University “Politehnica” of Bucharest

Faculty of Electronics, Telecommunications and Information Technology

Analysis of data transfer in LTE networks

Diploma Thesis

submitted in partial fulfillment of the requirements for the Degree of Engineer in the domain *Electronics and Telecommunications,* study program *Technologies and Communications Systems*

Thesis Advisor(s) Student

Ş.l. Dr. Ing. Şerban OBREJA *Albert Constantin IEPURE*

Dr. Ing. Oana BĂDIŢĂ

2013

University “Politehnica” of Bucharest
Faculty of Electronics, Telecommunications and Information Technology
Department of Telecommunications

Department Director's Approval:

Prof. Dr. Ing. Silviu CIOCHINĂ

DIPLOMA THESIS
of student Iepure S.I. Albert Constantin, 442 Ga

1. Thesis title: **Analysis of data transfer in LTE networks**

2. The student's original contribution will consist of (not including the documentation part): evaluating LTE network performance by using LTE System Level Simulator. The KPIs to be analyzed include the cell throughput and BLER, user throughput and BLER and the CQI, while varying several parameters as antenna system, channel bandwidth, scheduling algorithm, user speed, the number of users in a cell. One of the simulated scenarios will be replicated in a laboratory where two situations will be studied: the terminals will be located in a Faraday cage and the second case will deal with the terminals in open space. One of the downsides of the simulator is that the latency cannot be determined, while in the laboratory this parameter can be measured. Through this procedure the differences between the theoretical performances of the technology (simulator) and the ideal ones (laboratory) will be emphasized.

3. The project is based on knowledge mainly from the following 3-4 courses: Communication Networks, Data Communications, Antennas and Propagation, Microwaves.

4. The construction part of the project remains in the property of : *UPB*

5. The Intellectual Property upon the project belongs to: *Albert Constantin Iepure and Orange Romania S.A.*

6. The research is performed at the following location: *UPB and Orange Romania S.A.*

7. The thesis project was issued at the date: 26.11.2012

Thesis Advisor:

Ş.l. Dr. Ing. Şerban OBREJA

Dr. Ing. Oana BĂDIŢĂ

STUDENT:

Albert Constantin IEPURE

Statement of Academic Honesty

I hereby declare that the thesis "*Analysis of data transfer in LTE networks*", submitted to the Faculty of Electronics, Telecommunications and Information Technology in partial fulfillment of the requirements for the degree of Engineer in the domain *Electronics and Telecommunications*, study program *Technologies and Communications Systems*, is written by myself and was never before submitted to any other faculty or higher learning institution in Romania or any other country.

I declare that all information sources sources I used, including the ones I found on the Internet, are properly cited in the thesis as bibliographical references. Text fragments cited "as is" or translated from other languages are written between quotes and are referenced to the source. Reformulation using different words of a certain text is also properly referenced. I understand plagiarism constitutes an offence punishable by law.

I declare that all the results I present as coming from simulations and measurements I performed, together with the procedures used to obtain them, are real and indeed come from the respective simulations and measurements. I understand that data faking is an offence punishable according to the University regulations.

Bucharest, *17.06.2013*

Albert Constantin IEPURE

Albert

____

(student's signature)

Contents

List of figures

List of tables

List of Abbreviations

3GPP	3rd Generation Partnership Project
ABS	Almost Blank Subframe
ACK	Acknowledgement
AF	Application Function
AM	Acknowledge Mode
AMC	Adaptive Modulation and Coding
AMR	Adaptive Multi Rate
AMR-WB	Adaptive Multi Rate Wideband
ARP	Allocation and Retention Priority
ARQ	Automatic Repeat Request
BCCH	Broadcast Control Channel
BCH	Broadcast Channel
BER	Bit Error Rate
BLER	Block Error Rate
CBC	Cell Broadcast Center
CCCH	Common Control Channel
CDD	Cyclic Delay Diversity
CDMA	Code Division Multiple Access
CFI	Control Frame Indicator
CINR	Carrier to Interference and Noise Ratio
CMC	Connection Mobility Control
CQI	Channel Quality Indicator
CRC	Cyclic Redundancy Check
CS	Circuit Switched
CSFB	Circuit Switched Fallback
CSI	Channel State Information
DAB	Digital Audio Broadcasting
DCCH	Dedicated Control Channel
DFCA	Dynamic Frequency and Channel Allocation
DFT	Discrete Fourier Transform
DFTS-OFDM	Discrete Fourier Transform Spread OFDM
DL	Downlink
DL-SCH	Downlink Shared Channel
DM-RS	Demodulation Reference Signals
DRA	Dynamic Resource Allocation
DRB	Dedicated Radio Bearer
DRX	Discontinuous Reception
DTCH	Dedicated Traffic Channel
DTM	Dual Transfer Mode

DVB-T	Digital Video Broadcasting – Terrestrial
DwPTS	Downlink Pilot Time Slot
ECDF	Empirical Cumulative Distribution Function
ECM	EPS Connection Management
EFR	Enhanced Full Rate
EIR	Equipment Identity Register
eNB	evolved Node B
EPC	Evolved Packet Core
EPS	Evolved Packet System
E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FDD	Frequency-Division Duplexing
FEC	Forward Error Correction
FR	Full Rate
GBR	Guaranteed Bit Rate
GERAN	GSM/Edge Radio Access Network
GGSN	Gateway GPRS Support Node
GP	Guard Period
GSM	Global System for Mobile Communications
GTP-C	GPRS Tunneling Protocol Control Plane
GTP-U	GPRS Tunneling Protocol for the User Plane
HARQ	Hybrid Automatic Repeat Request
HPLMN	Home PLMN
HRPD	High Rate Packet Data
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
ICIC	Inter-Cell Interference Coordination
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electrical and Electronics Engineers
IMS	IP Multimedia Subsystem
IMT-2000	International Mobile Telecommunications-2000
IP	Internet Protocol
IRRRM	Inter RAT Radio Resource Management
ISI	Inter Symbol Interference
ITU	International Telecommunication Union
ITU-R	International Telecommunication Union Radiocommunication Sector
KPI	Key Performance Indicator
LAM	Link Adaptation Margin
LB	Load Balancing
LMSC	LAN/MAN Standards Committee
LTE	Long Term Evolution
MAC	Medium Access Control
MAX C/I	Maximum Rate

MBMS	Multimedia Broadcast/Multicast Service
MBSFN	Multi-Media Broadcast over a Single Frequency Network
MCCH	Multicast Control Channel
MCH	Multicast Channel
MIB	Master Information Block
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MME	Mobility Management Entity
MME/GW	Mobility Management Entity/Gateway
MMTel	Multi-Media Telephony
MRC	Maximum Ratio Combining
MSC	Mobile Switching Centre
MTCH	Multicast Traffic Channel
MU - MIMO	Multi-User MIMO
NACK	Negative Acknowledgement
NAS	Non Access Stratum
OFDM	Orthogonal Frequency Division Multiplex
OFDMA	Orthogonal Frequency-Division Multiple Access
OSC	Orthogonal Sub-Channel
PAPR	Peak-to-Average Power Ratio
PBCH	Physical Broadcast Channel
PCCH	Paging Control Channel
PCEF	Policy and Charging Enforcement Function
PCFICH	Physical Control Format Indicator Channel
PCH	Paging Channel
PCM	Pulse Coded Modulation
PCRF	Policy and Charging Rules Function
PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Convergence Protocol
PDN-GW	Public Data Network Gateway
PDSCH	Physical Downlink Shared Channel
PDU	Packet Data Unit
PELR	Packet Error Loss Rate
PFTF	Proportional Fair in Time and Frequency
PH	Power Headroom
PHICH	Physical Hybrid ARQ Indicator Channel
PLMN	Public Land Mobile Network
PMCH	Physical Multicast Channel
PMI	Precoding Matrix Indicator
PRACH	Physical Random Access Channel
PRB	Physical Resource Block
PS	Packet Switched

PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QAM	Quadrature Amplitude Modulation
QCI	QoS Class Indicator
QCI	Quality Class Indicator
QoS	Quality of Service
QPSK	Quadrature Phase-Shift Keying
RAB	Radio Access Bearer
RAC	Radio Admission Control
RACH	Random Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RBC	Radio Bearer Control
RE	Resource Element
RI	Rank Indicator
RIP	Received Interface Power
RLC	Radio Link Control
RNC	Radio Network Controller
RNTI	Radio Network Temporary Identifier
RoHC	Robust Header Compression
ROI	Region of Interest
RR	Round Robin
RRC	Radio Resource Control
RRM	Radio Resource Management
RS	Reference Signal
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indicator
S1AP	S1 Access Protocol
S1-MME	S1 control plane interface
S1-U	S1 user plane interface
SAP	Service Access Point
SCC AS	Service Centralization and Continuity Application Server
SC-FDMA	Single-Carrier Frequency-Division Multiple Access
SDI	Silence Description Information
SGSN	Serving GPRS Support Node
S-GW	Serving Gateway
SIMO	Single Input Multiple Output
SINR	Signal to Interference and Noise Ratio
SISO	Single Input Single Output
SMS	Short Message Service

SMSC	SMS Service Center
SNR	Signal to Noise Ratio
SPID	Subscriber Profile ID
SPS	Semi-Persistent Scheduling
SR	Scheduling Requests
SRS	Sounding Reference Signal
SR-VCC	Single Radio Voice Call Continuity
STN-SR	Special Transfer Number for SRVCC
SU - MIMO	Single User MIMO
TA	Timing Advance
TB	Transport Block
TCP	Transmission Control Protocol
TDD	Time Division Duplexing
TDMA	Time Division Multiple Access
TDSM	Time Domain Statistical Multiplexing
UDP	Universal Datagram Protocol
UE	User Equipment
UL	Uplink
UL-SCH	Uplink Shared Channel
UM	Unacknowledged Mode
UMB	Ultra Mobile Broadband
UMTS	Universal Mobile Telecommunications System
UpPTS	Uplink Pilot Time Slot
UTRAN	UMTS Terrestrial Radio Access Network
VoIP	Voice over IP
VoLGA	Voice over LTE via Generic Access
VoLTE	Voice over LTE
VPLMN	Visited PLMN
WCDMA	Wideband Code Division Multiple Access
WiMAX	Worldwide Interoperability for Microwave Access

Introduction

LTE (Long Term Evolution) of the Universal Mobile Telecommunication System (UMTS) also known as the Evolved Packet System (EPS) is a transient step in the field of mobile communications. Such a major change in the already employed technology is necessitated by the unceasing increase in demand for high speed connections on networks, low latency and delay, low error rates and resilience because modern users and network applications have become increasingly dependent on these requirements for efficient functionality and performance. 3GPP LTE (Third Generation Partnership Project Long Term Evolution) promotes high peak data rates for both uplink and downlink transmission, high spectral efficiency, low delay and latency, low bit error rates, to mention but a few. These functional and performance desiderates of 3GPP LTE can be met with a great measure of certainty, but a major change in the technologies that have been already used and the introduction of new ones lay the foundation of these goals that are finally put in practice. LTE leverages on a number of technologies namely MIMO (Multi Input Multiple Output) antennas, OFDM (Orthogonal Frequency Division Multiplexing) and OFDMA (Orthogonal Frequency Division Multiplexing Access) for the downlink, SC-FDMA (Single Carrier Frequency Division Multiple Access) for the uplink, support for various types of modulations: QPSK (Quadrature Phase Shift Keying), 16 QAM (Quadrature Amplitude Modulation) and 64 QAM. This was the major reason that determined me to choose this subject for may Bachelor thesis, because I wanted to get a more detailed insight into the technologies that stand at the basis of achieving the above-mentioned goals and these might as well serve as a basis for the next technologies that are to be deployed in the more recent or distant future.

This thesis work presents the key principles needed to understand how such important advancements in throughput, spectral efficiency, latency and low bit error rates could be achieved. In the first chapter, a list of requirements imposed by 3GPP in order to meet the objectives of LTE is presented, together with the general system architecture of the LTE network and the employed protocols that make the communication among network entities efficient. Afterwards, an overview of the key technologies used in LTE is presented, mainly focusing on: OFDMA, SC-FDMA (together with the main difference between them that determined that one of them to be used in the downlink, whereas the other in the uplink) and last, but not least – MIMO, which is a technology that involves the use of multiple antennas at the transmitter, receiver or both. In the fourth chapter, the most important radio aspects of LTE are presented, including the frequency bands intended for LTE, the six channel bandwidths available, the frame and slot structure, the use of reference signals and refarming – which is a concept that is based upon the idea that parts of the already used spectrum, especially the 1800 MHz part, can be reused for LTE. In chapter five, Radio Resource Management (RRM) aspects are discussed, together with scheduling and link adaptation in LTE and QoS (Quality of Service) concepts, in order to understand how the prioritization of different types of traffic (be it voice, video or gaming) is made possible. The seventh and eighth chapters discuss the challenges of carrying the circuit-switched specific voice traffic in a packet-switched network and the technologies used to benefit from the already implemented infrastructures of the current technologies in order to transport the LTE-specific voice traffic, as well as the Key Performance Indicators (KPIs) and measurements needed for LTE radio network optimization. In the last chapter, the student's original contribution is emphasized through an evaluation of the LTE network performance which is made by using the LTE System Level Simulator and Link Level Simulator of the Matlab simulator developed by Technische Universität Wien. The KPIs that are analyzed include the cell throughput and BLER (Block Error Rate), user throughput and BLER and the CQI (Channel Quality Information), while varying several parameters as antenna system, channel bandwidth, scheduling algorithm, user speed, the number of users in a cell. One of the simulated scenarios is replicated in the laboratory where two situations are studied: the terminals are located in a Faraday cage and the second case deals with the terminals in open space. In the simulator the latency cannot be determined, while in the laboratory this parameter can be measured. Through this procedure the differences between the theoretical performances of the technology (simulator) and the ideal ones (laboratory) will be emphasized. From these simulations it can be concluded that the user speed has a major impact on the throughput it experiences, the throughput decreasing when the user speed increases. The type of scheduler that is used also influences throughput.

The measurements of the practical part of this Bachelor thesis were carried out at the laboratory of Orange in Bucharest, Romania. My sincere gratitude goes to Dr. Ing. Oana Bădiţă, who was my thesis coordinator from Orange, for all the advice, guidance and support that was given to me. I am thankful to the supervisor at the university, Ş.l. Dr. Ing. Şerban Obreja for the initial advice and guidance and for reviewing my work.

1. 3GPP: requirements, objectives

The cellular wireless communications industry encountered a tremendous growth in the past decade with over four billion wireless subscribers worldwide. The first generation (1G) analog cellular systems supported voice communication, the roaming capability being limited. The second generation (2G) digital systems offered higher capacity and better voice quality in comparison with their analog counterparts. In addition to this, roaming became more widespread thanks to fewer standards and common spectrum allocations across countries particularly in Europe. The two mainly used second-generation (2G) cellular systems are GSM (Global System for Mobile Communications) and CDMA (Code Division Multiple Access). Similar to the 1G analog systems, 2G systems were primarily designed to support voice communication, but in subsequent releases of these standards, new capabilities were introduced to support data transmission, even if the data rates were lower than those employed by dial-up connections.

The ITU-R (International Telecommunication Union Radiocommunication Sector) initiative on IMT-2000 (International Mobile Telecommunications 2000) paved the way for evolution to 3G, specifications such as a peak data rate of 2 Mb/s and support for vehicular mobility being published under IMT-2000 initiative. Both the GSM and CDMA camps formed their own separate 3G partnership projects, namely 3GPP (3rd Generation Partnership Project) and 3GPP2, respectively, to develop IMT-2000 compliant standards based on the CDMA technology. The 3G standard in 3GPP is referred to as WCDMA (Wideband Code Division Multiple Access) because it uses a larger 5 MHz bandwidth in comparison with the 1.25 MHz bandwidth used in 3GPP2's CDMA2000 system.

