintaining the peak data rates and having minimal impact on the system performance. [50] shows that there is a negligible performance impact even with a 50% reduction in the soft buffer size. Therefore, for LTE, up to 50% soft buffer reduction is enabled by means of LBRM for the higher UE categories (3, 4 and 5 – see Section 1.3.4) while it is not applied to the lower UE categories (1 and 2). This reduction corresponds to a mother code rate of 2/3 for the largest Transport Block[12] (TB).

For each UE category, the soft buffer size is determined based on the instantaneous peak data rate supported per subframe (see Table 1.2) multiplied by eight (i.e. using eight HARQ processes) and applying the soft buffer reduction factor as described above [51]. For FDD, the soft buffer is split into eight equal partitions (one partition per TB) if the UE is configured to receive Physical Downlink Shared CHannel (PDSCH) transmissions based on transmission modes other than 3 or 4 (i.e. one TB per HARQ process), or sixteen equal partitions for transmission modes 3 or 4 (i.e. two TBs per HARQ process). It is important to note that all transmission modes are supported by all UE categories.[13] Therefore, a category 1 UE could be in transmission modes 3 or 4 but may be limited to rank-1 operation as it can support only one layer in spatial multiplexing (see Section 11.2.2). For simplicity and to facilitate dual-mode UE development, the soft buffer sizes for TDD were chosen to be the same as those for FDD. However, the number of downlink HARQ processes in TDD varies between 4 and 15 according to the downlink/uplink configuration (see Section 6.2); hence the soft buffer is adapted and split into $\min(M_{\text{DL\_HARQ}}, M_{\text{limit}})$ equal partitions if the UE is configured to receive PDSCH transmissions based on transmission modes other than 3 or 4, or $2 \cdot \min(M_{\text{DL\_HARQ}}, M_{\text{limit}})$ equal partitions for transmission modes 3 or 4, where $M_{\text{DL\_HARQ}}$ is the number of downlink HARQ processes, and M_{limit} is equal to 8. Thus, when the number of downlink HARQ processes is greater than 8, statistical soft buffer management can be used for efficient UE implementation [52].

A soft buffer size per code block segment is derived from the soft buffer size per TB, and LBRM then simply shortens the length of the VCB of the code block segments for certain larger sizes of TB, with the RV spacing being compressed accordingly. With LBRM, the effective mother code rate for a TB becomes a function of the TB size and the allocated UE soft buffer size. Since the eNodeB knows the soft buffer capability of the UE, it only transmits those code bits out of the VCB that can be stored in the UE's HARQ soft buffer for all (re)transmissions of a given TB.

[12] A transport block is equivalent to a MAC PDU – see Section 4.4.

[13] The only exceptions are that PDSCH transmission modes 7 and 8 (see Section 9.2.2.1) are optional for FDD UEs, and PUSCH transmission mode 2 (see Section 29.4.1) is optional for UE categories 1–7.

Soft buffer management in case of carrier aggregation in Release 10 is discussed in Section 28.4.2.

10.3.2.7 Overall Channel Coding Chain for Data

The overall flow diagram of the turbo coded channels in LTE is summarized in Figure 10.15.

The physical layer first attaches a 24-bit Cyclic Redundancy Check (CRC) to each TB received from the MAC layer. This is used at the receiver to verify correct reception and to generate the HARQ ACK/NACK feedback.

The TB is then segmented into 'code blocks' according to a rule which, given an arbitrary TB size, is designed to minimize the number of filler bits needed to match the available QPP sizes. This is accomplished by allowing two adjacent QPP sizes to be used when segmenting a TB, rather than being restricted to a single QPP size. The filler bits would then be placed in the first segment. However, in LTE the set of possible TB sizes is restricted such that the segmentation rule described above always results in a single QPP size for each segment with no filler bits.

Following segmentation, a further 24-bit CRC is attached to each code block if the TB was split into two or more code blocks. This code-block-level CRC can be utilized to devise

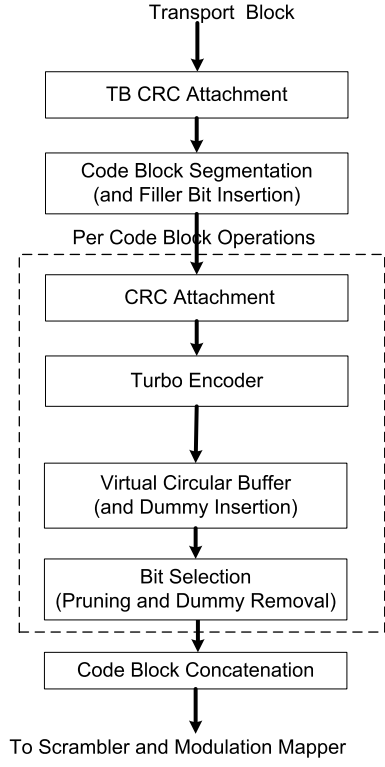


Figure 10.15: Flow diagram for turbo coded channels in LTE.

rules for early termination of the turbo decoding iterations to reduce decoding complexity. It is worth noting that the polynomial used for the code-block-level CRC is different from the polynomial used for the transport block CRC. This is a deliberate design feature in order to avoid increasing the probability of failing to detect errors as a result of the use of individual CRCs per code block; if all the code block CRCs pass, the decoder should still check the transport block CRC, which, being based on a different polynomial, is likely to detect an error which was not detected by a code-block CRC.

A major difference between LTE and HSDPA is that in LTE most of the channel coding operations on the PDSCH are performed at a code-block level, whereas in HSDPA only turbo coding is performed at the code block level. Although in the LTE specifications [10] the code block concatenation for transmission is done prior to the scrambler and modulation mapper, each code block is associated with a distinct set of modulation symbols, implying that the scrambling and modulation mapping operations may in fact be done individually for each code block, which facilitates an efficient pipelined implementation.

10.3.3 Channel Coding for Control Channels in LTE

Unlike the data, control information (such as is sent on the Physical Downlink Control CHannel (PDCCH) and Physical Broadcast CHannel (PBCH)!channel coding) is coded with a convolutional code, as the code blocks are significantly smaller and the additional complexity of the turbo coding operation is therefore not worthwhile.

The PDCCH is especially critical from a decoding complexity point of view, since a UE must decode a large number of potential control channel locations as discussed in Section 9.3.5.5. Another relevant factor in the code design for the PDCCH and the PBCH is that they both carry a relatively small number of bits, making the tail bits a more significant overhead. Therefore, it was decided to adopt a tail-biting convolutional code for LTE but, in order to limit the decoding complexity, using a new convolutional code with only 64 states instead of the 256-state convolutional code used in UMTS. These key differences are summarized in Table 10.8.

The LTE convolutional code, shown in Figure 10.2, offers slightly better performance for the target information block sizes, as shown in Figure 10.16. The initial value of the shift register of the encoder is set to the values corresponding to the last six information

Table 10.8: Differences between LTE and HSPA convolutional coding.

Property	LTE convolutional code	HSPA convolutional code
Number of states	64	256
Tailing method	Tail-biting	Tailed
Generators	[133, 171, 165](oct) $R = 1/3$	[561, 753](oct) $R = 1/2$ [557, 663, 711](oct) $R = 1/3$
Normalized decoding complexity	1/2 (assuming two iterations of decoding)	1
Rate matching	Circular buffer	Algorithmic calculation of rate-matching pattern

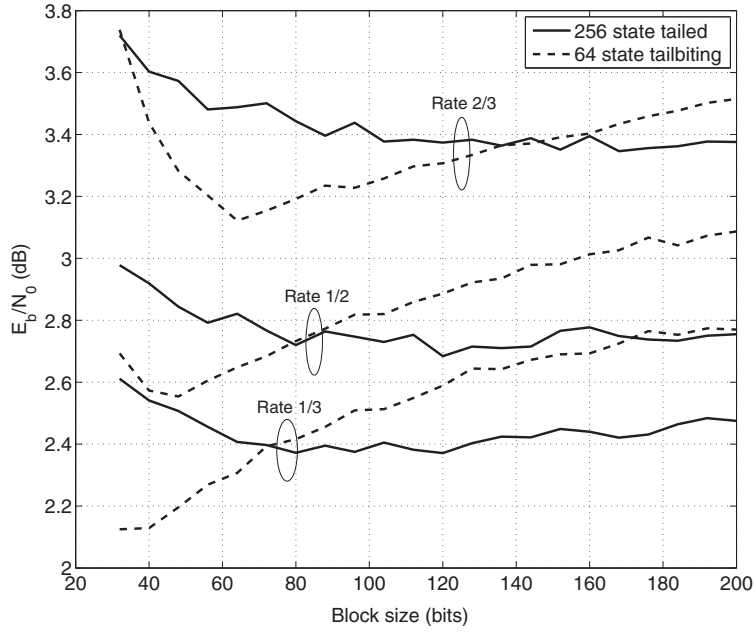


Figure 10.16: E_b/N_0 (dB) versus block size (bits) for BLER target of 1% for convolutional codes with rate $1/3$, $1/2$, and $2/3$. The HSPA convolutional code has 256-state code with tail bits and the 64-state LTE code is tail-biting.

bits in the input stream so that the initial and final states of the shift register are the same. The decoder can utilize a Circular Viterbi Algorithm [14] or MAP algorithm [53], with decoding complexity approximately twice that of the Viterbi decoder with two iterations (passes through the trellis). With only a quarter the number of states, the overall complexity of the LTE convolutional code can therefore be argued to be half that of the HSPA code, provided that only two iterations are used.

The rate-matching for the convolutional code in LTE uses a similar CB method as for the turbo code. A 32-column interleaver is used, with no interlacing in the CB (the three parity streams are concatenated in the CB). This structure gives good performance at higher code rates as well as lower code rates, and therefore LTE has no need for an additional (different) $R = 1/2$ generator polynomial as used in UMTS.

Other even smaller items of control information use block codes (for example, a Reed–Muller code for CQI, or a simple list of codewords for the Physical Control Format Indicator CHannel (PCFICH) – see Section 9.3.3). With small information words, block codes lend themselves well to a Maximum Likelihood (ML) decoding approach.

10.4 Conclusions

The LTE channel coding is a versatile design that has benefited from the decades of research and development in the area of iterative processing. Although the turbo codes used in LTE

and UMTS are of the same form as Berrou's original scheme, the LTE turbo code with its contention-free interleaver provides hardware designers with sufficient design flexibility to support the high data rates offered by LTE and LTE-Advanced. However, with increased support for parallelism comes the cost of routing the extrinsic values to and from the memory. The routing complexity in the turbo decoder with a large number of processors (e.g. $M = 64$) may in fact become comparable to that of an LDPC code with similar processing capability. Therefore, it is possible that the cost versus performance trade-offs between turbo and LDPC codes may be reinvestigated in the future. Nevertheless, it is clear that the turbo code will continue to shine for a long time to come.

References[14]

[1] A. J. Goldsmith and S. G. Chua, 'Adaptive Coded Modulation for Fading Channels'. *IEEE Trans. on Communications*, Vol. 46, pp. 595–602, May 1998.

[2] J. Hayes, 'Adaptive feedback communications'. *IEEE Trans. on Communication Technologies*, Vol. 16, pp. 15–22, February 1968.

[3] NTT DoCoMo, Fujitsu, Mitsubishi Electric Corporation, NEC, QUALCOMM Europe, Sharp, and Toshiba Corporation, 'R1-060039: Adaptive Modulation and Channel Coding Rate Control for Single-antenna Transmission in Frequency Domain Scheduling in E-UTRA Downlink', www.3gpp.org, 3GPP TSG RAN WG1, LTE ad-hoc meeting, Helsinki, Finland, January 2006.

[4] LG Electronics, 'R1-060051: Link Adaptation in E-UTRA Downlink', www.3gpp.org, 3GPP TSG RAN WG1, LTE ad-hoc meeting, Helsinki, Finland, January 2006.

[5] NTT DoCoMo, Fujitsu, Mitsubishi Electric Corporation, NEC, Sharp, and Toshiba Corporation, 'R1-060040: Adaptive Modulation and Channel Coding Rate Control for MIMO Transmission with Frequency Domain Scheduling in E-UTRA Downlink', www.3gpp.org, 3GPP TSG RAN WG1, LTE ad-hoc meeting, Helsinki, Finland, January 2006.

[6] Samsung, 'R1-060076: Adaptive Modulation and Channel Coding Rate', www.3gpp.org, 3GPP TSG RAN WG1, LTE ad-hoc meeting, Helsinki, Finland, January 2006.

[7] Huawei, 'R1-080555: Labelling Complexity of UE Selected Subbands on PUSCH', www.3gpp.org, 3GPP TSG RAN WG1, meeting 51bis, Sevilla, Spain, January 2008.

[8] Huawei, 'R1-080182: Labelling of UE-selected Subbands on PUSCH', www.3gpp.org, 3GPP TSG RAN WG1, meeting 51bis, Sevilla, Spain, January 2008.

[9] 3GPP Technical Specification 36.213, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures', www.3gpp.org.

[10] 3GPP Technical Specification 36.212, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and Channel Coding (FDD)', www.3gpp.org.

[11] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.

[12] J. H. Ma and J. W. Wolf, 'On Tail-Biting Convolutional Codes'. *IEEE Trans. on Communications*, Vol. 34, pp. 104–111, February 1986.

[13] R. V. Cox and C. V. Sundberg, 'An Efficient Adaptive Circular Viterbi Algorithm for Decoding Generalized Tailbiting Convolutional Codes'. *IEEE Trans. on Vehicular Technology*, Vol. 43, pp. 57–68, February 1994.

[14] H. Ma and W. J. Wolf, 'On Tail-biting Convolutional Codes'. *IEEE Trans. on Communications*, Vol. 34, pp. 104–111, February 1986.

[14] All web sites confirmed 1st March 2011.

[15] S. Lin and D. J. Costello, *Error and Control Coding: Fundamentals and Applications*. Second Edition, Prentice Hall: Englewood Cliffs, NJ, 2004.

[16] A. J. Viterbi, 'Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm'. *IEEE Trans. on Information Theory*, Vol. 13, pp. 260–269, April 1967.

[17] G. D. Forney Jr., 'The Viterbi Algorithm', in *Proc. of the IEEE*, March 1973.

[18] J. Haguenauer and P. Hoeher, 'A Viterbi algorithm with soft-decision outputs and its applications', in *Proc. of IEEE Globecom*, Dallas, TX, 1989.

[19] L. R. Bahl, J. Cocke, F. Jelinek and J. Raviv, 'Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate'. *IEEE Trans. on Information Theory*, Vol. 20, pp. 284–287, March 1974.

[20] C. Berrou and A. Glavieux, 'Near Optimum Error Correcting Coding and Decoding: Turbo-codes'. *IEEE Trans. on Communications*, Vol. 44, pp. 1261–1271, October 1996.

[21] C. E. Shannon, 'Communication in the Presence of Noise', in *Proc. IRE*, Vol. 37, pp. 10–21, January 1949.

[22] C. E. Shannon, *A Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1998.

[23] R. G. Gallager, 'Low-Density Parity-Check Codes'. Ph.D. Thesis, Cambridge, MA: MIT Press, 1963.

[24] S. Y. Chung, 'On the Construction of Some Capacity-Approaching Coding Scheme'. Ph.D. Thesis, Massachusetts Institute of Technology, September 2000.

[25] D. J. C. MacKay and R. M. Neal, 'Near Shannon Limit Performance of Low Density Parity Check Codes'. *IEEE Electronics Letters*, Vol. 32, pp. 1645–1646, August 1996.

[26] C. Berrou, P. Glavieux and A. Thitimajshima, 'Near Shannon Limit Error-correcting Coding and Decoding', in *Proc. IEEE International Conference on Communications*, Geneva, Switzerland, 1993.

[27] G. Battail, 'Coding for the Gaussian Channel: the Promise of Weighted-Output Decoding'. *International Journal of Satellite Communications*, Vol. 7, pp. 183–192, 1989.

[28] C. Berrou, 'The Ten-Year Old Turbo Codes Are Entering into Service'. *IEEE Communication Magazine*, Vol. 41, no. 8, 2003.

[29] T. J. Richardson and R. L Urbanke, 'The Capacity of Low-density Parity-check Codes Under Message-passing Decoding'. *IEEE Trans. on Information Theory*, Vol. 47, pp. 599–618, February 2001.

[30] M. M. Mansour and N. Shanbhag, 'High-throughput LDPC Decoders'. *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, Vol. 11, pp. 976–996, October 2003.

[31] T. J. Richardson, M. A Shokrollahi and R. L Urbanke, 'Design of Capacity-approaching Irregular Low-density Parity-check Codes'. *IEEE Trans. on Information Theory*, Vol. 47, pp. 619–637, February 2001.