The desired high-speed data transmissions claimed in the first release of the 3G standards were never achieved in practice. In this sense, a serious effort was made to enhance the 3G systems for efficient data support. The 3GPP2 first introduced the HRPD (High Rate Packet Data) system that employed advanced techniques that were optimized for data traffic such as channel sensitive scheduling, fast link adaptation and HARQ (Hybrid Automatic Repeat Request), etc. The HRPD system required the use of a separate 1.25 MHz carrier and did not support voice service. A similar direction was followed by the 3GPP which introduced HSPA (High Speed Packet Access) enhancement to the WCDMA system. In comparison with HRPD, HSPA supported both voice and data to be carried on the same 5 MHz carrier, the voice and data traffic being code multiplexed in the downlink. In a later release of HRPD, VoIP (Voice over IP) capabilities were introduced to provide both voice and data service on the same carrier.

In the same time as HSPA and HRPD systems were being developed and deployed, IEEE (Institute of Electrical and Electronics Engineers) 802 LMSC (LAN/MAN Standards Committee) introduced the IEEE 802.16e standard for mobile broadband wireless access, which made use of a different access technology named OFDMA (Orthogonal Frequency Division Multiple Access) and obtained better data rates and spectral efficiency than the ones provided by HSPA and HRPD. Even if the IEEE 802.16 family of standards is officially called WirelessMAN in IEEE, it has been called WiMAX (Worldwide Interoperability for Microwave Access) by WiMAX Forum, whose mission is to promote and certify the compatibility and interoperability of broadband wireless access products. The WiMAX system supporting mobility as in IEEE 802.16e standard is referred to as Mobile WiMAX. In addition to the radio technology advantage, the network architecture of Mobile WiMAX is a simpler one, based on IP (Internet Protocol).

The introduction of Mobile WiMAX determined both 3GPP and 3GPP2 to develop their proprietary version of beyond 3G systems based on the OFDMA technology and IP protocol-based network architecture. The beyond 3G system in 3GPP is called evolved UTRA and is also widely referred to as LTE (Long Term Evolution), while 3GPP2's version is called UMB (Ultra Mobile Broadband) [1].

3GPP LTE is one of several evolving 3G wireless standards loosely referred to as 3.9G. The first publicly available LTE service in the world was launched by TeliaSonera in Oslo on December 14 2009. Even if it is marketed as a 4G wireless service, LTE which is specified in the 3GPP Release 8 and 9 does not meet the technical requirements adopted by 3GPP. Yet, due to marketing pressures and the noteworthy improvements WIMAX, HSPA+ and LTE bring to the initial 3G technologies, ITU (International Telecommunication Union) decided that the previously mentioned technologies can be considered 4G technologies. It is the LTE Advanced standard that properly satisfies the requirements imposed by ITU-R and is considered by the institution as "True 4G".

The main drivers for LTE development are as following: reduced delay for connection establishment; reduced transmission latency for user plane data; increased bandwidth and bit rate per cell, also at the cell edge; reduced costs per bit for radio transmission; greater flexibility of spectrum usage; simplified network architecture; seamless mobility, including between various radio access technologies; reasonable power consumption for the mobile terminal [2].

The system supports flexible bandwidths thanks to OFDMA and SC-FDMA (Single-Carrier Frequency-Division Multiple Access) access schemes. In addition to FDD (Frequency Division Duplexing) and TDD (Time Division Duplexing), half-duplex FDD can support low cost UE (User Equipment) because in this mode of operation a UE is not required to transmit and receive at the same time, avoiding thus the need for a costly duplexer in the UE. The system is primarily optimized for low mobile speeds up to 15 km/h. However, the system specifications allow mobility support in excess of 350 km/h with some performance degradation. The uplink access is based on SC-FDMA that promises increased uplink coverage due to low PAPR (Peak-to-Average Power Ratio) relative to OFDMA.

Higher data rates are achieved through higher order modulation - up to 64 QAM (Quadrature Amplitude Modulation), MIMO (Multiple Input Multiple Output) transmission (up to 4x4 in the downlink) and a large allocated bandwidth.

One aspect to be noted is that the downlink is specified for SISO (Single Input Single Output) and MIMO antenna configurations at a fixed 64QAM modulation, whereas the uplink is specified only for SISO but with different modulation schemes. Lower rates are specified for specific UE categories, and performance requirements under non-ideal radio conditions have also been developed.

The LTE system uses as modulation schemes QPSK (Quadrature Phase-Shift Keying), 16-QAM and 64-QAM, turbo codes for channel coding and techniques such as: channel sensitive scheduling, link adaptation, power control, ICIC (Inter-Cell Interference Coordination) and HARQ.

The LTE system is described by a sub 5 ms latency for small IP packets and scalable channel bandwidths of 1.4, 3, 5, 10, 15, 20 MHz in both the uplink and the downlink direction [3].

Some specific targets that were set for LTE are listed below, according to 3GPP TR 25.913: increased peak data rate: 100Mbps for DL (Downlink) with 20MHz (2 Rx Antenna at UE), 50Mbps for UL (Uplink) with 20MHz; improved spectral efficiency: 5bps/Hz for DL and 2.5bps/Hz for UL (improvements over Release 6 HSPA of three to four times in the DL and two to three times in the UL); improved cell edge performance (in terms of bit rate); reduced latency; co-existence with legacy standards while evolving toward an all-IP network [4].

FDD downlink peak data rates (64QAM)			
Antenna configuration	SISO	2x2 MIMO	4x4 MIMO
Peak data rate Mbps	100	172.8	326.4
FDD uplink peak data rates (single antenna)			
Modulation depth	QPSK	16QAM	64QAM
Peak data rate Mbps	50	57.6	86.4

Table 1.1 LTE (FDD) downlink and uplink peak data rates [5]

2. System architecture and protocols

2.1 System architecture

GSM was designed to carry real time services in a circuit switched manner, data services being sent only over a circuit switched modem connection, achieving very low data rates. The first step towards an IP based packet switched solution was made through the transition of GSM to GPRS, by keeping the same air interface and access method: TDMA (Time Division Multiple Access).

In order to further increase data rates UMTS (Universal Mobile Telecommunications System) was developed with a new access network, based on CDMA. A circuit switched connection for real time services and a packet switched connection for data services were emulated by the access network in UMTS. In this scenario, an IP address is allocated to the UE when a data service is established and released when the service is released. Incoming data services still rely upon the circuit switched core for paging [6].

The LTE network architecture is designed in order to support packet-switched traffic with continuous mobility, quality of service and minimal latency. Voice through packet connections and data services are supported through this packet-switched approach. A flat and very simple network architecture results, containing only two types of nodes: eNB (evolved Node B) - which are the base stations of the network, and MME/GW (Mobility Management Entity/Gateway). This contrasts with the much more numerous network nodes in the hierarchical network architecture of the 3G system. A very important alteration that is worth observing is that RNC (Radio Network Controller) is removed from the data path and its functions are now assimilated by the eNB. A single node in the access network comes with a lot of benefits, of noteworthy importance being the reduced latency and the distribution of the RNC processing load among the multiple eNBs [7].

LTE has an always-on concept, in which the radio bearer is set up the moment a subscriber attaches to a network, all radio resources provided to subscribers by the E-UTRAN (Evolved Universal Terrestrial Radio Access Network) being shared. There are several types of bearer:

- the radio bearer, which is the point-to-point bidirectional connection for the user plane between the UE and the eNB;
- the S1 bearer, which is the user plane connection between the eNB and S-GW (Serving Gateway);
- the RAB (Radio Access Bearer), which is the user plane connection between the UE and S-GW;
- the S5/S8 bearer, which is the user plane connection between the S-GW and PDN-GW (Public Data Network Gateway);
- the EPS (Evolved Packet System)bearer, which is a virtual connection between UE and P-GW that identifies the data sent and received between these two end points with specific QoS (Quality of Service) attributes.