[32] S-Y. Chung, G. D., Jr. Forney, T. J. Richardson and R. Urbanke, 'On the Design of Low-density Parity-check Codes within 0.0045 dB of the Shannon Limit'. *IEEE Communication Letters*, Vol. 5, pp. 58–60, February 2001.

[33] J. Chen, R. M. Tanner, J. Zhang and M. P. C Fossorier, 'Construction of Irregular LDPC Codes by Quasi-Cyclic Extension'. *IEEE Trans. on Information Theory*, Vol. 53, pp. 1479–1483, 2007.

[34] M. Bickerstaff, L. Davis, C. Thomas, D. Garrett and C. Nicol, 'A 24 Mb/s Radix-4 LogMAP Turbo Decoder for 3GPP-HSDPA Mobile Wireless', in *Proc. of 2003 IEEE International Solid State Circuits Conference*, San Francisco, CA, 2003.

[35] Motorola, 'R1-060384: LDPC Codes for EUTRA', www.3gpp.org, 3GPP TSG RAN WG1, meeting 44, Denver, USA, February 2006.

[36] Samsung, 'R1-060334: LTE Channel Coding', www.3gpp.org, 3GPP TSG RAN WG1, meeting 44, Denver, USA, February 2006.

[37] Intel, ITRI, LG, Mitsubishi, Motorola, Samsung, and ZTE, 'R1-060874: Complexity Comparison of LDPC Codes and Turbo Codes', www.3gpp.org, 3GPP TSG RAN WG1, meeting 44bis, Athens, Greece, March 2006.

[38] T. K. Blankenship, B. Classon and V. Desai, 'High-Throughput Turbo Decoding Techniques for 4G', in *Proc. of International Conference 3G and Beyond*, pp. 137–142, 2002.

[39] Panasonic, 'R1-073357: Interleaver for LTE turbo code', www.3gpp.org, 3GPP TSG RAN WG1, meeting 47, Riga, Latvia, November 2006.

[40] Nortel and Samsung, 'R1-073265: Parallel Decoding Method for the Current 3GPP Turbo Interleaver', www.3gpp.org, 3GPP TSG RAN WG1, meeting 47, Riga, Latvia, November 2006.

[41] A. Tarable, S. Benedetto and G. Montorsi, 'Mapping Interleaving Laws to Parallel Turbo and LDPC Decoder Architectures'. *IEEE Trans. on Information Theory*, Vol. 50, pp. 2002–2009, September 2004.

[42] A. Nimbalker, T. K. Blankenship, B. Classon, T. Fuja, and D. J. Costello Jr, 'Contention-free Interleavers', in *Proc. IEEE International Symposium on Information Theory*, Chicago, IL, 2004.

[43] A. Nimbalker, T. K. Blankenship, B. Classon, T. Fuja and D. J. Costello Jr, 'Inter-window Shuffle Interleavers for High Throughput Turbo Decoding' in *Proc. of 3rd International Symposium on Turbo Codes and Related Topics*, Brest, France, 2003.

[44] C. Berrou, Y. Saouter, C. Douillard, S. Kerouedan and M. Jezequel, 'Designing Good Permutations for Turbo Codes: Towards a Single Model', in *Proc. IEEE International Conference on Communications*, Paris, France, 2004.

[45] O. Y. Takeshita, 'On Maximum Contention-free Interleavers and Permutation Polynomials over Integer Rings'. *IEEE Trans. on Information Theory*, Vol. 52, pp. 1249–1253, March 2006.

[46] A. Nimbalker, Y Blankenship, B. Classon, T. Fuja and T. K. Blankenship, 'ARP and QPP Interleavers for LTE Turbo Coding', in *Proc. IEEE Wireless Communications and Networking Conference*, 2008.

[47] Motorola, 'R1-063061: A Contention-free Interleaver Design for LTE Turbo Codes', www.3gpp.org, 3GPP TSG RAN WG1, meeting 47, Riga, Latvia, November 2006.

[48] Siemens, 'R1-030421: Turbo Code Irregularities in HSDPA', www.3gpp.org, 3GPP TSG RAN WG1, meeting 32, Marne La Vallée, Paris, May 2003.

[49] Motorola, 'R1-071795: Parameters for Turbo Rate-Matching', www.3gpp.org, 3GPP TSG RAN WG1, meeting 48bis, Malta, March 2007.

[50] Motorola, 'R1-080058: Limited Buffer Rate Matching Ű System performance', www.3gpp.org, 3GPP TSG RAN WG1, meeting 51bis, Seville, Spain, January 2008.

[51] Motorola, Nokia Siemens Networks 'R1-082123: Adjustments to UE Downlink Soft Buffer Sizes based on LTE transport block sizes', www.3gpp.org, 3GPP TSG RAN WG1, meeting 53, Kansas City, USA, May 2008.

[52] Ericsson, 'R1-082018: On soft buffer usage for LTE TDD', www.3gpp.org, 3GPP TSG RAN WG1, meeting 53, Kansas City, USA, May 2008.

[53] R. Y. Shao, S. Lin, and M. Fossorier, 'An Iterative Bidirectional Decoding Algorithm for Tail Biting Codes', in *Proc. of IEEE Information Theory Workshop (ITW)*, Kruger National Park, South Africa, 1999.

11

Multiple Antenna Techniques

Thomas Sälzer, David Gesbert, Cornelius van Rensburg, Filippo Tosato, Florian Kaltenberger and Tetsushi Abe

11.1 Fundamentals of Multiple Antenna Theory

11.1.1 Overview

The value of multiple antenna systems as a means to improve communications was recognized in the very early ages of wireless transmission. However, most of the scientific progress in understanding their fundamental capabilities has occurred only in the last 20 years, driven by efforts in signal and information theory, with a key milestone being achieved with the invention of so-called Multiple-Input Multiple-Output (MIMO) systems in the mid-1990s.

Although early applications of beamforming concepts can be traced back as far as 60 years in military applications, serious attention has been paid to the utilization of multiple antenna techniques in mass-market commercial wireless networks only since around 2000. Today, the key role which MIMO technology plays in the latest wireless communication standards for Personal, Wide and Metropolitan Area Networks (PANs, WANs and MANs) testifies to its importance. Aided by rapid progress in the areas of computation and circuit integration, this trend culminated in the adoption of MIMO for the first time in a cellular mobile network standard in the Release 7 version of HSDPA (High Speed Downlink Packet Access); soon after, the development of LTE broke new ground in being the first global mobile cellular system to be designed with MIMO as a key component from the start.

This chapter first provides the reader with the theoretical background necessary for a good understanding of the role and advantages promised by multiple antenna techniques in

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

wireless communications in general.[1] The chapter focuses on the intuition behind the main technical results and show how key progress in information theory yields practical lessons in algorithm and system design for cellular communications. As can be expected, there is still a gap between the theoretical predictions and the performance achieved by schemes that must meet the complexity constraints imposed by commercial considerations.

We distinguish between single-user MIMO and multi-user MIMO, although a common set of concepts captures the essential MIMO benefits in both cases. Single-user MIMO techniques dominated the algorithms selected for the first version of LTE, with multi-user MIMO becoming more established in Releases 9 and 10.

The second part of the chapter describes the actual methods adopted for LTE, paying particular attention to the combinations of theoretical principles and system design constraints that led to these choices.

While traditional wireless communications (Single-Input Single-Output (SISO)) exploit time- or frequency-domain pre-processing and decoding of the transmitted and received data respectively, the use of additional antenna elements at either the base station (eNodeB) or User Equipment (UE) side (on the downlink or uplink) opens an extra spatial dimension to signal precoding and detection. Space-time processing methods exploit this dimension with the aim of improving the link's performance in terms of one or more possible metrics, such as the error rate, communication data rate, coverage area and spectral efficiency (expressed in bps/Hz/cell).

Depending on the availability of multiple antennas at the transmitter and/or the receiver, such techniques are classified as Single-Input Multiple-Output (SIMO), Multiple-Input Single-Output (MISO) or MIMO. Thus in the scenario of a multi-antenna-enabled base station communicating with a single antenna UE, the uplink and downlink are referred to as SIMO and MISO respectively. When a multi-antenna terminal is involved, a full MIMO link may be obtained, although the term MIMO is sometimes also used in its widest sense, thus including SIMO and MISO as special cases. While a point-to-point multiple-antenna link between a base station and one UE is referred to as Single-User MIMO (SU-MIMO), Multi-User MIMO (MU-MIMO) features several UEs communicating simultaneously with a common base station using the same frequency- and time-domain resources.[2] By extension, considering a multicell context, neighbouring base stations sharing their antennas in virtual MIMO fashion [4] to communicate with the same set of UEs in different cells comes under the term Coordinated MultiPoint (CoMP) transmission/reception. This latter scenario is not supported in the first versions of LTE but it is being studied for possible inclusion in later releases of LTE-Advanced as described in Section 29.5. The overall evolution of MIMO concepts, from the simplest diversity setup to the advanced joint-processing CoMP technique, is illustrated in Figure 11.1.

Despite their variety and sometimes perceived complexity, single-user and multi-user MIMO techniques tend to revolve around just a few fundamental principles, which aim at leveraging some key properties of multi-antenna radio propagation channels. As introduced

[1] For more exhaustive tutorial information on MIMO systems, the reader is referred, for example, to [1–3].

[2] Note that, in LTE, a single eNodeB may in practice control multiple cells; in such a case, we consider each cell as an independent base station for the purpose of explaining the MIMO techniques; the simultaneous transmissions in the different cells address different UEs and are typically achieved using different fixed directional physical antennas; they are therefore not classified as multi-user MIMO.

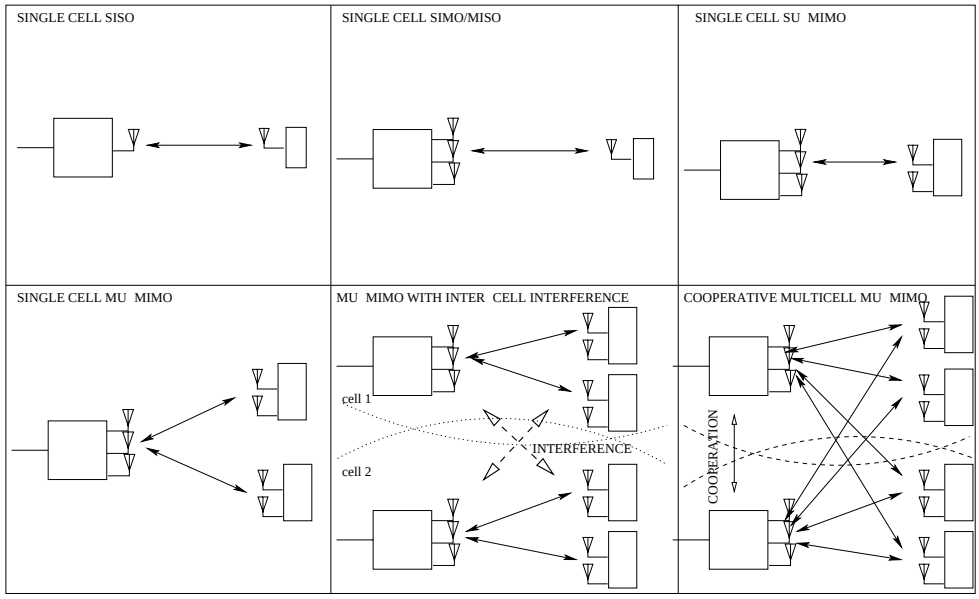


Figure 11.1: The evolution of MIMO technology, from traditional single antenna communication, to multi-user MIMO scenarios, to the possible multicell MIMO networks of the future.

in Section 1.3, there are basically three advantages associated with such channels (over their SISO counterparts):

- Diversity gain.
- Array gain.
- Spatial multiplexing gain.

Diversity gain corresponds to the mitigation of the effect of multipath fading, by means of transmitting or receiving over multiple antennas at which the fading is sufficiently decorrelated. It is typically expressed in terms of an *order*, referring either to the number of effective independent diversity branches or to the slope of the Bit Error Rate (BER) curve as a function of the Signal-to-Noise Ratio (SNR) (or possibly in terms of an SNR gain in the system's link budget).

While diversity gain is fundamentally related to improvement of the *statistics* of instantaneous SNR in a fading channel, array gain and multiplexing gain are of a different nature, rather being related to the geometry and the theory of vector spaces. Array gain corresponds to a spatial version of the well-known matched-filter gain in time-domain receivers, while multiplexing gain refers to the ability to send multiple data streams in parallel and to separate them on the basis of their spatial signature. The latter is much akin to the multiplexing of users separated by orthogonal spreading codes, timeslots or frequency assignments, with the great advantage that, unlike Code, Time or Frequency Division Multiple Access (CDMA, TDMA or FDMA), MIMO multiplexing does not come at the cost of bandwidth expansion; it does,

however, suffer the expense of added antennas and signal processing complexity and its gains strongly depend on the spatial characteristics of the propagation channel (see Chapter 20).

These aspects are analysed further in the following sections, considering first theoretically optimal transmission schemes and then popular MIMO approaches. First a base station to single-user communication model is considered (to cover single-user MIMO); then the techniques are generalized to multi-user MIMO.

The schemes adopted in LTE Releases 8 and 9, specifically for the downlink, are addressed subsequently. We focus on the Frequency Division Duplex (FDD) case. Discussion of some aspects of MIMO which are specific to Time Division Duplex (TDD) operation can be found in Section 23.5.

11.1.2 MIMO Signal Model

Let $\mathbf{Y}$ be a matrix of size $N \times T$ denoting the set of (possibly precoded) signals being transmitted from N distinct antennas over T symbol durations (or, in the case of some frequency-domain systems, T subcarriers), where T is a parameter of the MIMO algorithm (defined below). Thus the n^{th} row of $\mathbf{Y}$ corresponds to the signal emitted from the n^{th} transmit antenna. Let $\mathbf{H}$ denote the $M \times N$ channel matrix modelling the propagation effects from each of the N transmit antennas to any one of M receive antennas, over an arbitrary subcarrier whose index is omitted here for simplicity. We assume $\mathbf{H}$ to be invariant over T symbol durations. The matrix channel is represented by way of example in Figure 11.2. Then the $M \times T$ signal $\mathbf{R}$ received over T symbol durations over this subcarrier can be conveniently written as

$$\mathbf{R} = \mathbf{HY} + \mathbf{N} \tag{11.1}$$

where $\mathbf{N}$ is the additive noise matrix of dimension $M \times T$ over all M receiving antennas. We will use $\mathbf{h}_i$ to denote the i^{th} column of $\mathbf{H}$, which will be referred to as the *receive spatial signature* of (i.e. corresponding to) the i^{th} transmitting antenna. Likewise, the j^{th} row of $\mathbf{H}$ can be termed the *transmit spatial signature* of the j^{th} receiving antenna.

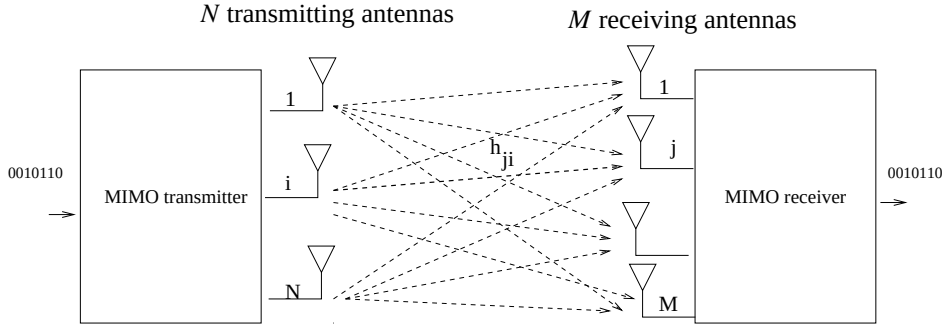


Figure 11.2: Simplified transmission model for a MIMO system with N transmit antennas and M receive antennas, giving rise to an $M \times N$ channel matrix, with MN links.

Let $\mathbf{X} = (x_1, x_2, \ldots, x_P)$ be a group of P QAM[3] symbols to be sent to the receiver over the T symbol durations; these symbols must be *mapped* to the transmitted signal $\mathbf{Y}$ before launching into the air. The choice of this mapping function $\mathbf{X} \rightarrow \mathbf{Y}(\mathbf{X})$ determines which one out of several possible MIMO transmission methods results, each yielding a different combination of the diversity, array and multiplexing gains. Meanwhile, the so-called *spatial* rate of the chosen MIMO transmission method is given by the ratio P/T.