The network entity responsible for radio interface transmission and reception is the eNB, which performs: radio channel modulation/demodulation, channel coding/decoding and multiplexing/demultiplexing. In each cell system information is broadcast on the radio interface DL to provide the UE with basic information that is needed to access the network.

At the eNB level the data is routed, multiplexed, ciphered/deciphered, segmented and reassembled. It can be said that on the E-UTRAN transport layer level, the eNB acts as an IP router and switch. On the control plane level, the eNB selects the MME (Mobility Management Entity) to which NAS (Non Access Stratum) signaling messages are routed.

The LTE base station performs all RRC (Radio Resource Control) functions such as broadcast of system information and RRC connection control including:

- paging of subscribers;
- set up, alteration and release of RRC connection, including the allocation of temporary UE identities, like RNTI (Radio Network Temporary Identifier);
- ciphering for both user plane and control plane traffic;
- intra-LTE handovers: both intra-frequency and inter-frequency;
- set-up, alteration and release of DRBs (Dedicated Radio Bearers) carrying user data;
- radio configuration control, mainly the assignment and adjustment of ARQ, HARQ (Hybrid ARQ) parameters and DRX (Discontinuous Reception) configuration parameters;
- QoS control;
- Recovery functions in order to re-establish radio connections after physical channel failures.

In addition to this, the RRC unit of the eNB performs all types of intra and inter-LTE measurements, such as:

- Set-up, alteration and release of measurements for intra-LTE intra-frequency, intra-LTE inter-frequency, inter-RAT (Radio Access Technology) mobility, transport channel quality, UE internal measurement reports;
- The estimation of reported measurement results and start of necessary handover procedures are also eNB functions (while in 3G UMTS all measurement evaluation and handover control functions have been implemented in the RNC).

The most important part for measuring the eNB performance is the UL/DL resource management and packet scheduling performed by the eNB. The eNB needs to handle different constraints like radio link quality, user priority, requested QoS and UE capabilities such as to make use of the available resources in the most efficient way.

The MME is in charge of the NAS connection with the UE. All NAS signaling messages are exchanged between the UE and MME to trigger further procedures in the core network.

The sole purpose of MME is signaling, implying that user IP packets do not pass through it. A separate network entity for signaling is advantageous in the sense that the network capacity for signaling and traffic can grow independent from one another.

A new function implemented in the E-UTRAN is NAS signaling security, whose purpose is to protect the signaling messages that could reveal the true subscriber’s identity and its location from unauthorized eavesdropping.

The MME also pages subscribers in the ECM (EPS Connection Management) IDLE state and is concerned with the management tracking area lists. The list of tracking areas is the list of locations where the UE will be paged.

To route the user plane data streams the MME will select the most appropriate PDN-GW and S-GW. It will also connect the E-UTRAN with the 3G UTRAN (UMTS Terrestrial Radio Access Network) using the S3 interface, namely the MME to SGSN (Serving GPRS Support Node).

Moreover, the MME performs handover management by selecting a target MME or SGSN for handovers to 2G or 3G 3GPP access networks. Also, it is the MME that establishes the connection to the HSS (Home Subscriber Server) across the S6a interface being thus responsible for roaming management and authentication of subscribers.

The MME also performs the bearer management function: it sets up, modifies and releases default and dedicated bearers.

The S-GW is the gateway which terminates the interface to the E-UTRAN. An LTE subscriber will always be connected to a single S-GW.

In the case of inter-eNB handover, the S-GW acts as the mobility anchor of the connection, remaining the same while the path for the transport of signaling and user plane will be switched onto the S1 interface. Once a handover is executed successfully and the associated UE has left the corresponding S-GW, the old S-GW will send one or more “end marker” packets to the source eNB, source SGSN, or source RNC of the handover to support the reordering of user plane packets in these network elements.

The S-GW routes the traffic between the 2G/3G SGSN and the PDN-GW of the EPC (Evolved Packet Core), which means that when it comes to inter-RAT handover involving the S4 interface, the S-GW acts as the GGSN (Gateway GPRS Support Node).

If the UE is in IDLE mode the S-GW buffers user plane packets that will be sent to the UE once a successful paging response occurs. The paging via the S1 and Uu interfaces is also triggered by the S-GW.

The S-GW provides connectivity and software implementations for lawful interception, as well as for charging function for UL and DL per UE, PDN and QCI (QoS Class Indicator). These functions are used to charge the operator’s own subscribers as well as roaming users (inter-operator charging).

On the IP transport layer the S-GW acts as a packet router, transparently forwarding user plane packets in the UL and DL direction, marking their underlying transport units with parameters like DiffServ Code Point, based on the QCI of the associated EPS bearer.

The PDN-GW interfaces with external packet data networks, such as the Internet and the IMS (IP Multimedia Subsystem), perform several IP functions such as address allocation, policy enforcement, packet filtering and routing. It is possible for a user to be connected to more than just one PDN-GW if the user has access to more than one packet data network. Though, it is not possible for the same UE to simultaneously open connections to a PDN-GW and to a GGSN in the 3G packet-switched domain, according to 3GPP standards.

The main function of the PDN-GW is to establish, maintain and delete GTP tunnels to S-GW or SGSN in the case of inter-RAT mobility. The PDN-GW assigns the user’s dynamic IP addresses and routes the user plane packets, providing in addition to this functions for lawful interception, policy/QoS control, and charging [8].

Each eNB is connected to one or more MMEs which, at their turn, are connected to a S-GW that may be collocated (encompassed in the same hardware) with the MME. The interface between the eNB and MME is called S1 interface. In the case where the MME and S-GW are not collocated, the S1-MME (S1 control plane interface) connects the eNB and MME, while S1-U (S1 user plane interface) connects the eNB with the S-GW.

In case where one eNB is connected to multiple MMEs, these MMEs form a so-called MME pool and the corresponding network functionality is called S1 flex.

In the LTE system, there is no centralized intelligent controller. The main reason for the distribution of intelligence among eNBs is to quicken the connection set-up and reduce the necessary time for a handover to be realized. For the end-user the connection set-up time for a real time data session is most of the time crucial, especially in on-line gaming. The time for a handover to take place is essential for real-time services where handovers that take too much time determine the end-users to prematurely end calls.

A few benefits arise from the presence of a single node in the access network, such as reduced latency and the distribution of RNC processing load into multiple eNBs.

The X2 interface is used to interconnect eNBs, as it can be seen in Figure 2.1. The main role of this connectivity is the intra-E-UTRAN handover.

It must be noted that only the base stations and their physical connections (wires or fibers) are defined by 3GPP as the E-UTRAN, while MME and S-GW are seen as elements of the EPC network [9]. The E-UTRAN and EPC form together the EPS, which is entirely IP based, meaning that both real time and non-real time services will be carried by the IP protocol, the IP address being allocated when the mobile is switched on and released when it is switched-off.

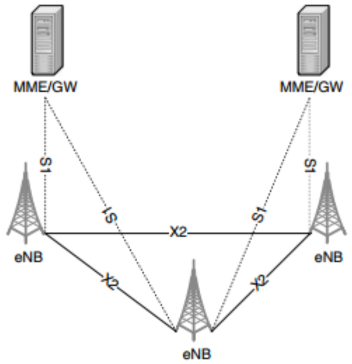

Figure 2.1 Network architecture [10]

2.2 Interfaces

The E-UTRAN is an all-IP network, as previously mentioned. The figure below shows the network elements that are typically involved in the signaling procedures and routing of payload data from the UE to the PDN-GW and vice versa. In the figure we can also observe the reference points for inter-RAT handover (and inter-RAT packet routing) between E-UTRAN, UTRAN and GERAN (GSM/EDGE Radio Access Network).