Note that, in the most general case, the considered transmit (or receive) antennas may be attached to a single transmitting (or receiving) device (base station or UE), or distributed over different devices. The symbols in $(x_1, x_2, \ldots, x_P)$ may also correspond to the data of one or possibly multiple users, giving rise to the single-user MIMO or multi-user MIMO models.

11.1.3 Single-User MIMO Techniques

Several classes of SU-MIMO transmission methods are discussed below, both optimal and suboptimal.

11.1.3.1 Optimal Transmission over MIMO Systems

The optimal way of communicating over the MIMO channel involves a channel-dependent precoder, which fulfils the roles of both transmit beamforming and power allocation across the transmitted streams, and a matching receive beamforming structure. Full channel knowledge is therefore required at the transmit side for this method to be applicable. Consider a set of $P = NT$ symbols to be sent over the channel. The symbols are separated into N streams (or layers) of T symbols each. Stream i consists of symbols $[x_{i,1}, x_{i,2}, \ldots, x_{i,T}]$. Note that in an ideal setting, each stream may adopt a distinct code rate and modulation. This is discussed in more detail below. The transmitted signal can now be written as

$$\mathbf{Y}(\mathbf{X}) = \mathbf{V}\mathbf{P}\bar{\mathbf{X}} \tag{11.2}$$

where

$$\bar{\mathbf{X}} = \begin{pmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,T} \\ \vdots & \vdots & & \vdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,T} \end{pmatrix} \tag{11.3}$$

$\mathbf{V}$ is an $N \times N$ transmit beamforming matrix and $\mathbf{P}$ is a $N \times N$ diagonal power-allocation matrix with $\sqrt{p_i}$ as its i^{th} diagonal element, where p_i is the power allocated to the i^{th} stream. Of course, the power levels must be chosen so as not to exceed the available transmit power, which can often be conveniently expressed as a constraint on the total normalized transmit power P_t.[4] Under this model, the information-theoretic capacity of the MIMO channel in bps/Hz can be obtained as [3]

$$C_{\text{MIMO}} = \log_2 \det(I + \rho \mathbf{H}\mathbf{V}\mathbf{P}^2\mathbf{V}^{\text{H}}\mathbf{H}^{\text{H}}) \tag{11.4}$$

where $\{\cdot\}^{\text{H}}$ denotes the Hermitian operator for a matrix or vector and ρ is the so-called transmit SNR, given by the ratio of the transmit power over the noise power.

[3] Quadrature Amplitude Modulation.

[4] In practice, there may be a limit on the maximum transmission power from each antenna.

The optimal (capacity-maximizing) precoder (**VP**) in Equation (11.4) is obtained by the concatenation of *singular vector beamforming* and the so-called *waterfilling power allocation*.

Singular vector beamforming means that **V** should be a unitary matrix (i.e. $\mathbf{V}^{\mathrm{H}}\mathbf{V}$ is the identity matrix of size N) chosen such that $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$ is the Singular-Value Decomposition[5] (SVD) of the channel matrix **H**. Thus the i^{th} right singular vector of **H**, given by the i^{th} column of **V**, is used as a transmit beamforming vector for the i^{th} stream. At the receiver side, the optimal beamformer for the i^{th} stream is the i^{th} left singular vector of **H**, obtained as the i^{th} row of $\mathbf{U}^{\mathrm{H}}$:

$$\mathbf{u}_i^{\mathrm{H}}\mathbf{R} = \lambda_i \sqrt{p_i}[x_{i,1}, x_{i,2}, \ldots, x_{i,T}] + \mathbf{u}_i^{\mathrm{H}}\mathbf{N} \tag{11.5}$$

where λ_i is the i^{th} singular value of **H**.

Waterfilling power allocation is the optimal power allocation and is given by

$$p_i = [\mu - 1/(\rho\lambda_i^2)]_+ \tag{11.6}$$

where $[x]_+$ is equal to x if x is positive and zero otherwise. μ is the so-called 'water level', a positive real variable which is set such that the total transmit power constraint is satisfied.

Thus the optimal SU-MIMO multiplexing scheme uses SVD-based transmit and receive beamforming to decompose the MIMO channel into a number of parallel non-interfering subchannels, dubbed 'eigen-channels', each one with an SNR being a function of the corresponding singular value λ_i and chosen power level p_i.

Contrary to what would perhaps be expected, the philosophy of optimal power allocation across the eigen-channels is *not* to equalize the SNRs, but rather to render them more unequal, by 'pouring' more power into the better eigen-channels, while allocating little power (or even none at all) to the weaker ones because they are seen as not contributing enough to the total capacity. This waterfilling principle is illustrated in Figure 11.3.

The underlying information-theoretic assumption here is that the information rate on each stream can be adjusted finely to match the eigen-channel's SNR. In practice, this is done by selecting a suitable Modulation and Coding Scheme (MCS) for each stream.

11.1.3.2 Beamforming with Single Antenna Transmitter or Receiver

In the case where either the receiver or the transmitter is equipped with only a single antenna, the MIMO channel exhibits only one active eigen-channel, and hence multiplexing of more than one data stream is not possible.

In *receive* beamforming, $N = 1$ and $M > 1$ (assuming a single stream). In this case, one symbol is transmitted at a time, such that the symbol-to-transmit-signal mapping function is characterized by $P = T = 1$, and $\mathbf{Y}(\mathbf{X}) = \mathbf{X} = x$, where x is the one QAM symbol to be sent. The received signal vector is given by

$$\mathbf{R} = \mathbf{H}x + \mathbf{N} \tag{11.7}$$

The receiver combines the signals from its M antennas through the use of weights $\mathbf{w} = [w_1, \ldots, w_M]$. Thus the received signal after antenna combining can be expressed as

$$z = \mathbf{w}\mathbf{R} = \mathbf{w}\mathbf{H}s + \mathbf{w}\mathbf{N} \tag{11.8}$$

[5]The reader is referred to [5] for an explanation of generic matrix operations and terminology.

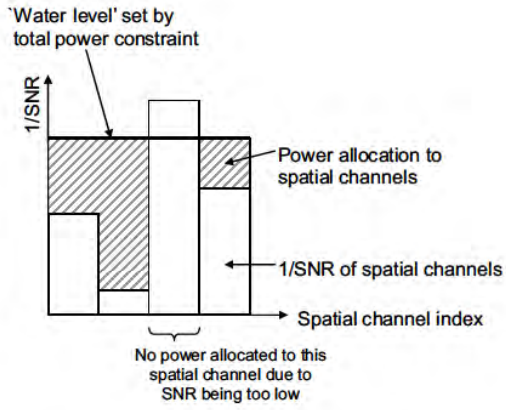


Figure 11.3: The waterfilling principle for optimal power allocation.

After the receiver has acquired a channel estimate (as discussed in Chapter 8), it can set the beamforming vector $\mathbf{w}$ to its optimal value to maximize the received SNR. This is done by aligning the beamforming vector with the UE's channel, via the so-called Maximum Ratio Combining (MRC) $\mathbf{w} = \mathbf{H}^{\mathrm{H}}$, which can be viewed as a spatial version of the well-known matched filter. Note that cancellation of an interfering signal can also be achieved, by selecting the beamforming vector to be orthogonal to the channel from the interference source. These simple concepts are illustrated in Figure 11.4.

The maximum ratio combiner provides a factor of M improvement in the received SNR compared to the $M = N = 1$ case – i.e. an array gain of $10\log_{10}(M)$ dB in the link budget.

In *transmit* beamforming, $M = 1$ and $N > 1$. The symbol-to-transmit-signal mapping function is characterized by $P = T = 1$, and $\mathbf{Y}(\mathbf{X}) = \mathbf{w}x$, where x is the one QAM symbol to be sent and $\mathbf{w}$ is the transmit beamforming vector of size $N \times 1$, computed based on channel knowledge, which is itself often obtained via a receiver-to-transmitter feedback link.[6] Assuming perfect channel knowledge at the transmitter side, the SNR-maximizing solution is given by the transmit MRC, which can be seen as a matched prefilter,

$$\mathbf{w} = \frac{\mathbf{H}^{\mathrm{H}}}{\|\mathbf{H}\|} \tag{11.9}$$

where the normalization by $\|\mathbf{H}\|$ enforces a total power constraint across the transmit antennas. The transmit MRC pre-filter provides a similar gain as its receive counterpart, namely $10\log_{10}(N)$ dB in average SNR improvement.

[6] In some situations, other techniques such as receive/transmit channel reciprocity may be used, as discussed in Section 23.5.2.

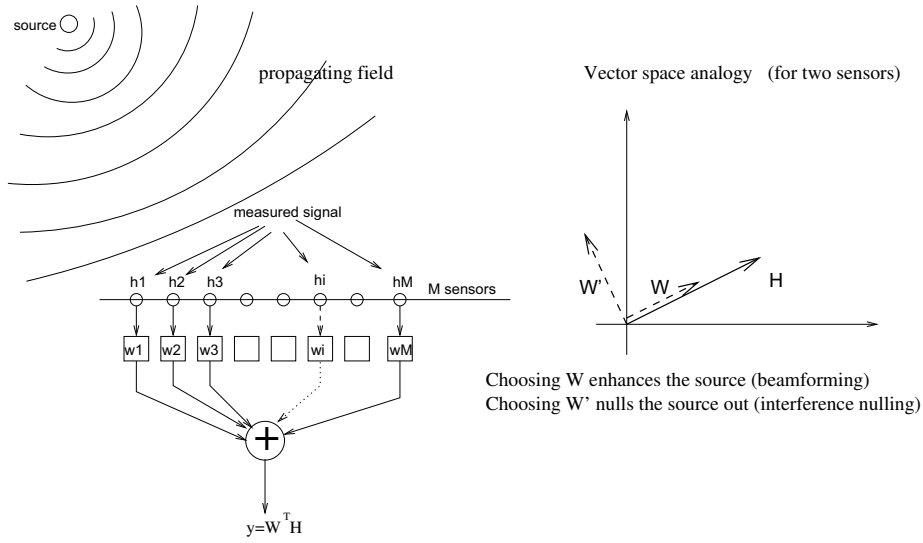


Figure 11.4: The beamforming and interference cancelling concepts.

11.1.3.3 Spatial Multiplexing without Channel Knowledge at the Transmitter

When $N > 1$ and $M > 1$, multiplexing of up to $\min(M, N)$ streams is theoretically possible even without channel knowledge at the transmitter. Assume for instance that $M \geq N$. In this case, one considers N streams, each transmitted using a different transmit antenna. As the transmitter does not have knowledge of the matrix $\mathbf{H}$, the precoder is simply the identity matrix. In this case, the symbol-to-transmit-signal mapping function is characterized by $P = NT$ and by

$$\mathbf{Y}(\mathbf{X}) = \bar{\mathbf{X}} \tag{11.10}$$

At the receiver, a variety of linear and non-linear detection techniques may be implemented to recover the symbol matrix $\bar{\mathbf{X}}$. A low-complexity solution is offered by the linear case, whereby the receiver superposes N beamformers $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N$ so that the i^{th} stream $[x_{i,1}, x_{i,2}, \ldots, x_{i,T}]$ is detected as follows,

$$\mathbf{w}_i\mathbf{R} = \mathbf{w}_i\mathbf{H}\bar{\mathbf{X}} + \mathbf{w}_i\mathbf{N} \tag{11.11}$$

The design criterion for the beamformer $\mathbf{w}_i$ can be interpreted as a compromise between single-stream beamforming and cancelling of interference (created by the other $N - 1$ streams). Inter-stream interference is fully cancelled by selecting the Zero-Forcing (ZF) receiver given by

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_N \end{pmatrix} (\mathbf{H}^{\text{H}}\mathbf{H})^{-1}\mathbf{H}^{\text{H}}. \tag{11.12}$$

However, for optimal performance, $\mathbf{w}_i$ should strike a balance between alignment with respect to $\mathbf{h}_i$ and orthogonality with respect to all other signatures $\mathbf{h}_k$, $k \neq i$. Such a balance is achieved by, for example, a Minimum Mean-Squared Error (MMSE) receiver.

Beyond classical linear detection structures such as the ZF or MMSE receivers, more advanced but non-linear detectors can be exploited which provide a better error rate performance at the chosen SNR operating point, at the cost of extra complexity. Examples of such detectors include the Successive Interference Cancellation (SIC) detector and the Maximum Likelihood Detector (MLD) [3]. The principle of the SIC detector is to treat individual streams, which are channel-encoded, as layers which are 'peeled off' one by one by a processing sequence consisting of linear detection, decoding, remodulating, re-encoding and subtraction from the total received signal **R**. On the other hand, the MLD attempts to select the most likely set of all streams, simultaneously, from **R**, by an exhaustive search procedure or a lower-complexity equivalent such as the sphere-decoding technique [3].

Multiplexing gain

The multiplexing gain corresponds to the multiplicative factor by which the spectral efficiency is increased by a given scheme. Perhaps the single most important requirement for MIMO multiplexing gain to be achieved is for the various transmit and receive antennas to experience a sufficiently different channel response. This translates into the condition that the spatial signatures of the various transmitters (the $\mathbf{h}_i$'s) (or receivers) be sufficiently decorrelated and linearly independent to make the channel matrix **H** invertible (or, more generally, well-conditioned). An immediate consequence of this condition is the limitation to $\min(M, N)$ of the number of independent streams which may be multiplexed into the MIMO channel, or, more generally, to rank(**H**) streams. As an example, single-user MIMO communication between a four-antenna base station and a dual antenna UE can, at best, support multiplexing of two data streams, and thus a doubling of the UE's data rate compared with a single stream.

11.1.3.4 Diversity Techniques

Unlike the basic multiplexing scenario in Equation (11.10), where the design of the transmitted signal matrix **Y** exhibits no redundancy between its entries, a diversity-oriented design will feature some level of repetition between the entries of **Y**. For 'full diversity', each transmitted symbol $x_1, x_2, \ldots, x_P$ must be assigned to each of the transmit antennas at least once in the course of the T symbol durations. The resulting symbol-to-transmit-signal mapping function is called a Space-Time Block Code (STBC). Although many designs of STBC exist, additional properties such as the orthogonality of matrix **Y** allow improved performance and easy decoding at the receiver. Such properties are realized by the Alamouti space-time code [6], explained in Section 11.2.2.1. The total diversity order which can be realized in the N to M MIMO channel is MN when entries of the MIMO channel matrix are statistically uncorrelated. The intuition behind this is that $MN - 1$ represents the number of SISO links simultaneously in a state of severe fading which the system can sustain while still being able to convey the information to the receiver. The diversity order is equal to this number plus one. As in the previous simple multiplexing scheme, an advantage of diversity-oriented transmission is that the transmitter does not need knowledge of the channel **H**, and therefore no feedback of this parameter is necessary.

Diversity versus multiplexing trade-off

A fundamental aspect of the benefits of MIMO lies in the fact that any given multiple antenna configuration has a limited number of degrees of freedom. Thus there exists a compromise between reaching full beamforming gain in the detection of a desired stream of data and the perfect cancelling of undesired, interfering streams. Similarly, there exists a trade-off between the number of streams that may be multiplexed across the MIMO channel and the amount of diversity that each one of them will enjoy. Such a trade-off can be formulated from an information-theoretic point of view [7]. In the particular case of spatial multiplexing of N streams over an N to M antenna channel, with $M \geq N$, and using a linear detector, it can be shown that each stream can enjoy a maximum diversity order of $M - N + 1$.

To some extent, increasing the spatial load of MIMO systems (i.e. the number of spatially multiplexed streams) is akin to increasing the user load in CDMA systems. This correspondence extends to the fact that an optimal load level exists for a given target error rate in both systems.

11.1.4 Multi-User MIMO Techniques

11.1.4.1 Comparing Single-User and Multi-User MIMO

The set of MIMO techniques featuring data streams being communicated to (or from) antennas located on distinct UEs is referred to as Multi-User MIMO (MU-MIMO), which is one of the more recent developments of MIMO technology.

Although the MU-MIMO situation is just as well described by our model in Equation (11.1), the MU-MIMO scenario differs in a number of crucial ways from its single-user counterpart. We first explain these differences qualitatively, and then present a brief survey of the most important MU-MIMO transmission techniques.