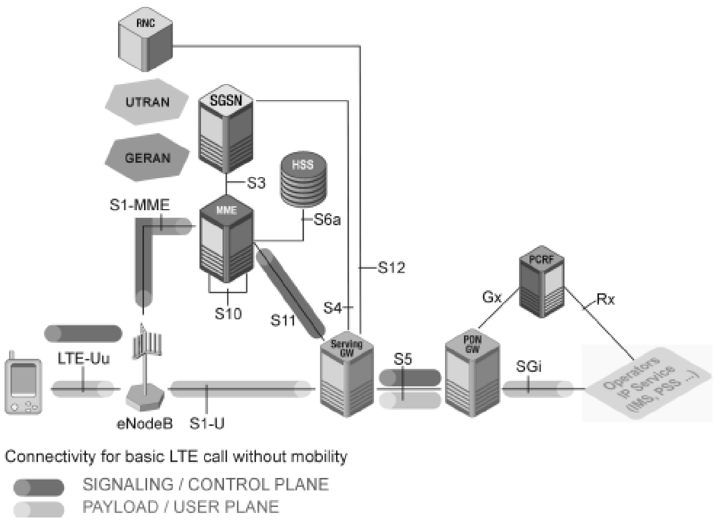

Figure 2.2 Connection via E-UTRAN non-roaming architecture [11]

The point of reference in the figure below can be described as follows:

- **S1-MME**: is the reference point for the control plane protocol between the E-UTRAN and MME. The protocol used at this reference point is S1AP (S1 Access Protocol).
- **S1-U**: is the reference point between the E-UTRAN and S-GW for the bearer user plane tunneling and inter-eNB path switching during handover. The protocol used at this reference point is GTP-U (GPRS Tunneling Protocol User Plane).
- **S3**: is the reference point between the MME and SGSN, the latter may serve UTRAN, GERAN or both. On S3 passes control plane information for user and

bearer information exchange for inter-RAT handover. If the connection was initially set up in the E-UTRAN and is handed over to UTRAN/GERAN the corresponding user plane streams are routed across the S4 reference point. In the case of UTRAN/GERAN to E-UTRAN handover, depending on the S-GW which can also act as the anchor for UTRAN/GERAN traffic, the user plane tunnel can be easily switched between S4 and S1-U during the handover. The protocol used at this reference point is the GTP-C (GPRS Tunneling Protocol Control Plane).

- **S4**: provides related control and mobility support between the GPRS core and the 3GPP anchor function of the S-GW using GTP-C. Moreover, if a direct tunnel across S12 is not established, it provides user plane tunneling using GTP-U.
- **S5**: provides user plane tunneling and tunnel management between the S-GW and PDN-GW. It is employed in case of S-GW relocation caused by UE mobility and if the S-GW needs to connect to a non-collocated PDN-GW for the required PDN connectivity. The protocol used at this reference point is GTP for both the control plane and user plane.
- **S6a**: is established between the MME and HSS and it enables the transfer of subscription and authentication data for authorizing user access to the network. The protocol used at this reference point is the DIAMETER protocol, which is the successor of RADIUS, a protocol for granting access and authentication.
- **Gx**: this point provides transfer of (QoS) policy and charging rules from the PCRF (Policy and Charging Rules Function) to the PCEF (Policy and Charging Enforcement Function) in the PDN-GW. This implies that a set of rules for charging the transmission of a particular user data stream, which is called service flow, will be requested by the PDN-GW when the bearer is established and the PCRF will provide the required parameters for the charging process. Particularly, it will signal which of the following charging models are applicable: volume-based charging, time-based charging, volume and time-based charging, event-based charging, no charging (if the user pays at a monthly flat rate). Also, information about prepaid limits and other thresholds can be included.
- **S8**: is used only in the case of roaming subscribers. It is the inter-PLMN (Public Land Mobile Network) reference point providing the user plane and control plane between the S-GW in the VPLMN (Visited PLMN) and the PDN-GW in the HPLMN (Home Public Land Mobile Network). It is based on GTP and can be compared to the Gp interface defined for GERAN GPRS.
- **S9**: is also used by roaming subscribers only. It provides transfer of (QoS) policy and charging control information between the home PCRF and the visited PCRF in order to support the local breakout function. For example: the case of a prepaid limit that can only be known by the home PCRF and must be provided to the visited PCRF to allow roaming services for a particular user.
- **S10**: is the reference point between MMEs for MME relocation and MME-to-MME information transfer, providing mobility functions for intra-E-UTRAN handover/relocation. Basically, signaling procedures on this interface are triggered by UE mobility. This kind of MME relocation in 3GPP 23.401 is called S1 handover. Hence, S10 is seen as special kind of S1 interface and the S1AP protocol is used at this reference point.
- **S11**: is the reference point between the MME and S-GW. The protocol used at this reference point is the GTP-C, the corresponding user plane being routed across S1-U.
- **S12**: its use is an operator configuration option, allowing only GTP-U monitoring. It is located between the RNC in the 3G UTRAN and the S-GW for user plane

tunneling when a “direct tunnel” is established. The protocol used at this reference point is GTP-U.

- **S13**: it enables a UE identity check procedure between the MME and EIR (Equipment Identity Register). Usually there is no EIR installed in public networks due to the high administrative efforts, but this network element is not uncommon to be found in private networks.
- **SGi**: is the reference point between the PDN-GW and the packet data network, which can be an operator external public or private packet data network or an intra-operator packet data network, for example for the provision of IMS services. For many user plane connections SGi is the interface to the public Internet. Usually, the whole TCP/IP (Transmission Control Protocol/ Internet Protocol) suite can be monitored at this point.
- **Rx**: is between the AF (Application Function) and the PCRF. It is for instance mandatory if real-time communication services such as VoIP are to be charged in a different manner than common PS (Packet Switched) data transfer.
- **SBc**: lies between the CBC (Cell Broadcast Center) and MME for warning message delivery and control functions. This interface is used to broadcast warning messages to subscribers, like bush fire or tsunami alarms [12].

2.3 Layer 2 structure

According to 3GPP specifications, LTE’s Layer 2 structure consists of three sub-layers, which are: MAC (Medium Access Control), RLC (Radio Link Control) and PDCP (Packet Data Convergence Protocol). The hierarchy of these layers is illustrated in Figure 2.3 and 2.4. The SAP (Service Access Point) between the Physical layer and the MAC sub-layer provides the transport channels while the SAP between the MAC and RLC sub-layers provide the logical channels. The MAC sub-layer multiplexes RLC links, performs scheduling and priority handling serving via logical channels.

The main difference between downlink and uplink structures arises from the fact that in the downlink, the MAC sub-layer manages the priority both among UEs and among the logical channels of a single UE. The other functions performed by the MAC sub-layer in both downlink and uplink include transport format selection, mapping between the logical and the transport channels, HARQ, multiplexing of RLC PDUs (Packet Data Units) and padding.

The main services and functions provided by the RLC sub-layer consist of segmentation, ARQ, in-sequence delivery and duplicate detection etc. The in-sequence delivery of upper layer PDUs is not guaranteed during a handover. The reliability of RLC is configurable to either AM (Acknowledge Mode) or UM (Unacknowledged Mode) transfers. The latter can be used for radio bearers that can allow some loss. In the former-mentioned mode, ARQ functionality of RLC retransmits transport blocks that fail recovery by HARQ. The recovery at HARQ may fail due to hybrid ARQ NACK (Negative Acknowledgement) to ACK (Acknowledgement) error or because the maximum number of retransmission attempts is reached. In this case, the relevant transmitting ARQ entities are notified and potential retransmissions and re-segmentation can be initiated.

The PDCP layer performs functions such as ciphering, header compression and decompression (performed using RoHC protocol) and in-sequence delivery and duplicate detection at handover for RLC AM.

There are three groups into which the various data channels may be grouped:

- **Physical channels**: transmission channels that carry user data and control messages;
- **Transport channels**: the physical layer transport channels offer information transfer to MAC and higher layers;

- **Logical channels**: provide services for the MAC layer within the LTE protocol structure.

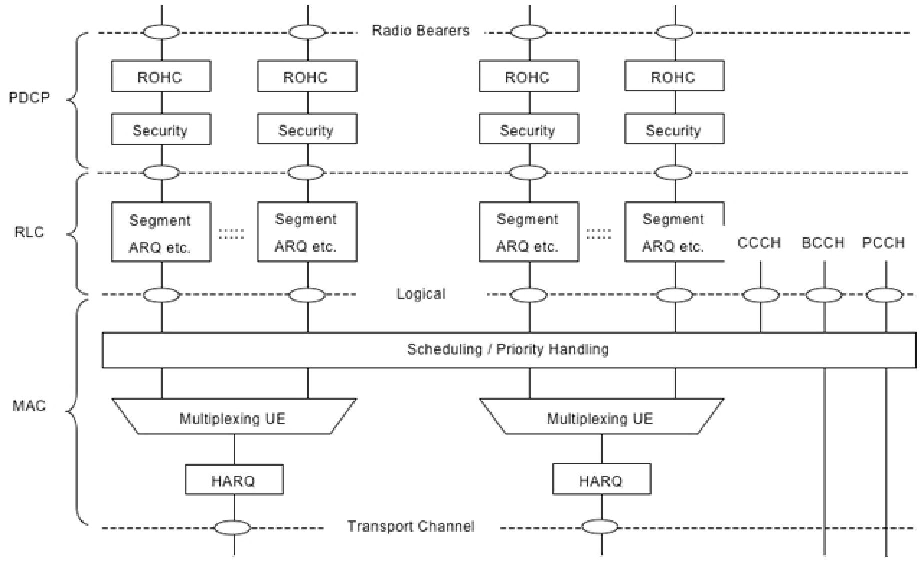

Figure 2.3 Downlink Layer 2 Structure [13]

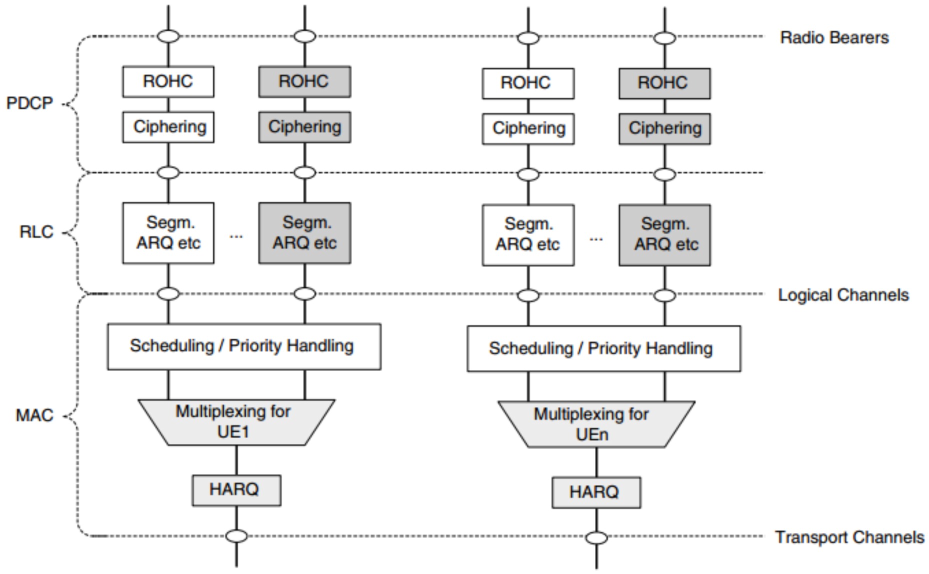

Figure 2.4 Uplink Layer 2 Structure [14]

2.3.1 Downlink logical, transport and physical channels

As stated by 3GPP, numerous types of data transfer services are offered by MAC. Each logical channel type is defined according to the type of information to be transferred. Logical channels can be classified into two groups:

- **Control Channels** (intended for the transfer of control plane information)
- **Traffic Channels** (intended for the transfer of user plane information).

The figure below depicts the mapping between logical channels, transport channels and physical channels for downlink:

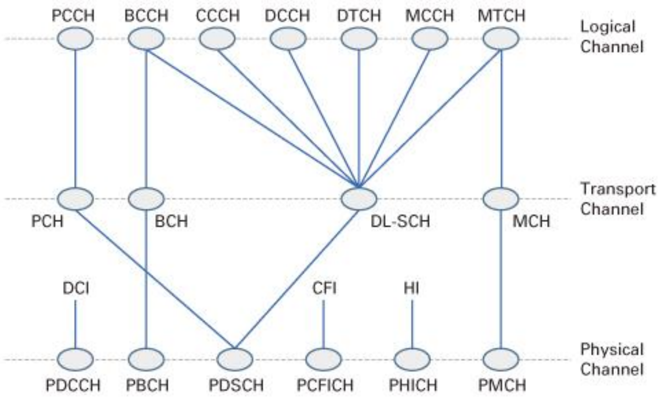

Figure 2.5 Downlink logical, transport and physical channels mapping [15]

In the downlink, five control channels and two traffic channels are defined as follows:

- **PCCH (Paging Control Channel)**: is the downlink control channel used for paging information and system information change notifications, being used when the network does not know about the location cell of the UE.
- **BCCH (Broadcast Control Channel)**: is the downlink channel that conveys system control information.
- **CCCH (Common Control Channel)**: carries control information between the network and the UEs. It is used for UEs that have no RRC connection with the network.
- **DCCH (Dedicated Control Channel)**: a point-to-point bi-directional channel that carries information between the network and the UE. It is used for UEs that have an RRC connection with the network.
- **MCCH (Multicast Control Channel)**: is used for the transmission of MBMS (Multimedia Broadcast/Multicast Service) control information (information needed for multicast reception), being used solely by those UEs receiving MBMS.

The two traffic channels employed in the downlink are:

- **DTCH (Dedicated Traffic Channel)**: is a point-to-point channel dedicated to a single UE for the transmission of user data.
- **MTCH (Multicast Traffic Channel)**: is a point-to-multipoint channel used for the transmission of user traffic to UEs receiving MBMS (multicast data)

The PCCH is mapped to a transport channel referred to as PCH (Paging Channel), which is practically used to convey the PCCH. It is also DRX enabled in order to assist UE power saving.

The BCCH is mapped to either a transport channel referred to as a BCH (Broadcast Channel) or to the DL-SCH (Downlink Shared Channel). The BCH is characterized by a fixed pre-defined format as this is the first channel the UE receives after acquiring synchronization to the cell and is broadcasted over the entire coverage area of the cell. In LTE, the broadcast channel is used to transmit the System Information field necessary for system access.

The MCCH and MTCH are either mapped to a transport channel called MCH (Multicast Channel) or to the DL-SCH. The MCH supports MBSFN (Multi-Media Broadcast over a Single Frequency Network) combining of MBMS transmission from multiple cells, being used to transmit MCCH information to set up multicast transmissions.

The other logical channels, namely CCCH, DCCH and DTCH are mapped to the transport channel DL-SCH, which is the main channel for downlink data transfer. The DLSCH is characterized by support for adaptive modulation/coding, HARQ, power control, semi-static/dynamic resource allocation, DRX, MBMS transmission and multi-antenna technologies. All the four-downlink transport channels have the requirement to be broadcast in the entire coverage area of a cell [16].

The BCH is mapped to a physical channel referred to as PBCH (Physical Broadcast Channel), which carries system information for UEs requiring access to the network. The messages transported over this channel are called MIB (Master Information Block) messages. The employed modulation scheme is always QPSK because it offers a high resistance to noise and the information bits are coded and rate matched - the bits are then scrambled using a scrambling sequence that is cell-specific to prevent interference with data from other cells. The MIB message on the PBCH is mapped onto the central 72 subcarriers or six central resource blocks regardless of the overall system bandwidth. A PBCH message is repeated every 40 ms, being transmitted over four sub-frames. The 40 ms timing is detected blindly without requiring any explicit signaling. Also, each sub-frame transmission of BCH is self-decodable and UEs with good channel conditions may not need to wait for reception of all the four sub-frames for PBCH decoding. The PBCH transmissions have 14 information bits, 10 spare bits, and 16 CRC (Cyclic Redundancy Check) bits.