In MU-MIMO, K UEs are selected for simultaneous communication over the same time-frequency resource, from a set of U active UEs in the cell. Typically, K is much smaller than U. Each UE is assumed to be equipped with J antennas, so the selected UEs together form a set of $M = KJ$ UE-side antennas. Since the number of streams that may be communicated over an N to M MIMO channel is limited to $\min(M, N)$ (if complete interference suppression is intended using linear combining of the antennas), the upper bound on the number of streams in MU-MIMO is typically dictated by the number of base station antennas N. The number of streams which may be allocated to each UE is limited by the number of antennas J at that UE. For instance, with single-antenna UEs, up to N streams can be multiplexed, thus achieving the maximum multiplexing gain, with a distinct stream being allocated to each UE. This is in contrast to SU-MIMO, where the transmission of N streams necessitates that the UE be equipped with at least N antennas. Therefore a great advantage of MU-MIMO over SU-MIMO is that the MIMO multiplexing benefits are preserved even in the case of low-cost UEs with a small number of antennas. As a result, it is generally assumed that, in MU-MIMO, it is the base station which bears the burden of spatially separating the UEs, be it on the uplink or the downlink. Thus the base station performs receive beamforming from several UEs on the uplink and transmit beamforming towards several UEs on the downlink.

Another fundamental contrast between SU-MIMO and MU-MIMO comes from the difference in the underlying channel model. While in SU-MIMO the decorrelation between the spatial signatures of the antennas requires rich multipath propagation or the use of

orthogonal polarizations, in MU-MIMO the decorrelation between the signatures of the different UEs occurs naturally due to fact that the separation between such UEs is typically large relative to the wavelength.

However, all these benefits of MU-MIMO depend on the level of Channel State Information at the Transmitter (CSIT) that the eNodeB receives from each UE. In the case of SU-MIMO, it has been shown that even a small number of feedback bits per antenna can be very beneficial in steering the transmitted energy accurately towards the UE's antenna(s) [8–11]. More precisely, in SU-MIMO channels, the accuracy of CSIT only causes an SNR offset, but does not affect the slope of the cell throughput-versus-SNR curve (i.e. the multiplexing gain). Yet for the MU-MIMO downlink, the level of Channel State Information (CSI) available at the transmitter does affect the multiplexing gain, because a MU-MIMO system with finite-rate feedback is essentially interference-limited after the crucial interference rejection processing has been carried out by the transmitter. Hence, providing accurate channel feedback is considerably more important for MU-MIMO than for SU-MIMO. On the other hand, in a system with a large number of UEs, the uplink resources required for accurate CSI feedback can become large and must be carefully controlled in the system design.

11.1.4.2 Techniques for Single-Antenna UEs

In considering the case of MU-MIMO for single-antenna UEs, it is worth noting that the number of antennas available to a UE for transmission is typically less than the number available for reception. We therefore examine first the uplink scenario, followed by the downlink.

With a single antenna at each UE, the MU-MIMO uplink scenario is very similar to the one described by Equation (11.10): because the UEs in mobile communication systems such as LTE typically cannot cooperate and do not have knowledge of the uplink channel coefficients, no precoding can be applied and each UE simply transmits an independent message. Thus, if K UEs are selected for transmission in the same time-frequency resource, with each UE k transmitting symbol s_k, the received signal at the base station, over a single $T = 1$ symbol period, is written as

$$\mathbf{R} = \mathbf{H}\bar{\mathbf{X}} + \mathbf{N} \tag{11.13}$$

where $\bar{\mathbf{X}} = (x_1, \ldots, x_K)^T$. In this case, the columns of $\mathbf{H}$ correspond to the receive spatial signatures of the different UEs. The base station can recover the transmitted symbol information by applying beamforming filters, for example using MMSE or ZF solutions. Note that no more than N UEs can be served (i.e. $K \leq N$) if inter-user interference is to be suppressed fully.

MU-MIMO in the uplink is sometimes referred to as 'Virtual MIMO' as, from the point of view of a given UE, there is no knowledge of the simultaneous transmissions of the other UEs. This transmission mode and its implications for LTE are discussed in Section 16.6.2.

On the downlink, which is illustrated in Figure 11.5, the base station must resort to transmit beamforming in order to separate the data streams intended for the various UEs. Over a single $T = 1$ symbol period, the signal received by UEs 1 to K can be written compactly as

$$\mathbf{R} = (r_1, \ldots, r_K)^T = \mathbf{HVP}\bar{\mathbf{X}} + \mathbf{N} \tag{11.14}$$

This time, the *rows* of $\mathbf{H}$ correspond to the transmit spatial signatures of the various UEs. $\mathbf{V}$ is the transmit beamforming matrix and $\mathbf{P}$ is the (diagonal) power allocation matrix selected

such that it fulfils the total normalized transmit power constraint P_t. To cancel out fully the inter-user interference when $K \leq N$, a transmit ZF beamforming solution may be employed (although this is not optimal due to the fact that it may require a high transmit power if the channel is ill-conditioned). Such a solution would be given by $\mathbf{V} = \mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}$.

Note that regardless of the channel realization, the power allocation must be chosen to satisfy any power constraints at the base station, for example such that $\mathrm{trace}(\mathbf{V}\mathbf{P}\mathbf{P}^{\mathrm{H}}\mathbf{V}^{\mathrm{H}}) = P_t$. It is important to note that ZF precoding requires accurate channel knowledge at the transmitter, which in most cases necessitates terminal feedback in the uplink.

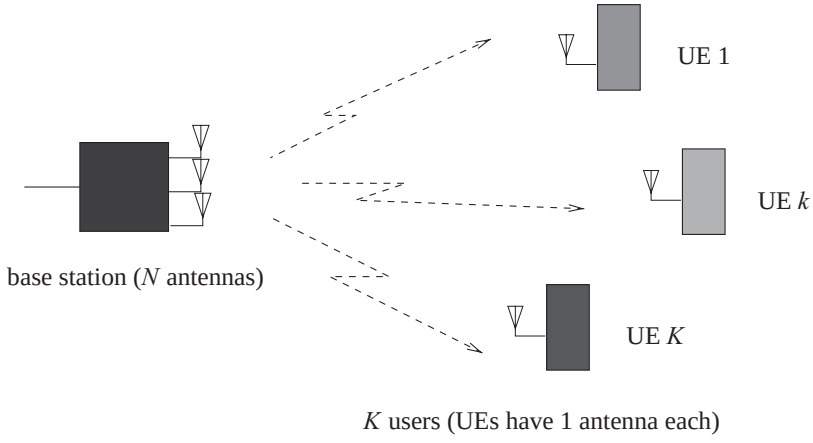


Figure 11.5: An MU-MIMO scenario in the downlink with single-antenna UEs: the base station transmits to K selected UEs simultaneously. Their signals are separated by multiple-antenna precoding at the base station side, based on channel knowledge.

11.1.4.3 Techniques for Multiple-Antenna UEs

The principles presented above for single-antenna UEs can be generalized to the case of multiple-antenna UEs. There could, in theory, be essentially two ways of exploiting the additional antennas at the UE side. In the first approach, the multiple antennas are simply treated as multiple *virtual* UEs, allowing high-capability terminals to receive or transmit more than one stream, while at the same time spatially sharing the channel with other UEs. For instance, a four-antenna base station could theoretically communicate in a MU-MIMO fashion with two UEs equipped with two antennas each, allowing two streams per UE, resulting in a total multiplexing gain of four. Another example would be that of two single-antenna UEs, receiving one stream each, and sharing access with another two-antenna UE, the latter receiving two streams. Again, the overall multiplexing factor remains limited to the number of base-station antennas.

The second approach for making use of additional UE antennas is to treat them as extra degrees of freedom for the purpose of strengthening the link between the UE and the base

station. Multiple antennas at the UE may then be combined in MRC fashion in the case of the downlink, or, in the case of the uplink, space-time coding could be used. Antenna selection is another way of extracting more diversity out of the uplink channel, as discussed in Section 16.6.

11.1.4.4 Comparing Single-User and Multi-User Capacity

To illustrate the gains of multi-user multiplexing over single-user transmission, we compare the sum-rate achieved by both types of system from an information-theoretic standpoint, for single antenna UEs. We compare the Shannon capacity in single-user and multi-user scenarios both for an idealized synthetic channel and for a channel obtained from real measurement data.

The idealized channel model assumes that the entries of the channel matrix $\mathbf{H}$ in Equation (11.13) are independently and identically distributed (i.i.d.) Rayleigh fading. For the measured channel case, a channel sounder[7] was used to perform real-time wideband channel measurements synchronously for two UEs moving at vehicular speed in an outdoor semi-urban hilly environment with Line-Of-Sight (LOS) propagation predominantly present. The most important parameters of the platform are summarized in Table 11.1.

Table 11.1: Parameters of the measured channel for SU-MIMO/MU-MIMO comparison. More details can be found in [12, 13].

Parameter	Value
Centre frequency	1917.6 MHz
Bandwidth	4.8 MHz
Base station transmit power	30 dBm
Number of antennas at base station	4 (2 cross polarized)
Number of UEs	2
Number of antennas at UE	1
Number of subcarriers	160

The sum-rate capacity of a two-UE MU-MIMO system (calculated assuming a ZF precoder as described in Section 11.1.4) is compared with the capacity of an equivalent MISO system serving a single UE at a time (i.e. in TDMA), employing beamforming (see Section 11.1.3.2). The base station has four antennas and the UE has a single antenna. Full CSIT is assumed in both cases.

Figure 11.6 shows the ergodic (i.e. average) sum-rate of both schemes in both channels. The mean is taken over all frames and all subcarriers and subsequently normalized to bps/Hz. It can be seen that in both the ideal and the measured channels, MU-MIMO yields a higher sum-rate than SU-MISO in general. In fact, at high SNR, the multiplexing gain of the MU-MIMO system is two while it is limited to one for the SU-MISO case.

However, for low SNR, the SU-MISO TDMA and MU-MIMO schemes perform very similarly. This is because a sufficiently high SNR is required to excite more than one MIMO

[7] The Eurecom MIMO OpenAir Sounder (EMOS) [12].

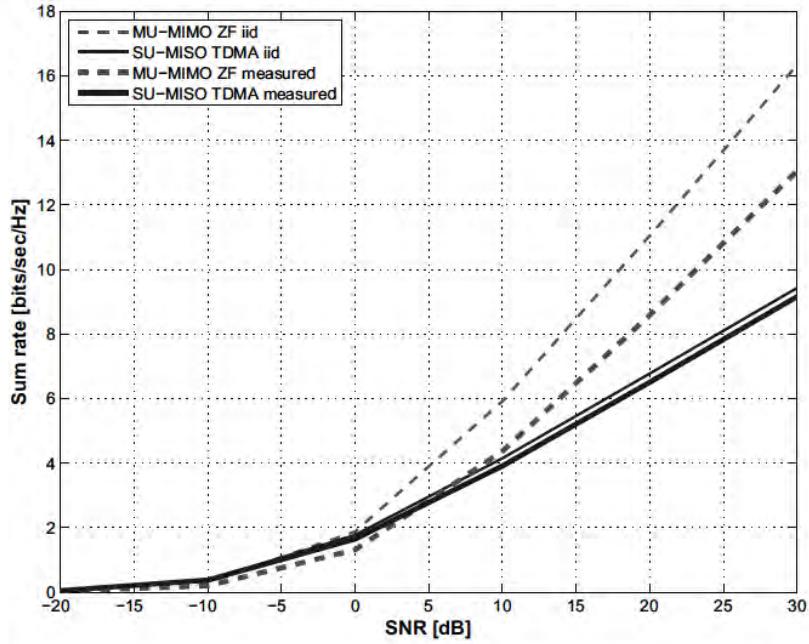


Figure 11.6: Ergodic sum-rate capacity of SU-MISO TDMA and MU-MIMO with two UEs, for an i.i.d. Rayleigh fading channel and for a measured channel.

transmission mode. Interestingly, the performance of both SU-MISO TDMA and MU-MIMO is slightly worse in the measured channels than in the idealized i.i.d. channels. This can be attributed to the correlation of the measured channel in time (due to the relatively slow movement of the users), in frequency (due to the LOS propagation), and in space (due to the transmit antenna correlation). In the MU-MIMO case, the difference between the i.i.d. and the measured channel is much higher than in the single-user TDMA case, since these correlation effects result in a rank-deficient channel matrix.

11.2 MIMO Schemes in LTE

Building on the theoretical background of the previous section, the MIMO schemes adopted for LTE Releases 8 and 9 are now reviewed and explained. These schemes relate to the downlink unless otherwise stated. The MIMO enhancements introduced for LTE-Advanced in Release 10 are explained in Chapter 29.

11.2.1 Practical Considerations

First, a few important practical constraints are briefly reviewed which affect the real-life performance of the theoretical MIMO systems considered above. Such constraints include practical deployment and antenna configurations, propagation conditions, channel knowledge, and implementation complexity. These aspects are often decisive for assessing the performance of a particular transmission strategy in a given propagation and system setting and were fundamental to the selection of MIMO schemes for LTE.

It was argued above that the full MIMO benefits (array gain, diversity gain and multiplexing gain) assume ideally decorrelated antennas and full-rank MIMO channel matrices. In this regard, the propagation environment and the antenna design (e.g. the spacing and polarization) play a significant role. If the antenna separation at the base station and the UE is small relative to the wavelength, strong correlation will be observed between the spatial signatures (especially in a LOS situation), limiting the usefulness of spatial multiplexing. However, this can to some extent be circumvented by means of dual-polarized antennas, whose design itself provides orthogonality even in LOS situations.[8] Beyond such antennas, the condition of spatial signature independence with SU-MIMO can only be satisfied with the help of rich random multipath propagation. In diversity-oriented schemes, the invertibility of the channel matrix is not required, yet the entries of the channel matrix should be statistically decorrelated. Although a greater LOS to non-LOS energy ratio will tend to correlate the fading coefficients on the various antennas, this effect will be counteracted by the reduction in fading delivered by the LOS component.

Another source of discrepancy between theoretical MIMO gains and practically achieved performance lies in the (in-)ability of the receiver and, whenever needed, the transmitter, to acquire reliable knowledge of the propagation channel. At the receiver, channel estimation is typically performed over a finite sample of Reference Signals (RSs), as discussed in Chapter 8. In the case of transmit beamforming and MIMO precoding, the transmitter then has to acquire this channel knowledge from the receiver usually through a limited feedback link, which causes further degradation to the available CSIT.

The potential advantages of MU-MIMO over SU-MIMO include robustness with respect to the propagation environment and the preservation of spatial multiplexing gain even in the case of UEs with small numbers of antennas. However, these advantages rely on the ability of the base station to compute the required transmit beamformer, which requires accurate CSIT; if no CSIT is available and the fading statistics are identical for all the UEs, the MU-MIMO gains disappear and the SU-MIMO strategy becomes optimal [14]. As a consequence, one of the most difficult challenges in making MU-MIMO practical for cellular applications, and particularly for an FDD system, is devising feedback mechanisms that allow for accurate CSI to be delivered efficiently by the UEs to the base station. This requires the use of appropriate *codebooks* for quantization (see Section 11.2.2.3). A recent account of the literature on this subject may be found in [15].

Another issue which arises for practical implementations of MIMO schemes is the interaction between the physical layer and the scheduling protocol. As noted in Section 11.1.4.1, in both uplink and downlink cases the number of UEs which can be served in a MU-MIMO fashion is typically limited to $K = N$, assuming linear combining structures. Often

[8] Horizontal and vertical polarizations, or, better, +45° and −45° polarizations, give a twofold multiplexing capability even in LOS.

one may even decide to limit K to a value strictly less than N to preserve some degrees of freedom for per-user diversity. As the number of active users U will typically exceed K, a selection algorithm must be implemented to identify which set of users will be scheduled for simultaneous transmission over a particular time-frequency Resource Block (RB). This algorithm is not specified in LTE and various approaches are possible; as discussed in Chapter 12, a combination of rate maximization and QoS constraints will typically be considered. It is important to note that the choice of UEs that will maximize the sum-rate (the sum over the K individual rates for a given subframe) is one that favours UEs exhibiting not only good instantaneous SNR but also spatial separability among their signatures.

11.2.2 Single-User Schemes

In this section, we examine the downlink multi-antenna schemes adopted in LTE for individual UEs. We consider first the diversity schemes used on the transmit side, then the single-user spatial multiplexing transmission modes, and finally beamforming using precoded UE-specific RSs (including the combination of beamforming and spatial multiplexing as introduced in Release 9).