The PCH and DLSCH are mapped to a physical channel referred to as PDSCH (Physical Downlink Shared Channel), which is the main data bearing channel which is allocated to users on a dynamic and opportunistic basis. The PDSCH carries data in TBs (Transport Blocks) which correspond to a MAC PDU, being passed from the MAC layer to the PHY layer once per TTI which is 1 ms (in order to meet low latency requirements). A convolutional turbo coder is used for forward error correction in order to guard against propagation channel errors. The data is mapped to spatial layers according to the type of multi-antenna technique (closed loop spatial multiplexing, open-loop, spatial multiplexing, transmit diversity) and then mapped to a modulation symbol which includes QPSK, 16 QAM and 64 QAM.

The MCH is mapped to PMCH (Physical Multicast Channel), which is the multi-cell MBSFN transmission channel. It is designed for a single-frequency network and it requires that the base stations transmit with tight time synchronization the same modulated symbols. The PMCH is transmitted only in specific dedicated sub-frames where the PDSCH is not transmitted.

The three stand-alone physical control channels are:

- **PCFICH (Physical Control Format Indicator Channel)**: is transmitted every sub-frame and carries the CFI (Control Frame Indicator) which includes the number of OFDM (Orthogonal Frequency Division Multiplex) symbols used for control channel transmission in each sub-frame (typically 1, 2, or 3). The 32-bit long CFI is mapped to 16 Resource Elements in the first OFDM symbol of each downlink frame using QPSK modulation.
- **PDCCH (Physical Downlink Control Channel)**: is used to inform the UEs about the resource allocation of PCH and DL-SCH as well as modulation, coding and HARQ information related to DL-SCH. A maximum of three or four OFDM symbols can be used for PDCCH. With dynamic indication of number of OFDM symbols used for PDCCH via PCFICH, the unused OFDM symbols among the three or four PDCCH OFDM symbols can be used for data transmission. QPSK modulation is used for the PDCCH.
- **PHICH (Physical Hybrid ARQ Indicator Channel):** is used to carry HARQ ACK/NAK which indicates to the UE whether the eNB correctly received the uplink user data carried on the PUSCH (Physical Uplink Shared Channel). BPSK modulation is used with a repetition factor of 3 for robustness [17].

2.3.2 Uplink logical, transport and physical channels

The figure below depicts the mapping between logical channels, transport channels and physical channels for uplink:

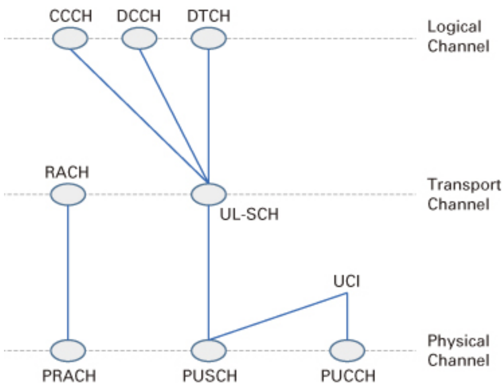

Figure 2.6 Uplink logical, transport and physical channels mapping [18]

In the uplink only two control channels and a single traffic channel are defined as follows:

- **CCCH**: carries control information between the network and the UEs. It is used for UEs that have no RRC connection with the network.
- **DCCH**: a point-to-point bi-directional channel that carries information between the network and the UE. It is used for UEs that have an RRC connection with the network.
- **DTCH**: is a point-to-point channel dedicated to a single UE for the transmission of user data.

All the three uplink logical channels are mapped to a transport channel named UL-SCH (Uplink Shared Channel), which is the main channel for uplink data transfer. The UL-SCH supports adaptive modulation/coding, HARQ, power control and semi-static/dynamic resource allocation.

Another transport channel defined for the uplink is referred to as RACH (Random Access Channel), which is used for random access requirements and transmission of limited control information from a UE with possibility of collisions with transmissions from other UEs.

The physical channel to which the RACH is mapped is called PRACH (Physical Random Access Channel), and it is used for random access functions. This channel carries the random access preamble a UE sends to access the network in non-synchronized mode and used to allow the UE to synchronize timing with the eNB. This is the only non-synchronized transmission that the UE can make within LTE (the downlink and uplink propagation delays are unknown when PRACH is used and therefore it cannot be synchronized.

The PRACH instance is made up of two sequences: a cyclic prefix and a guard period. The preamble sequence may be repeated to enable the eNB to decode the preamble when link conditions are poor.

The UL-SCH transport channel is mapped to the physical channel called PUSCH, which is the uplink counterpart of PDSCH. This channel carries user data. It supports QPSK and 16 QAM modulation, with 64QAM being optional. The uplink scheduling interval is 1 ms, similar to the downlink. However, it is possible to merge a group of 4 TTIs to improve performance at cell edge and reduce higher layer protocol overhead by segmenting MAC PDU for transmission over multiple TTIs.

The only stand-alone uplink physical channel referred to as PUCCH (Physical Uplink Control Channel) provides the various control signaling requirements. A number of various PUCCH formats are defined to enable the channel to carry the required information in the most efficient format for the particular scenario encountered. It is used to carry HARQ ACK/NACK for uplink transmissions, uplink CQI (Channel Quality Indicator), MIMO feedback (Rank Indicator, Precoding Matrix Indicator), SRs (Scheduling Requests) for uplink transmission [18].

3. Key technologies

The 3G systems nowadays make use of a WCDMA scheme within a 5 MHz bandwidth in both the downlink and the uplink directions. In WCDMA, multiple users that employ orthogonal Walsh codes are multiplexed on to the same carrier. In the downlink, the information transmitted on different Walsh codes is orthogonal when it is received at the UE. The reason for this is the fact that the signal is transmitted from a fixed location, namely the NodeB (in the downlink) and all the Walsh codes are synchronized when they are received at the UE. Consequently, in the absence of multi-path propagation, transmissions on different codes do not interfere with each other. Yet, in the presence of multi-path propagation, which is characteristic for cellular environments, the Walsh codes at reception are no longer orthogonal and interfere with each other resulting in inter-user and/or inter-symbol interference. The multi-path interference can be eliminated by using a more complex receiver such as linear minimum mean square error receiver.

The multi-path interference problem of WCDMA becomes even more disturbing for larger bandwidths such as 10 and 20 MHz as those that are required by LTE for support of higher data rates. This is because chip rate increases for larger bandwidths and hence more signals affected by multi-path propagation must be resolved due to shorter chip times, determining a major increase in the complexity of the receiver. Another solution would be to use multiple 5 MHz WCDMA carriers to support 10 and 20 MHz bandwidths. However, transmitting and receiving multiple carriers causes the Node B and UE to become more complex. Another factor that hindered the use of WCDMA for LTE was the unavailability of flexible bandwidth support as the supported bandwidths could only be multiples of 5 MHz and also bandwidths smaller than 5 MHz could not be supported.

The LTE requirements and scalability conflicts with WCDMA determined a new access scheme to be used in the downlink. LTE is based on OFDM with cyclic prefix in the downlink, and on SC-FDMA with cyclic prefix in the uplink. It supports both FDD and TDD duplex modes for transmission on paired and unpaired spectrum.

3.1 Multiple access technology in the downlink: OFDM and OFDMA

OFDM and the corresponding multiple access scheme OFDMA is the result of an evolution spanning over decades of digital transmission techniques and advances in the digital signal processing methods and technologies. Initial applications of OFDM were in the direction of broadcast systems, especially digital radio and video broadcasting standards in Europe, like DAB (Digital Audio Broadcasting) and DVB-T (Digital Video Broadcasting – Terrestrial), and then in IEEE standards like 802.11a for Wireless LAN and WiMAX [19].