11.2.2.1 Transmit Diversity Schemes

The theoretical aspects of transmit diversity were discussed in Section 11.1. Here we discuss the two main transmit diversity techniques defined in LTE. In LTE, transmit diversity is only defined for 2 and 4 transmit antennas and one data stream, referred to in LTE as one *codeword* since one transport block CRC[9] is used per data stream. To maximize diversity gain, the antennas typically need to be uncorrelated, so they need to be well separated relative to the wavelength or have different polarization. Transmit diversity is valuable in a number of scenarios, including low SNR, or for applications with low delay tolerance. Diversity schemes are also desirable for channels for which no uplink feedback signalling is available (e.g. Multimedia Broadcast/Multicast Services (MBMS) described in Chapter 13, Physical Broadcast CHannel (PBCH) in Chapter 9 and Synchronization Signals in Chapter 7), or when the feedback is not sufficiently accurate (e.g. at high speeds); transmit diversity is therefore also used as a fallback scheme in LTE when transmission with other schemes fails.

In LTE the multiple-antenna scheme is independently configured for the control channels and the data channels, and in the case of the data channels (Physical Downlink Shared CHannel – PDSCH) it is assigned independently per UE.

In this section we discuss in detail the transmit diversity techniques of Space-Frequency Block Codes (SFBCs) and Frequency-Switched Transmit Diversity (FSTD), as well as the combination of these schemes as used in LTE. These transmit diversity schemes may be used in LTE for the PBCH and Physical Downlink Control CHannel (PDCCH), and also for the PDSCH if it is configured in transmit diversity mode[10] for a UE.

Another transmit diversity technique which is commonly associated with OFDM is Cyclic Delay Diversity (CDD). CDD is not used in LTE as a diversity scheme in its own right but

[9] Cyclic Redundancy Check.

[10] PDSCH transmission mode 2 – see Section 9.2.2.1. As noted above, transmit diversity may also be used as a fallback in other transmission modes.

rather as a precoding scheme for spatial multiplexing on the PDSCH; we therefore introduce it in Section 11.2.2.2 in the context of spatial multiplexing.

Space-Frequency Block Codes (SFBCs)

If a physical channel in LTE is configured for transmit diversity operation using two eNodeB antennas, pure SFBC is used. SFBC is a frequency-domain version of the well-known Space-Time Block Codes (STBCs), also known as Alamouti codes [6]. This family of codes is designed so that the transmitted diversity streams are orthogonal and achieve the optimal SNR with a linear receiver. Such orthogonal codes only exist for the case of two transmit antennas.

STBC is used in UMTS, but in LTE the number of available OFDM symbols in a subframe is often odd while STBC operates on pairs of adjacent symbols in the time domain. The application of STBC is therefore not straightforward for LTE, while the multiple subcarriers of OFDM lend themselves well to the application of SFBC.

For SFBC transmission in LTE, the symbols transmitted from the two eNodeB antenna ports on each pair of adjacent subcarriers are defined as follows:

$$\begin{bmatrix} y^{(0)}(1) & y^{(0)}(2) \\ y^{(1)}(1) & y^{(1)}(2) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \\ -x_2^* & x_1^* \end{bmatrix} \tag{11.15}$$

where $y^{(p)}(k)$ denotes the symbols transmitted from antenna port p on the k^{th} subcarrier.

Since no orthogonal codes exist for antenna configurations beyond 2×2, SFBC has to be modified in order to apply it to the case of 4 transmit antennas. In LTE, this is achieved by combining SFBC with FSTD.

Frequency Switched Transmit Diversity (FSTD) and its Combination with SFBC

General FSTD schemes transmit symbols from each antenna on a different set of subcarriers. In LTE, FSTD is only used in combination with SFBC for the case of 4 transmit antennas, in order to provide a suitable transmit diversity scheme where no orthogonal rate-1 block code exists. The LTE scheme is in fact a combination of two 2×2 SFBC schemes mapped to independent subcarriers as follows:

$$\begin{bmatrix} y^{(0)}(1) & y^{(0)}(2) & y^{(0)}(3) & y^{(0)}(4) \\ y^{(1)}(1) & y^{(1)}(2) & y^{(1)}(3) & y^{(1)}(4) \\ y^{(2)}(1) & y^{(2)}(2) & y^{(2)}(3) & y^{(2)}(4) \\ y^{(3)}(1) & y^{(3)}(2) & y^{(3)}(3) & y^{(3)}(4) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_3 & x_4 \\ -x_2^* & x_1^* & 0 & 0 \\ 0 & 0 & -x_4^* & x_3^* \end{bmatrix} \tag{11.16}$$

where, as previously, $y^{(p)}(k)$ denotes the symbols transmitted from antenna port p on the k^{th} subcarrier. Note that the mapping of symbols to antenna ports is different in the 4 transmit-antenna case compared to the 2 transmit-antenna SFBC scheme. This is because the density of cell-specific RSs on the third and fourth antenna ports is half that of the first and second antenna ports (see Section 8.2.1), and hence the channel estimation accuracy may be lower on the third and fourth antenna ports. Thus, this design of the transmit diversity scheme avoids concentrating the channel estimation losses in just one of the SFBC codes, resulting in a slight coding gain.

11.2.2.2 Spatial Multiplexing Schemes

We begin by introducing some terminology used to describe spatial multiplexing in LTE:

- A spatial *layer* is the term used in LTE for one of the different streams generated by spatial multiplexing as described in Section 11.1. A layer can be described as a mapping of symbols onto the transmit antenna ports.[11] Each layer is identified by a precoding vector of size equal to the number of transmit antenna ports and can be associated with a radiation pattern.
- The *rank* of the transmission is the number of layers transmitted.[12]
- A *codeword* is an independently encoded data block, corresponding to a single Transport Block (TB) delivered from the Medium Access Control (MAC) layer in the transmitter to the physical layer, and protected with a CRC.

For ranks greater than 1, two codewords can be transmitted. Note that the number of codewords is always less than or equal to the number of layers, which in turn is always less than or equal to the number of antenna ports.

In principle, a SU-MIMO spatial multiplexing scheme can either use a single codeword mapped to all the available layers, or multiple codewords each mapped to one or more different layers. The main benefit of using only one codeword is a reduction in the amount of control signalling required, both for Channel Quality Indicator (CQI) reporting, where only a single value would be needed for all layers, and for HARQ ACK/NACK[13] feedback, where only one ACK/NACK would have to be signalled per subframe per UE. In such a case, the MLD receiver is optimal in terms of minimizing the BER. At the opposite extreme, a separate codeword could be mapped to each of the layers. The advantage of such a mapping is that significant gains are possible by using SIC, albeit at the expense of more signalling being required. An MMSE-SIC receiver can be shown to approach the Shannon capacity [7]. Note that an MMSE receiver is viable for both transmitter structures. For LTE, a middle-way was adopted whereby at most two codewords are used, even if four layers are transmitted. The codeword-to-layer mapping is static, since only minimal gains were shown for a dynamic mapping method. The mappings are shown in Table 11.2. Note that in LTE all RBs belonging to the same codeword use the same MCS, even if a codeword is mapped to multiple layers.

Precoding

The PDSCH transmission modes for open-loop spatial multiplexing[14] and closed-loop spatial multiplexing[15] use precoding from a defined ‘codebook’ to form the transmitted layers. Each codebook consists of a set of predefined precoding matrices, with the size of the set being a trade-off between the number of signalling bits required to indicate a particular matrix in the codebook and the suitability of the resulting transmitted beam direction.

[11] See Section 8.2 for an explanation of the concept of an antenna port.

[12] Note that the rank of a matrix is the number of linearly independent rows or columns, or equivalently the dimension of the subspace generated by the rows or columns of the matrix. Hence the rank of the channel matrix may be higher than the number of transmitted layers.

[13] Hybrid Automatic Repeat reQuest ACKnowledgement/Negative ACKnowledgement.

[14] PDSCH transmission mode 3 – see Section 9.2.2.1.

[15] PDSCH transmission modes 4 and 6 – see Section 9.2.2.1.

Table 11.2: Codeword-to-layer mapping in LTE.

Transmission rank	Codeword 1	Codeword 2
Rank 1	Layer 1	
Rank 2	Layer 1	Layer 2
Rank 3	Layer 1	Layer 2 and Layer 3
Rank 4	Layer 1 and Layer 2	Layer 3 and Layer 4

In the case of closed-loop spatial multiplexing, a UE feeds back to the eNodeB the index of the most desirable entry from a predefined codebook. The preferred precoder is the matrix which would maximize the capacity based on the receiver capabilities. In a single-cell, interference-free environment the UE will typically indicate the precoder that would result in a transmission with an effective SNR following most closely the largest singular values of its estimated channel matrix.

Some important properties of the LTE codebooks are as follows:

- **Constant modulus property.** LTE uses precoders which mostly comprise pure phase corrections – that is, with no amplitude changes. This is to ensure that the Power Amplifier (PA) connected to each antenna is loaded equally. The one exception (which still maintains the constant modulus property) is using an identity matrix as the precoder. However, although the identity precoder may completely switch off one antenna on one layer, since each layer is still connected to one antenna at constant power the net effect across the layers is still constant modulus to the PA.
- **Nested property.** The nested property is a method of arranging the codebooks of different ranks so that the lower rank codebook is comprised of a subset of the higher rank codebook vectors. This property simplifies the CQI calculation across different ranks. It ensures that the precoded transmission for a lower rank is a subset of the precoded transmission for a higher rank, thereby reducing the number of calculations required for the UE to generate the feedback. For example, if a specific index in the codebook corresponds to columns 1, 2 and 3 from the precoder $\mathbf{W}$ in the case of a rank 3 transmission, then the same index in the case of rank 2 transmission must consist of either columns 1 and 2 or columns 1 and 3 from $\mathbf{W}$.
- **Minimal 'complex' multiplications.** The 2-antenna codebook consists entirely of a QPSK[16] alphabet, which eliminates the need for any complex multiplications since all codebook multiplications use only ± 1 and $\pm j$. The 4-antenna codebook does contain some QPSK entries which require a $\sqrt{2}$ magnitude scaling as well; it was considered that the performance gain of including these precoders justified the added complexity.

The 2-antenna codebook in LTE comprises one 2×2 identity matrix and two DFT (Discrete Fourier Transform) matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix} \tag{11.17}$$

[16] Quadrature Phase Shift Keying.

Here the columns of the matrices correspond to the layers.

The 4 transmit antenna codebook in LTE uses a Householder generating function:

$$\mathbf{W}_{\mathrm{H}} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^{\mathrm{H}}/\mathbf{u}^{\mathrm{H}}\mathbf{u}, \tag{11.18}$$

which generates unitary matrices from input vectors $\mathbf{u}$, which are defined in [16, Table 6.3.4.2.3-2]. The advantage of using precoders generated with this equation is that it simplifies the CQI calculation (which has to be carried out for each individual precoder in order to determine the preferred precoder) by reducing the number of matrix inversions. This structure also reduces the amount of control signalling required, since the optimum rank-1 version of $\mathbf{W}_{\mathrm{H}}$ is always the first column of $\mathbf{W}_{\mathrm{H}}$; therefore the UE only needs to indicate the preferred precoding matrix, and not also the individual vector within it which would be optimal for the case of rank-1 transmission.

Note that both DFT and Householder matrices are unitary.

Cyclic Delay Diversity (CDD)

In the case of *open*-loop spatial multiplexing,[17] the feedback from the UE indicates only the *rank* of the channel, and not a preferred precoding matrix. In this mode, if the rank used for PDSCH transmission is greater than 1 (i.e. more than one layer is transmitted), LTE uses Cyclic Delay Diversity (CDD) [17]. CDD involves transmitting the same set of OFDM symbols on the same set of OFDM subcarriers from multiple transmit antennas, with a different delay on each antenna. The delay is applied before the Cyclic Prefix (CP) is added, thereby guaranteeing that the delay is cyclic over the Fast Fourier Transform (FFT) size. This gives CDD its name.

Adding a time delay is identical to applying a phase shift in the frequency domain. As the same time delay is applied to all subcarriers, the phase shift will increase linearly across the subcarriers with increasing subcarrier frequency. Each subcarrier will therefore experience a different beamforming pattern as the non-delayed subcarrier from one antenna interferes constructively or destructively with the delayed version from another antenna. The diversity effect of CDD therefore arises from the fact that different subcarriers will pick out different spatial paths in the propagation channel, thus increasing the frequency-selectivity of the channel. The channel coding, which is applied to a whole transport block across the subcarriers, ensures that the whole transport block benefits from the diversity of spatial paths.

Although this approach does not optimally exploit the channel in the way that ideal precoding would (by matching the precoding to the eigenvectors of the channel), it does help to ensure that any destructive fading is constrained to individual subcarriers rather than affecting a whole transport block. This can be particularly beneficial if the channel information at the transmitter is unreliable, for example due to the feedback being limited or the UE velocity being high.

The general principle of the CDD technique is illustrated in Figure 11.7. The fact that the delay is added before the CP means that any delay value can be used without increasing the overall delay spread of the channel. By contrast, if the delay had been added after the addition of the CP, then the usable delays would have had to be kept small in order to ensure that the delay spread of the delayed symbol is no more than the maximum channel delay spread, in order to obviate any need to increase the CP length.

[17] PDSCH transmission mode 3 – see Section 9.2.2.1.

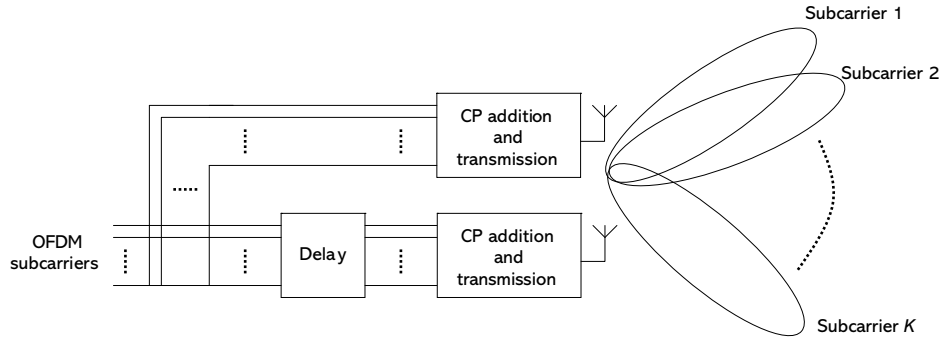


Figure 11.7: Principle of Cyclic Delay Diversity.

The time-delay/phase-shift equivalence means that the CDD operation can be implemented as a frequency-domain precoder for the affected antenna(s), where the precoder phase changes on a per-subcarrier basis according to a fixed linear function. In general, the implementation designer can choose whether to implement CDD in the time domain or the frequency domain. However, one advantage of a frequency-domain implementation is that it is not limited to delays corresponding to an integer number of samples.

As an example of CDD for a case with 2 transmit antenna ports, we can express mathematically the received symbol r_k on the k^{th} subcarrier as

$$r_k = h_{1k}x_k + h_{2k}e^{j\phi k}x_k \tag{11.19}$$

where h_{pk} is the channel from the p^{th} transmit antenna and $e^{j\phi k}$ is the phase shift on the k^{th} subcarrier due to the delay operation. We can see clearly that on some subcarriers the symbols from the second transmit antenna will add constructively, while on other subcarriers they will add destructively. Here, $\phi = 2\pi d_{\text{cdd}}/N$, where N is the FFT size and d_{cdd} is the delay in samples.

The number of peaks and troughs created by CDD in the received signal spectrum across the subcarriers therefore depends on the length of the delay d_{cdd}: as d_{cdd} is increased, the number of peaks and troughs in the spectrum also increases.

In LTE, the frequency-domain phase shift values ϕ are π, $2\pi/3$ and $\pi/2$ for 2-, 3- and 4-layer transmission respectively. For a size-2048 FFT, for example, these values correspond to $d_{\text{cdd}} = 1024,\ 682.7,\ 512$ sample delays respectively.

For the application of CDD in LTE, the eNodeB transmitter combines CDD delay-based phase shifts with additional precoding using fixed unitary DFT-based precoding matrices.

The application of precoding in this way is useful when the channel coefficients of the antenna ports are correlated, since then *virtual* antennas (formed by fixed, non-channel-dependent precoding) will typically be uncorrelated. On its own, the benefit of CDD is reduced by antenna correlation. The use of uncorrelated virtual antennas created by fixed precoding can avoid this problem in correlated channels, while not degrading the performance if the individual antenna ports are uncorrelated.

For ease of explanation, the above discussion of CDD has been in terms of a rank-1 transmission – i.e. with a single layer. However, in practice, CDD is only applied in LTE when the rank used for PDSCH transmission is greater than 1. In such a case, each layer benefits independently from CDD in the same way as for a single layer. For example, for a rank-2 transmission, the transmission on the second antenna port is delayed relative to the first antenna port for each layer. This means that symbols transmitted on both layers will experience the delay and hence the increased frequency selectivity.

For multilayer CDD operation, the mapping of the layers to antenna ports is carried out using precoding matrices selected from the spatial multiplexing codebooks described earlier. As the UE does not indicate a preferred precoding matrix in the open loop spatial multiplexing transmission mode in which CDD is used, the particular spatial multiplexing matrices selected from the spatial multiplexing codebooks in this case are predetermined.

In the case of 2 transmit antenna ports, the predetermined spatial multiplexing precoding matrix **W** is always the same (the first entry in the 2 transmit antenna port codebook, which is the identity matrix). Thus, the transmitted signal can be expressed as follows:

$$\begin{bmatrix} y^{(0)}(k) \\ y^{(1)}(k) \end{bmatrix} = \mathbf{W}\mathbf{D}_2\mathbf{U}_2\mathbf{x} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ 0 & e^{j\phi_1 k} \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} x^{(0)}(i) \\ x^{(1)}(i) \end{bmatrix} \tag{11.20}$$

In the case of 4 transmit antenna ports, ν different precoding matrices are used from the 4 transmit antenna port codebook where ν is the transmission rank. These ν precoding matrices are applied in turn across groups of ν subcarriers in order to provide additional decorrelation between the spatial streams.

11.2.2.3 Beamforming Schemes

The theoretical aspects of beamforming were described in Section 11.1. This section explains how it is implemented in LTE Release 8 and Release 9.

In LTE, two beamforming techniques are supported for the PDSCH:

- **Closed-loop rank-1 precoding.**[18] This mode amounts to beamforming since only a single layer is transmitted, exploiting the gain of the antenna array. However, it can also be seen as a special case of the SU-MIMO spatial multiplexing using codebook based precoding discussed in Section 11.2.2.2. In this mode, the UE feeds channel state information back to the eNodeB to indicate suitable precoding to apply for the beamforming operation. This feedback information is known as Precoding Matrix Indicators (PMIs) and is described in detail in Section 11.2.2.4.

- **Beamforming with UE-specific RSs.**[19] In this case, the MIMO precoding is not restricted to a predefined codebook, so the UE cannot use the cell-specific RS for PDSCH demodulation and precoded UE-specific RS are therefore needed.

In this section we focus on the latter case which supports a single transmitted layer in LTE Release 8 and was extended to support dual-layer beamforming in Release 9. These modes are primarily mechanisms to extend cell coverage by concentrating the eNodeB power in the directions in which the radio channel offers the strongest path(s) to reach the UE. They are

[18] PDSCH transmission mode 6 – see Section 9.2.2.1.

[19] PDSCH transmission mode 7 (from Release 8) and 8 (introduced in Release 9) – see Section 9.2.2.1.

typically implemented with an array of closely spaced antenna elements (or pairs of cross-polarized elements) for creating directional transmissions. Due to the fact that no precoding codebook is used, an arbitrary number of transmit antennas is theoretically supported; in practical deployments up to 8 antennas are typically used (e.g. an array of 4 cross-polarized pairs). The signals from the correlated antenna elements are phased appropriately so that they add up constructively at the location where the UE is situated. The UE is not really aware that it is receiving a precoded directional transmission rather than a cell-wide transmission (other than by the fact that the UE is directed to use the UE-specific RS as the phase reference for demodulation). To the UE, the phased array of antennas 'appears' as just one antenna port.

In LTE Release 8, only a single UE-specific antenna port and associated RSs is defined (see Section 8.2.2). As a consequence the beamformed PDSCH can only be transmitted with a single spatial layer. The support of two UE-specific antenna ports was introduced in Release 9 (see Section 8.2.3), thus enabling beamforming with two spatial layers with a direct one-to-one mapping between the layers and antenna ports. The UE-specific RS patterns for the two Release 9 antenna ports were designed to be orthogonal in order to facilitate separation of the spatial layers at the receiver. The two associated PDSCH layers can then be transmitted to a single user in good propagation conditions to increase its data rate, or to multiple users (MU-MIMO) to increase the system capacity. LTE Release 10 further extends the support for beamforming with UE-specific RSs to higher numbers of spatial layers, as described in Chapter 29.

When two layers are transmitted over superposed beams they share the available transmit power of the eNodeB, so the increase in data rate comes at the expense of a reduction of coverage. This makes fast rank adaptation (which is also used in the codebook-based spatial multiplexing transmission mode) an important mechanism for the dual-layer beamforming mode. To support this, it is possible to configure the UE to feed back a Rank Indicator (RI) together with the PMI (see Section 11.2.2.4).

It should also be noted that beamforming in LTE can only be applied to the PDSCH and not to the downlink control channels, so although the range of a given data rate on the PDSCH can be extended by beamforming, the overall cell coverage may still be limited by the range of the control channels unless other measures are taken. The trade-off between throughput and coverage with beamforming in LTE is illustrated in Figure 11.8.

Like any precoding technique, these beamforming modes require reliable CSIT which can, in principle, be obtained either from feedback from the UE or from estimation of the uplink channel.

In the single-layer beamforming mode of Release 8, the UE does not feed back any precoding-related information, and the eNodeB deduces this information from the uplink, for example using Direction Of Arrival (DOA) estimations. It is worth noting that in this case calibration of the eNodeB RF paths may be necessary, as discussed in Section 23.5.2.

The dual-layer beamforming mode of Release 9 provides the possibility for the UE to help the eNodeB to derive the beamforming precoding weights by sending PMI feedback in the same way as for codebook-based closed-loop spatial multiplexing. For beamforming, this is particularly suitable for FDD deployments where channel reciprocity cannot be exploited as effectively as in the case of TDD. Nevertheless, this PMI feedback can only provide partial CSIT, since it is based on measurements of the common RSs which, in beamforming deployments, are usually transmitted from multiple antenna elements with broad beam patterns; furthermore, the PMI feedback uses the same codebook as for closed-loop spatial

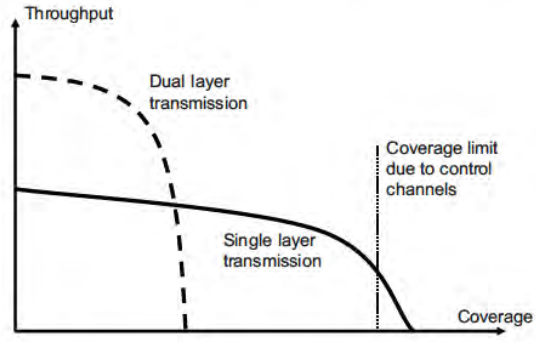


Figure 11.8: Trade-off between throughput and coverage with beamforming.

multiplexing and is therefore constrained by quantization while the beamforming precoding is not. As a consequence, additional uplink channel estimates are still required if the full benefits of beamforming are to be realized.

For the Channel Quality Indicator (CQI) feedback for link adaptation, the eNodeB must take into account the fact that in the beamforming modes the channel quality experienced by the UE on the beamformed PDSCH resources will typically be different (hopefully better) than on the cell-specific RSs from which the CQI is derived. The UE-specifc RSs cannot be used for CQI estimation because they are only present in the RBs in which beamforming transmission is used; UE-specific RSs that are precoded for a different UE are obviously not useful for channel estimation. The eNodeB has to adjust the MCS used for transmission relative to the CQI reports by estimating the expected beamforming gain depending on the number of transmitted layers; the HARQ acknowledgement rate can be useful in deriving such an adjustment.

It is worth mentioning that beamforming also affects the interference caused to neighbouring cells. If the beamforming is intermittent, it can result in a problem often known as the 'flash-light effect', where strong intermittent interference may disturb the accuracy of the UEs' CQI reporting in adjacent cells. This effect was shown to have the potential to cause throughput reductions in HSDPA [18]. In LTE the possibility for frequency-domain scheduling (as discussed in Section 12.5) provides an additional degree of freedom to avoid such issues. In addition, beamforming may even be coordinated to take advantage of the spatial interference structure created by beamforming in multiple cells, especially among the cells controlled by one eNodeB.

11.2.2.4 Feedback Computation and Signalling

In order to support MIMO operation. the UE can be configured to report Precoding Matrix Indicators (PMIs) and Rank Indicators(RIs), in addition to CQI reporting as discussed in Section 10.2.1.

In the case of codebook-based spatial multiplexing transmissions, the precoding at the eNodeB transmitter is applied relative to the transmitted phase of the cell-specific RSs for each antenna port. Thus if the UE knows the precoding matrices that could be applicable (as defined in the configured codebook) and it knows the transfer function of the channels from the different antenna ports (by making measurements on the RSs), it can determine which **W** is most suitable under the current radio conditions and signal this to the eNodeB. The preferred **W**, whose index constitutes the PMI report, is the precoder that maximizes the aggregate number of data bits which could be received across all layers. The number of RBs to which each PMI corresponds in the frequency domain depends on the feedback mode configured by the eNodeB:

- **Wideband PMI.** The UE reports a single PMI corresponding to the preferred precoder assuming transmission over the whole system bandwidth. This is applicable for PMI feedback that is configured to be sent periodically (on the PUCCH[20] unless the PUSCH[21] is transmitted), and also for PMI feedback that is sent aperiodically in conjunction with UE-selected sub-band CQI reports on the PUSCH (see Section 10.2.1).
- **Sub-band PMI.** The UE reports one PMI for each sub-band across the whole system bandwidth, where the sub-band size is between 4 and 6 RBs depending on the system bandwidth (as shown in Table 10.3). This is applicable for PMI feedback that is sent aperiodically on the PUSCH in conjunction with a wideband CQI report. It is well suited to a scenario where the eNodeB wishes to schedule a wideband data transmission to a UE while the channel direction varies across the bandwidth; the precoder used by the eNodeB can change from sub-band to sub-band within one transport block.
- **UE-selected sub-band PMI.** The UE selects a set of M preferred sub-bands, each of size k RBs (where M and k are the same as for UE-selected sub-band CQI feedback as given in Table 10.4) and reports a single PMI corresponding to the preferred precoder assuming transmission over all of the M selected sub-bands. This is applicable in conjunction with UE-selected sub-band CQI reports on the PUSCH.

The eNodeB is not bound to use the precoder requested by the UE, but clearly if the eNodeB chooses another precoder then the eNodeB will have to adjust the MCS accordingly since the reported CQI could not be assumed to be applicable.

The eNodeB may also restrict the set of precoders which the UE may evaluate and report. This is known as *codebook subset restriction*. It enables the eNodeB to prevent the UE from reporting precoders which are not useful, for example in some eNodeB antenna configurations or in correlated fading scenarios. In the case of open-loop spatial multiplexing, codebook subset restriction amounts simply to a restriction on the rank which the UE may report.

For each PDSCH transmission to a UE, the eNodeB indicates via a 'Transmitted Precoding Matrix Indicator' (TPMI) in the downlink assignment message[22] on the PDCCH whether it is applying the UE's preferred precoder, and if not, which precoder is used. This enables the UE to derive the correct phase reference relative to the cell-specific RSs in order to

[20] Physical Uplink Control CHannel.
[21] Physical Uplink Shared CHannel.
[22] The TPMI is sent in DCI Formats 2, 2A and 2B – see Section 9.3.5.1.

demodulate the PDSCH data. The ability to indicate that the eNodeB is applying the UE's preferred precoder is particularly useful in the case of sub-band PMI reporting, as it enables multiple precoders to be used across the bandwidth of a wideband transmission to a UE without incurring a high signalling overhead to notify the UE of the set of precoders used.

The UE can also be configured to report the channel rank via a RI, which is calculated to maximize the capacity over the entire bandwidth, jointly selecting the preferred precoder(s) to maximize the capacity on the assumption of the selected transmission rank. The CQI values reported by the UE correspond to the preferred transmission rank and precoders, to enable the eNodeB to perform link adaptation and multi-user scheduling as discussed in Sections 10.2 and 12.1. The number of CQI values reported normally corresponds to the number of codewords supported by the preferred transmission rank. Further, the CQI values themselves will depend on the assumed transmission rank: for example, the precoding matrix for layer 1 will usually be different depending on whether or not the UE is assuming the presence of a second layer.

Although the UE indicates the transmission rank that would maximize the downlink data rate, the eNodeB can also indicate to the UE that a different transmission rank is being used for a PDSCH transmission. This gives flexibility to the eNodeB, since the UE does not know the amount of data in the downlink buffer; if the amount of data in the buffer is small, the eNodeB may prefer to use a lower rank with higher reliability.

11.2.3 Multi-User MIMO

The main focus of MIMO in the first release of LTE was on achieving transmit antenna diversity or single-user multiplexing gain, neither of which requires such a sophisticated CSI feedback mechanism as MU-MIMO. As pointed out in Section 11.2.1, the availability of accurate CSIT is the main challenge in making MU-MIMO schemes attractive for cellular applications.

Nevertheless, basic support for MU-MIMO was provided in Release 8, and enhanced schemes were added in Releases 9 and 10.

The MU-MIMO scheme supported in LTE Release 8[23] is based on the same codebook-based scheme as is defined for SU-MIMO (see Section 11.2.2.2). In particular the same 'implicit' feedback calculation mechanism is used, whereby the UE makes hypotheses on the precoder the eNodeB would use for transmission on the PDSCH and reports a recommended precoding matrix (PMI) corresponding to the best hypothesis – i.e. an *implicit* representation of the CSI. The recommended precoder typically depends on the type of receiver implemented by the UE, e.g. MRC, linear MMSE or MMSE Decision Feedback Equalizer (DFE). However, in the same way as for SU-MIMO, the eNodeB is not bound to the UE's recommendation and can override the PMI report by choosing a different precoder from the codebook and signalling it back to the scheduled UEs using a TPMI on the PDCCH.

In the light of the similarities between implicit channel feedback by precoder recommendation and explicit channel vector quantization, it is instructive to consider the implications of the fact that the LTE 2-antenna codebook and the first eight entries in the 4-antenna codebook are derived from a DFT matrix (as noted in Section 11.2.2.2). The main reasons for this choice are as follows:

[23] PDSCH transmission mode 5 – see Section 9.2.2.1.

- Simplicity in the PMI/CQI calculation, which can be done in some cases without complex multiplications;
- A DFT matrix nicely captures the characteristics of a highly correlated MISO channel.[24]

In LTE Release 8, a UE configured in MU-MIMO mode assumes that an eNodeB transmission on the PDSCH would be performed on one layer using one of the rank-1 codebook entries defined for two or four antenna ports. Therefore, the MU-MIMO transmission is limited to a single layer and single codeword per scheduled UE, and the RI and PMI feedback are the same as for SU-MIMO rank 1.

The eNodeB can then co-schedule multiple UEs on the same RB, typically by pairing UEs that report orthogonal PMIs. The eNodeB can therefore signal to those UEs (using one bit in the DCI format 1D message on the PDCCH – see Section 9.3.5.1) that the PDSCH power is reduced by 3 dB relative to the cell-specific RSs. This assumes that the transmission power of the eNodeB is equally divided between the two codewords and implies that in practice a maximum of two UEs may be co-scheduled in Release 8 MU-MIMO mode.

Apart from this power offset signalling, MU-MIMO in Release 8 is fully transparent to the scheduled UEs, in that they are not explicitly aware of how many other UEs are co-scheduled or of which precoding vectors are used for any co-scheduled UEs.

CQI reporting for MU-MIMO in Release 8 does not take into account any interference from transmissions to co-scheduled users. The CQI values reported are therefore equivalent to those that would be reported for rank-1 SU-MIMO. It is left to the eNodeB to estimate a suitable adjustment to apply to the MCS if multiple UEs are co-scheduled.

In Release 9, the support for MU-MIMO is enhanced by the introduction of dual-layer beamforming in conjunction with precoded UE-specific RSs, as discussed in Sections 8.2.3 and 11.2.2.3. This removes the constraint on the eNodeB to use the feedback codebook for precoding, and gives flexibility to utilize other approaches to MU-MIMO user separation, for example according to a zero-forcing beamforming criterion. Note that in the limit of a large UE population, a zero-forcing beamforming solution converges to a unitary precoder because, with high probability, there will be N UEs with good channel conditions reporting orthogonal channel signatures.

The use of precoded UE-specific RSs removes the need to signal explicitly the details of the chosen precoder to the UEs. Instead, the UEs estimate the effective channel (i.e. the combination of the precoding matrix and physical channel), from which they can derive the optimal antenna combiner for their receivers.

The UE-specific RS structure described in Section 8.2.3 allows up to 4 UEs to be scheduled for MU-MIMO transmissions from Release 9.

Furthermore, the precoded UE-specific RSs mean that for Release 9 MU-MIMO it is not necessary to signal explicitly a power offset depending on the number of co-scheduled UEs. As the UE-specific RSs for two UEs are code-division multiplexed, the eNodeB transmission power is shared between the two antenna ports in the same way as the transmitted data, so the UE can use the UE-specific RSs as both the phase reference and the amplitude reference for demodulation. Thus the MU-MIMO scheduling is fully transparent in Release 9.

[24] It is not difficult to see that if the first N rows are taken from a DFT matrix of size N_q, each of the N_q column vectors comprises the spatial signature of the i^{th} element of a uniform linear antenna array with spacing l, such that $\frac{l}{\lambda}\sin\theta = \frac{i}{N_q}$ [7].

11.2.4 MIMO Performance

The multiple antenna schemes supported by LTE contribute much to its spectral efficiency improvements compared to UMTS.

Figure 11.9 shows a typical link throughput performance comparison between the open-loop SU-MIMO spatial multiplexing mode of LTE Release 8 with rank feedback but no PMI feedback, and closed-loop SU-MIMO spatial multiplexing with PMI feedback. Four transmit and two receive antennas are used. Both modes make use of spatial multiplexing gain to enhance the peak data rate. Closed-loop MIMO provides gain from the channel-dependent precoding, while open-loop MIMO offers additional robustness by SFBC for single-layer transmissions and CDD for more than one layer.

The gain of closed-loop precoding can be observed in low mobility scenarios, while the benefit of open-loop diversity can clearly be seen in the case of high mobility.

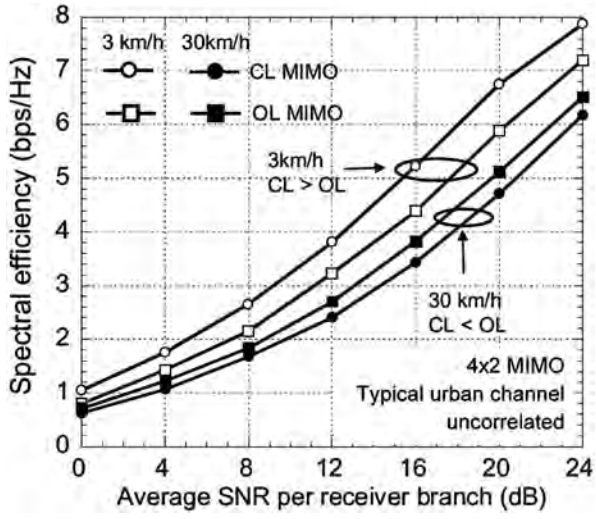


Figure 11.9: Example of closed-loop and open-loop MIMO performance.

11.3 Summary

In this chapter we have reviewed the predominant families of MIMO techniques (both single-user and multi-user) of relevance to LTE. LTE breaks new ground in drawing on such approaches to harness the power of MIMO systems not just for boosting the peak per-user data rate but also for improving overall system capacity and spectral efficiency. Single-User MIMO techniques are well-developed in Release 8, providing the possibility to benefit from precoding while avoiding a high control signalling overhead. Support for beamforming and Multi-User MIMO is significantly enhanced in Release 9. LTE-Advanced further develops and enhances these techniques, as outlined in Chapter 29.

References[25]

[1] D. Gesbert, M. Shafi, P. Smith, D. Shiu and A. Naguib, 'From Theory to Practice: An Overview of MIMO Space-time Coded Wireless Systems'. *IEEE Journal on Selected Areas in Communications*. Special Issue on MIMO systems, guest edited by the authors, April 2003.

[2] H. Bölcskei, D. Gesbert, C. Papadias and A. J. van der Veen (eds), *Space-Time Wireless Systems: From Array Processing to MIMO Communications*. Cambridge: Cambridge University Press, 2006.

[3] A. Goldsmith, *Wireless Communications*. Cambridge: Cambridge University Press, 2005.

[4] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone and W. Yu, 'Multi-Cell MIMO Cooperative Networks: A New Look at Interference', to appear in *IEEE Journal on Selected Areas in Communications*, December 2010.

[5] G. Golub and C. van Loan, *Matrix Computations*. Balitmore, MA: John Hopkins University Press, 1996.

[6] S. M. Alamouti, 'A Simple Transmitter Diversity Technique for Wireless Communications'. *IEEE Journal on Selected Areas in Communications*, Vol. 16, pp. 1451–1458, October 1998.

[7] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge: Cambridge University Press, 2004.

[8] A. Narula, M. J. Lopez, M. D. Trott and G. W. Wornell, 'Efficient Use of Side Information in Multiple Antenna Data Transmission over Fading Channels'. *IEEE Journal on Selected Areas in Communications*, Vol. 16, pp. 1223–1436, October 1998.

[9] D. Love, R. Heath Jr and T. Strohmer, 'Grassmannian Beamforming for Multiple-input Multiple-output Wireless Systems'. *IEEE Trans. on Information Theory*, Vol. 49, pp. 2735–2747, October 2003.

[10] D. Love, R. Heath Jr, W. Santipach, and M. Honig, 'What is the Value of Limited Feedback for MIMO Channels?'. *IEEE Communication Magazine*, Vol. 42, pp. 54–59, October 2003.

[11] N. Jindal, 'MIMO Broadcast Channels with Finite Rate Feedback'. *IEEE Trans. Information Theory*, Vol. 52, pp. 5045–5059, November 2006.

[12] F. Kaltenberger, L. S. Cardoso, M. Kountouris, R. Knopp and D. Gesbert, 'Capacity of Linear Multi-user MIMO Precoding Schemes with Measured Channel Data', in *Proc. IEEE Workshop on Sign. Proc. Adv. in Wireless Comm. (SPAWC)*, Recife, Brazil, July 2008.

[13] F. Kaltenberger, M. Kountouris, D. Gesbert and R. Knopp, 'Correlation and Capacity of Measured Multi-user MIMO Channels', in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Cannes, France, September 2008.

[14] Caire, G. and Shamai, S. (Shitz), 'On the Achievable Throughput of a Multi-antenna Gaussian Broadcast Channel'. *IEEE Trans. on Information Theory*, Vol. 49, pp. 1691–1706, July 2003.

[15] R. Heath, V. Lau, D. Love, D. Gesbert, B. Rao and A. Andrews, (eds), 'Exploiting Limited Feedback in Tomorrow's Wireless Communications Networks'. Special issue of *IEEE Journal on Selected Areas in Communications*, October 2008.

[16] 3GPP Technical Specification 36.211, 'Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation', www.3gpp.org.

[17] Y. Li, J. Chuang and N. Sollenberger, 'Transmitter Diversity for OFDM Systems and Its Impact on High-Rate Data Wireless Networks'. *IEEE Journal on Selected Areas in Communications*, Vol. 17, pp. 1233–1243, July 1999.

[18] A. Osseiran and A. Logothetis, 'Closed Loop Transmit Diversity in WCDMA HS-DSCH', in *Proc. IEEE Vehicular Technology Conference*, May 2005.

[25] All web sites confirmed 1st March 2011.

12

Multi-User Scheduling and Interference Coordination

Issam Toufik and Raymond Knopp

12.1 Introduction

The eNodeB in an LTE system is responsible, among other functions, for managing resource scheduling for both uplink and downlink channels. The ultimate aim of this function is typically to fulfil the expectations of as many users of the system as possible, taking into account the Quality of Service (QoS) requirements of their respective applications.

A typical single-cell cellular radio system is shown in Figure 12.1, comprising K User Equipments (UEs) communicating with one eNodeB over a fixed total bandwidth B. In the uplink, each UE has several data queues corresponding to different uplink logical channel groups, each with different delay and rate constraints. In the same way, in the downlink the eNodeB may maintain several buffers per UE containing dedicated data traffic, in addition to queues for broadcast services. The different traffic queues in the eNodeB would typically have different QoS constraints. We further assume for the purposes of this analysis that only one user is allocated a particular Resource Block (RB) in any subframe, although in practice multi-user MIMO schemes may result in more than one UE per RB, as discussed in Sections 11.2.3 and 16.6.2 for the downlink and uplink respectively.

The goal of a resource scheduling algorithm in the eNodeB is to allocate the RBs and transmission powers for each subframe in order to optimize a function of a set of performance metrics, for example maximum/minimum/average throughput, maximum/minimum/average delay, total/per-user spectral efficiency or outage probability. In the downlink the resource allocation strategy is constrained by the total transmission power of the eNodeB, while in the uplink the main constraints on transmission power in different RBs arises from a multicell view of inter-cell interference and the power headroom of the UEs.

In this chapter we provide an overview of some of the key families of scheduling algorithms which are relevant for an LTE system and highlight some of the factors that an eNodeB can advantageously take into account.

LTE – The UMTS Long Term Evolution: From Theory to Practice, Second Edition.
Stefania Sesia, Issam Toufik and Matthew Baker.

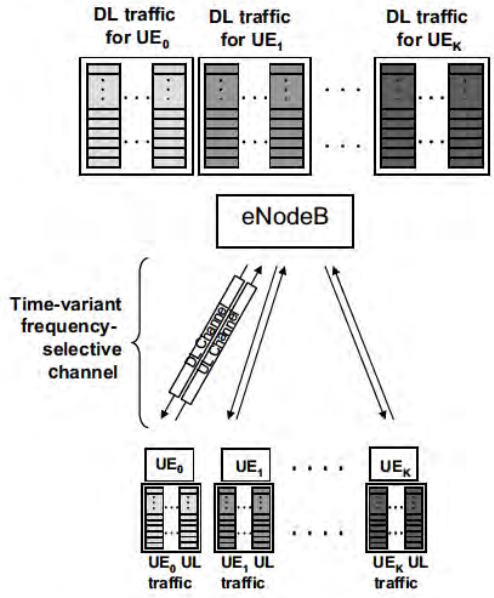


Figure 12.1: A typical single-cell cellular radio system.

12.2 General Considerations for Resource Allocation Strategies

The generic function of a resource scheduler, as shown for the downlink case in Figure 12.2, is to schedule data to a set of UEs on a shared set of physical resources.

It is worth noting that the algorithm used by the resource scheduler is also tightly coupled with the Adaptive Modulation and Coding (AMC) scheme and the retransmission protocol (Hybrid Automatic Repeat reQuest, HARQ, see Sections 4.4 and 10.3.2.5). This is due firstly to the fact that, in addition to dynamic physical resource allocation, the channel measurements are also used to adapt the Modulation and Coding Scheme (MCS) (i.e. transmission spectral-efficiency) as explained in Section 10.2. Secondly, the queue dynamics, which impact the throughput and delay characteristics of the link seen by the application, depend heavily on the HARQ protocol and transport block sizes. Moreover, the combination of channel coding and retransmissions provided by HARQ enables the spectral efficiency of an individual transmission in one subframe to be traded off against the number of subframes in which retransmissions take place. Well-designed practical scheduling algorithms will necessarily consider all these aspects.

In general, scheduling algorithms can make use of two types of measurement information to inform the scheduling decisions, namely Channel State Information (CSI) and traffic measurements (volume and priority). These are obtained either by direct measurements at

the eNodeB, or via feedback signalling channels, or a combination of both. The amount of feedback used is an important consideration, since the availability of accurate CSI and traffic information helps to maximize the data rate in one direction at the expense of more overhead in the other. This fundamental trade-off, which is common to all feedback-based resource scheduling schemes, is particularly important in Frequency Division Duplex (FDD) operation where uplink-downlink reciprocity of the radio channels cannot be assumed (see Chapter 20). For Time Division Duplex (TDD) systems, coherence between the uplink and downlink channels may be used to assist the scheduling algorithm, as discussed in Section 23.5.

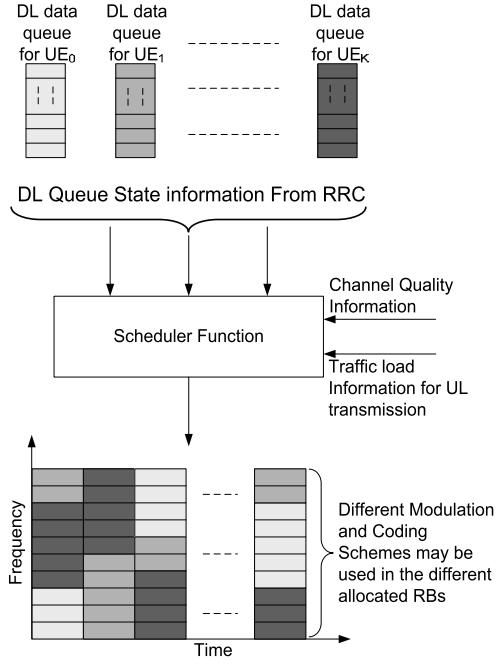


Figure 12.2: Generic view of a wideband resource scheduler.

Based on the available measurement information, the eNodeB resource scheduler must manage the differing requirements of all the UEs in the cells under its control to ensure that sufficient radio transmission resources are allocated to each UE within acceptable latencies to meet their QoS requirements in a spectrally efficient way. The details of this process are not standardized as it is largely internal to the eNodeB, allowing for vendor-specific algorithms to be developed which can be optimized for specific scenarios in conjunction with network operators. However, the key inputs available to the resource scheduling process are common.

In general, two extremes of scheduling algorithm may be identified: *opportunistic scheduling* and *fair scheduling*. The former is typically designed to maximize the sum of the transmitted data rates to all users by exploiting the fact that different users experience different channel gains and hence will experience good channel conditions at different times and frequencies. A fundamental characteristic of mobile radio channels is the fading effects

arising from the mobility of the UEs in a multipath propagation environment, and from variations in the surrounding environment itself (see Chapter 20). In [1–3] it is shown that, for a multi-user system, significantly more information can be transmitted across a fading channel than a non-fading channel for the same average signal power at the receiver. This principle is known as *multi-user diversity*. With proper dynamic scheduling, allocating the channel at each given time instant only to the user with the best channel condition in a particular part of the spectrum can yield a considerable increase in the total throughput as the number of active users becomes large.

The main issue arising from opportunistic resource allocation schemes is the difficulty of ensuring fairness and the required QoS. Users' data cannot always wait until the channel conditions are sufficiently favourable for transmission, especially in slowly varying channels. Furthermore, as explained in Chapter 1, it is important that network operators can provide reliable wide area coverage, including to stationary users near the cell edge – not just to the users which happen to experience good channel conditions by virtue of their proximity to the eNodeB.

The second extreme of scheduling algorithm, fair scheduling, therefore pays more attention to latency and achieving a minimum data rate for each user than to the total data rate achieved. This is particularly important for real-time applications such as Voice-over-IP (VoIP) or video-conferencing, where a certain minimum rate must be guaranteed independently of the channel state.

In practice, most scheduling algorithms fall between the two extremes outlined above, including elements of both to deliver the required mix of QoS. A variety of metrics can be used to quantify the degree of fairness provided by a scheduling algorithm, the general objective being to avoid heavily penalizing the cell-edge users in an attempt to give high throughputs to the users with good channel conditions. One example is based on the Cumulative Distribution Function (CDF) of the throughput of all users, whereby a system may be considered sufficiently fair if the CDF of the throughput lies to the right-hand side of a particular line, such as that shown in Figure 12.3. Another example is the *Jain index* [4], which gives an indication of the variation in throughput between users. It is calculated as:

$$J = \frac{\overline{T}^2}{\overline{T}^2 + \mathrm{var}(T)}, \quad 0 \le J \le 1, \tag{12.1}$$

where $\overline{T}$ and $\mathrm{var}(T)$ are the mean and variance respectively of the average user throughputs. The more similar the average user throughputs ($\mathrm{var}(T) \rightarrow 0$), the higher the value of J.

Other factors also need to be taken into account, especially the fact that, in a coordinated deployment, individual cells cannot be considered in isolation – nor even the individual set of cells controlled by a single eNodeB. The eNodeBs should take into account the interference generated by co-channel cells, which can be a severe limiting factor, especially for cell-edge users. Similarly, the performance of the system as a whole can be enhanced if each eNodeB also takes into account the impact of the transmissions of its own cells on the neighbouring cells. These inter-cell aspects, and the corresponding inter-eNodeB signalling mechanisms provided in LTE, are discussed in detail in Section 12.5.

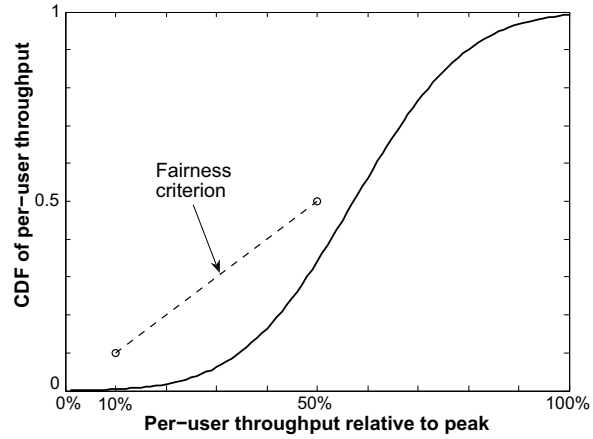


Figure 12.3: An example of a metric based on the throughput CDF over all users for scheduler fairness evaluation (10–50 metric).

12.3 Scheduling Algorithms

Multi-user scheduling finds its basis at the interface between information theory and queueing theory, in the theory of capacity-maximizing resource allocation. Before establishing an algorithm, a capacity-related metric is first formulated and then optimized across all possible resource allocation solutions satisfying a set of predetermined constraints. Such constraints may be physical (e.g. bandwidth and total power) or QoS-related.

Information theory offers a range of possible capacity metrics which are relevant in different system operation scenarios. Two prominent examples are explained here, namely the so-called *ergodic capacity* and *delay-limited capacity*, corresponding respectively to soft and hard forms of rate guarantee for the user.

12.3.1 Ergodic Capacity

The ergodic capacity (also known as the Shannon capacity) is defined as the maximum data rate which can be sent over the channel with asymptotically small error probability, averaged over the fading process. When CSI is available at the Transmitter (i.e. CSIT), the transmit power and mutual information[1]between the transmitter and the receiver can be varied depending on the fading state in order to maximize the average rates. The ergodic capacity metric considers the long-term average data rate which can be delivered to a user when the user does not have any latency constraints.

[1]The mutual information bewteen two random variables X and Y with probability density function $p(X)$ and $p(Y)$ respectively is defined as $I(X;Y) = \mathbb{E}_{X,Y}\,[\log\, p(X;Y)/(p(X)p(Y))]$, where $\mathbb{E}[.]$ is the expectation function [5].

12.3.1.1 Maximum Rate Scheduling

It has been shown in [3, 6] that the maximum total ergodic sum rate $\sum_{k=1}^{K} R_k$, where R_k is the total rate allocated to user k, is achieved with a transmission power given by

$$P_k(m, f) = \begin{cases} \left[\frac{1}{\lambda_k} - \frac{P_N}{|H_k(m, f)|^2}\right]^+ & \text{if } |H_k(m, f)|^2 \geq \frac{\lambda_k}{\lambda_{k'}}|H_{k'}(m, f)|^2 \\ 0 & \text{otherwise} \end{cases} \tag{12.2}$$

where $[x]^+ = \max(0, x)$, $H_k(m, f)$ is the channel gain of user k in RB m of subframe f and λ_k are constants which are chosen in order to satisfy an average per-user power constraint.

This approach is known in the literature as *maximum sum rate* scheduling. The result in Equation (12.2) shows that the maximum sum rate is achieved by orthogonal multiplexing where in each subchannel (i.e. each RB in LTE) only the user with the best channel gain is scheduled. This orthogonal scheduling property is in line with, and thus justifies, the philosophy of the LTE multiple-access schemes. The input power spectra given by Equation (12.2) are water-filling formulae in both frequency and time (i.e. allocating more power to a scheduled user when his channel gain is high and less power when it is low).[2]

A variant of this resource allocation strategy with no power control (relevant for cases where the power control dynamic range is limited or zero) is considered in [2]. This allocation strategy is called 'maximum-rate constant-power' scheduling, where only the user with the best channel gain is scheduled in each RB, but with no adaptation of the transmit power. It is shown in [2] that most of the performance gains offered by the maximum rate allocation in Equation (12.2) are due to multi-user diversity and not to power control, so an on-off power allocation can achieve comparable performance to maximum sum rate scheduling. This not only allows some simplification of the scheduling algorithm but also is well suited to a downlink scenario where the frequency-domain dynamic range of the transmitted power in a given subframe is limited by constraints such as the dynamic range of the UE receivers and the need to transmit wideband reference signals for channel estimation.

12.3.1.2 Proportional Fair Scheduling

The ergodic sum rate corresponds to the optimal rate for traffic which has no delay constraint. This results in an unfair sharing of the channel resources. When the QoS required by the application includes latency constraints, such a scheduling strategy is not suitable and other fairer approaches need to be considered. One such approach is the well-known Proportional Fair Scheduling (PFS) algorithm. PFS schedules a user when its instantaneous channel quality is high relative to its own average channel condition over time. It allocates user $\hat{k}_m$ in RB m in any given subframe f if [7]

$$\hat{k}_m = \underset{k=1,\ldots,K}{\operatorname{argmax}} \frac{R_k(m, f)}{T_k(f)} \tag{12.3}$$

where $T_k(f)$ denotes the long-term average throughput of user k computed in subframe f and $R_k(m, f) = \log(1 + \mathrm{SNR}_k(m, f))$ is the achievable rate by user k in RB m and subframe f.

[2] Water-filling strategies are discussed in more detail in the context of MIMO in Section 11.1.3.

The long-term average user throughputs are recursively computed by

$$T_k(f) = \left(1 - \frac{1}{t_c}\right) T_k(f-1) + \frac{1}{t_c} \sum_{m=1}^{M} R_k(m, f) \mathcal{I}\{\hat{k}_m = k\} \tag{12.4}$$

where t_c is the time window over which fairness is imposed and $\mathcal{I}\{\cdot\}$ is the indicator function equal to one if $\hat{k}_m = k$ and zero otherwise. A large time window t_c tends to maximize the total average throughput; in fact, in the limit of a very long time window, PFS and maximum-rate constant-power scheduling result in the same allocation of resources. For small t_c, the PFS tends towards a round-robin[3] scheduling of users [7].

Several studies of PFS have been conducted in the case of WCDMA/HSDPA[4], yielding insights which can be applied to LTE. In [8] the link level system performance of PFS is studied, taking into account issues such as link adaptation dynamic range, power and code resources, convergence settings, signalling overhead and code multiplexing. [9] analyses the performance of PFS in the case of VoIP, including a comparison with round-robin scheduling for different delay budgets and packet-scheduling settings. A study of PFS performance in the case of video streaming in HSDPA can be found in [10].

12.3.2 Delay-Limited Capacity

Even though PFS introduces some fairness into the system, this form of fairness may not be sufficient for applications which have a very tight latency constraint. For these cases, a different capacity metric is needed; one such example is referred to as the 'delay-limited capacity'. The delay-limited capacity (also known as zero-outage capacity) is defined as the transmission rate which can be guaranteed in all fading states under finite long-term power constraints. In contrast to the ergodic capacity, where mutual information between the transmitter and the receiver vary with the channel, the powers in delay-limited capacity are coordinated between users and RBs with the objective of maintaining constant mutual information independently of fading states. The delay-limited capacity is relevant to traffic classes where a given data rate must be guaranteed throughout the connection time, regardless of the fading dips.

The delay-limited capacity for a flat-fading multiple-access channel is characterized in [11]. The wideband case is analysed in [7, 12].

It is shown in [7] that guaranteeing a delay-limited rate incurs only a small throughput loss in high Signal to Noise Ratio (SNR) conditions, provided that the number of users is large. However, the solution to achieve the delay-limited capacity requires non-orthogonal scheduling of the users, with successive decoding in each RB. This makes this approach basically unsuitable for LTE.

It is, however, possible to combine orthogonal multiple access with hard QoS requirements. Examples of algorithms achieving such a compromise can be found in [12, 13]. It can be shown that even under hard fairness constraints it is possible to achieve performance which is very close to the optimal unfair policy; thus hard fairness constraints do not necessarily introduce a significant throughput degradation, even with orthogonal resource allocation, provided that the number of users and the system bandwidth are large. This may be a relevant

[3] A round-robin approach schedules each user in turn, irrespective of their prevailing channel conditions.
[4] Wideband Code Division Multiple Access / High Speed Downlink Packet Access.

scenario for a deployment targeting VoIP users, where densities of several hundred users per cell are foreseen (as mentioned in Section 1.2), with a latency constraint typically requiring each packet to be successfully delivered within 50 ms.

In general this discussion of scheduling strategies assumes that all users have an equal and infinite queue length – often referred to as a full-buffer traffic model. In practice this is not the case, especially for real-time services, and information on users' queue lengths is necessary to guarantee system stability. If a scheduling algorithm can be found which keeps the average queue length bounded, the system is said to be stabilized [14]. One approach to ensuring that the queue length remains bounded is to use the queue length to set the priority order in the allocation of RBs. This generally works for lightly loaded systems. In [15] and [16] it is shown that in wideband frequency-selective channels, low average packet delay can be achieved even if the fading is very slow.

12.4 Considerations for Resource Scheduling in LTE

In LTE, each logical channel has a corresponding QoS description which should influence the behaviour of the eNodeB resource scheduling algorithm. Based on the evolution of the radio and traffic conditions, this QoS description could potentially be updated for each service in a long-term fashion. The mapping between the QoS descriptions of different services and the resource scheduling algorithm in the eNodeB can be a key differentiating factor between radio network equipment manufacturers.

An important constraint for the eNodeB scheduling algorithm is the accuracy of the eNodeB's knowledge of the channel quality for the active UEs in the cell. The manner in which such information is provided to the scheduler in LTE differs between uplink and downlink transmissions. In practice, for the downlink this information is provided through the feedback of Channel Quality Indicators (CQIs)[5] by UEs as described in Chapter 10, while for the uplink the eNodeB may use Sounding Reference Signals (SRSs) or other signals transmitted by the UEs to estimate the channel quality, as discussed in Chapter 15. The frequency with which CQI reports and SRS are transmitted is configurable by the eNodeB, allowing for a trade-off between the signalling overhead and the availability of up-to-date channel information. If the most recent CQI report or SRS was received a significant time before the scheduling decision is taken, the performance of the scheduling algorithm can be significantly degraded.

In order to perform frequency-domain scheduling, the information about the radio channel needs to be frequency-specific. For this purpose, the eNodeB may configure the CQI reports to relate to specific sub-bands to assist the downlink scheduling, as explained in Section 10.2.1. Uplink frequency-domain scheduling can be facilitated by configuring the SRS to be transmitted over a large bandwidth. However, for cell-edge UEs the wider the transmitted bandwidth the lower the available power per RB; this means that accurate frequency-domain scheduling may be more difficult for UEs near the cell edge. Limiting the SRS to a subset of the system bandwidth will improve the channel quality estimation on these RBs but restrict the ability of the scheduler to find an optimal scheduling solution for all users. In general, provided that the bandwidth over which the channel can be estimated for

[5] More generally, in the context of user selection for MU-MIMO, full CSI is relevant, to support user pairing as well as rate scheduling.

scheduling purposes is greater than the intended scheduling bandwidth for data transmission by a sufficient factor, a useful element of multi-user diversity gain may still be achievable.

As noted above, in order to support QoS- and queue-aware scheduling, it is necessary for the scheduler to have not only information about the channel quality, but also information on the queue status. In the LTE downlink, knowledge of the amount of buffered data awaiting transmission to each UE is inherently available in the eNodeB; for the uplink, Section 4.4.2.2 explains the buffer status reporting mechanisms available to transfer such information to the eNodeB.

12.5 Interference Coordination and Frequency Reuse

One limiting aspect for system throughput performance in cellular networks is inter-cell interference, especially for cell-edge users. Careful management of inter-cell interference is particularly important in systems such as LTE which are designed to operate with a frequency reuse factor of one.

The scheduling strategy of the eNodeB may therefore include an element of Inter-Cell Interference Coordination (ICIC), whereby interference from and to the adjacent cells is taken into account in order to increase the data rates which can be provided for users at the cell edge. This implies for example imposing restrictions on what resources in time and/or frequency are available to the scheduler, or what transmit power may be used in certain time/frequency resources.

The impact of interference on the achievable data rate for a given user can be expressed analytically. If a user k is experiencing no interference, then its achievable rate in an RB m of subframe f can be expressed as

$$R_{k,\text{no-Int}}(m,f) = W\log\left(1+\frac{P^s(m,f)|H_k^s(m,f)|^2}{P_N}\right) \tag{12.5}$$

where $H_k^s(m,f)$ is the channel gain from the serving cell s to user k, $P^s(m,f)$ is the transmit power from cell s, P_N is the noise power and W is the bandwidth of one RB (i.e. 180 kHz). If neighbouring cells are transmitting in the same time-frequency resources, then the achievable rate for user k reduces to

$$R_{k,\text{Int}}(m,f) = W\log\left(1+\frac{P^s(m,f)|H_k^s(m,f)|^2}{P_N+\sum_{i\neq s}P^i(m,f)|H_k^i(m,f)|^2}\right) \tag{12.6}$$

where the indices i denote interfering cells.

The rate loss for user k can then be expressed as

$$\begin{aligned} R_{k,\text{loss}}(m,f) &= R_{k,\text{no-Int}}(m,f) - R_{k,\text{Int}}(m,f) \\ &= W\log\left(\frac{1+\text{SNR}}{1+\left[\frac{1}{\text{SNR}}+\frac{\sum_{i\neq s}P^i(m,f)|H_k^i(m,f)|^2}{P^s(m,f)|H_k^s(m,f)|^2}\right]^{-1}}\right) \end{aligned} \tag{12.7}$$

Figure 12.4 plots the rate loss for user k as a function of the total inter-cell interference to signal ratio $\alpha = \left(\sum_{i\neq s}P^i(m,f)|H_k^i(m,f)|^2\right)/\left(P^s(m,f)|H_k^s(m,f)|^2\right)$, with SNR = 0 dB. It can

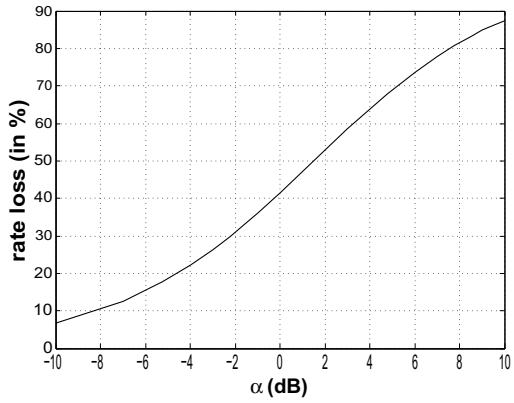


Figure 12.4: User's rate loss due to interference.

easily be seen that for a level of interference equal to the desired signal level (i.e. $\alpha \approx 0$ dB), user k experiences a rate loss of approximately 40%.

In order to demonstrate further the significance of interference and power allocation depending on the system configuration we consider two examples of a cellular system with two cells (s_1 and s_2) and one active user per cell (k_1 and k_2 respectively). Each user receives the wanted signal from its serving cell, while the inter-cell interference comes from the other cell.

In the first example, each user is located near its respective eNodeB (see Figure 12.5(a)). The channel gain from the interfering cell is small compared to the channel gain from the serving cell ($|H_{k_1}^{s_1}(m,f)| \gg |H_{k_1}^{s_2}(m,f)|$ and $|H_{k_2}^{s_2}(m,f)| \gg |H_{k_2}^{s_1}(m,f)|$). In the second example (see Figure 12.5(b)), we consider the same scenario but with the users now located close to the edge of their respective cells. In this case the channel gain from the serving cell and the interfering cell are comparable ($|H_{k_1}^{s_1}(m,f)| \approx |H_{k_1}^{s_2}(m,f)|$ and $|H_{k_2}^{s_2}(m,f)| \approx |H_{k_2}^{s_1}(m,f)|$).

The capacity of the system with two eNodeBs and two users can be written as

$$R_{\text{Tot}} = W\left(\log\left(1 + \frac{P^{s_1}|H_{k_1}^{s_1}(m,f)|^2}{P_N + P^{s_2}|H_{k_1}^{s_2}(m,f)|^2}\right) + \log\left(1 + \frac{P^{s_2}|H_{k_2}^{s_2}(m,f)|^2}{P_N + P^{s_1}|H_{k_2}^{s_1}(m,f)|^2}\right)\right) \tag{12.8}$$

From this equation, it can be noted that the optimal transmit power operating point in terms of maximum achievable throughput is different for the two considered cases. In the first scenario, the maximum throughput is achieved when both eNodeBs transmit at maximum power, while in the second the maximum capacity is reached by allowing only one eNodeB to transmit. It can in fact be shown that the optimal power allocation for maximum capacity for this situation with two base stations is binary in the general case; this means that either both base stations should be operating at maximum power in a given RB, or one of them should be turned off completely in that RB [17].

From a practical point of view, this result can be exploited in the eNodeB scheduler by treating users in different ways depending on whether they are cell-centre or cell-edge users. Each cell can then be divided into two parts – inner and outer. In the inner part, where users experience a low level of interference and also require less power to communicate with the