The basic OFDM principle is to divide the available spectrum into narrowband parallel channels referred to as subcarriers and transmit information on these channels at a reduced signaling rate. The purpose is to let each channel experience almost flat-fading simplifying the channel equalization process. Each data symbol is then transmitted on one subcarrier, making it robust against multipath propagation. Further advantages of OFDM are the very efficient spectrum usage and, with digital signal processing being cost-effective and flexible, also low-complexity application of the MIMO principle. The name OFDM comes from the fact that the frequency responses of the subchannels are overlapping and orthogonal. The OFDM symbol signal has much larger amplitude variations than the individual subcarriers, making this technology have a high peak to average signal ratio, since the OFDM waveform in time domain is a superposition of sinusoids with frequencies which are n-times the frequency of the “lowest” subcarrier [20].

An OFDM signal consists of a certain number of closely spaced modulated carriers. When modulating a carrier with data, voice, etc., then sidebands in the frequency domain extend either side. It is necessary for a receiver to be able to receive the whole signal to be able to successfully demodulate the data. As a result, when modulated signals are transmitted close to one another they must be set apart in such a way that the receiver can extract them using a non-ideal filter and there must be a guard interval between them, which is not the case with OFDM: even if the sidebands from each carrier overlap, they can still be received without interference because they are orthogonal to each other. This is accomplished by having the carrier spacing equal to the reciprocal of the symbol period. As the carrier spacing is equal to the inverse of the symbol period implies that a whole number of cycles will be found in the symbol period and their contribution will sum up to zero, which implies that there is no interference contribution.

A question of major importance for both the OFDM transmitter and receiver is that they must be linear. Any non-linearity will cause interference between the carriers as a result of inter-modulation distortion, which will introduce unwanted signals that will cause interference and affect the orthogonality of the transmitted subcarriers.

The data to be transmitted on an OFDM signal is spread across the orthogonal subcarriers, each one of them taking a part of the payload, reducing thus the data rate taken by each carrier. The lower data rate has the advantage that interference from reflections is greatly diminished, which is achieved by adding a guard interval into the system, ensuring that the data is only sampled when the signal is stable and not perturbed by delayed replicas of itself that would alter the timing and phase of the signal. The guard interval represents the interval of time between the moments the direct signal and the last reflection is received.

The distribution of the data across a large number of subcarriers has some other advantages: nulls caused by multi-path propagation or interference on a given frequency only affect a small number of the subcarriers, the unperturbed ones being received correctly. By using error-coding techniques, a large extent of the corrupted data will be reconstructed within the receiver. This can be performed because the error correction code is transmitted in a different part of the signal [21].

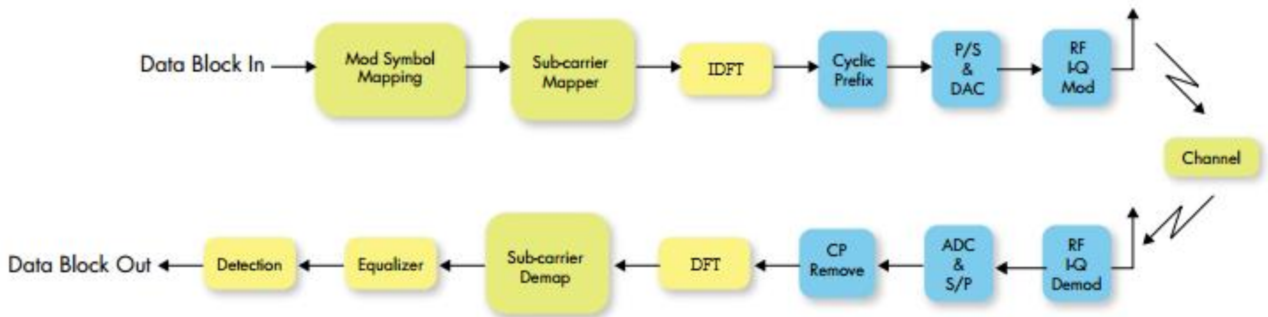

Figure 3.1 OFDM transceiver [22]

As it can be seen in the figure above, an OFDM transmitter includes a baseband modulator, subcarrier mapping, inverse discrete Fourier transformer, cyclic prefix addition, parallel-to-serial conversion, a digital-to-analog converter, followed by an in-phase and quadrature (I-Q) radio frequency modulator and finally by an antenna that transmits the signal over the medium. OFDM transmits a block of data symbols concurrently over one OFDM symbol. An OFDM symbol is the time used to transmit all the subcarriers that are modulated by the block of input data symbols.

The baseband modulator converts the input binary bits into a set of multi-level complex numbers that corresponds to different modulations schemes such as BPSK, QPSK, 16-QAM or 64-QAM. The type of modulation scheme that is being used at a particular moment depends on the SNR (Signal to Noise Ratio) of the received signal and the receiver’s ability to decode them correctly (the higher the SNR, the higher the order of the modulation scheme, i.e BPSK is used in very noisy channels). These modulated symbols are then mapped to subcarriers. An IDFT (Inverse Discrete Fourier Transform) block is used to transform the modulated subcarriers in frequency domain to time domain samples.

Generally, the same modulation format is used in all subcarriers to keep the control information overhead small. In harsh and time varying channel conditions it is possible to have different modulation formats over multiple subcarriers. It is known that in a broadband system, the channel is frequency selective over the large system bandwidth, meaning that the signal fading on each subcarrier is independent. Also, the level of interference on each subcarrier can also be different and varies uniquely with time. Thus, we have a different SNR on each of the subcarriers. Hence, adapting the modulation scheme on each of these subcarriers depending on channel conditions (which represents channel dependent scheduling) would help to maximize the overall system throughput.

The orthogonality of OFDM subcarriers can be lost when the signal passes through a time-dispersive radio channel due to inter-OFDM symbol interference. However, a cyclic extension of the OFDM signal is performed to avoid this interference. In cyclic prefix extension, the last part of the OFDM signal is added as cyclic prefix at the beginning of the OFDM signal, making it periodic. The cyclic prefix length is generally chosen a bit larger than the maximum delay spread of the wireless channel. For example, the LTE system uses an OFDM symbol of 66 microseconds plus another 5 microseconds of cyclic prefix. This means that it is susceptible to a maximum dispersion of 5 microseconds due to multi-path propagation.

At the receiver, the prefix part of the symbol is discarded as it may contain ISI (Inter Symbol Interference) from its previous symbol. Hence, the effect of ISI caused by the multipath signal propagation is bypassed. However, the prefix represents the overhead in an OFDM system, as it does not carry any useful information.

The block of complex samples are then serialized in the time domain and converted to analog signals. The RF section modulates the I-Q samples and prepares the signal to be sent.

A corresponding receiver does the inverse operations of the transmitter in the reverse order. A typical OFDM receiver includes a radio frequency I-Q demodulator, analog-to-digital converter, serial-to-parallel converter, cyclic prefix remover, discrete Fourier transformer, sub-carrier de-mapper, equalizer and detector.

This technology has a series of advantages, among which of noteworthy importance we can find:

- High spectral efficiency in comparison with other modulation schemes with double sideband, spread spectrum, etc.
- It can easily adapt to harsh radio conditions without a complex equalization in the time domain
- Robustness to co-channel narrowband interferences
- Robustness to ISI and fading caused by multi-path propagation
- Low sensitivity at temporal synchronization errors

Of course, there are also disadvantages, the most disturbing being:

- The closely spaced subcarriers makes OFDM sensitive to frequency errors and phase noise
- For the same reason as the one stated above, OFDM is also sensitive to Doppler shift, which causes interference between subcarriers
- High values for PAPR, which needs linear transmitter circuits, that have low efficiency
- Efficiency loss caused by the cyclic prefix/guard interval

- The modulation with double sideband of each subcarrier determines a low spectral efficiency and requires greater power for the transmitter in order to obtain the same coverage as obtained with vestigial sideband modulation [22].

With standard OFDM, very narrow UE-specific transmissions can suffer from narrowband fading and interference. That is the main reason why 3GPP chose for the downlink OFDMA, which incorporates elements TDMA.

OFDMA is a multi-user version of OFDM, where the multiple access is realized through the assignation of subcarriers to individual users. Thus, simultaneous transmissions are possible, but with low rates from multiple users. As a result, a more robust system with increased capacity is obtained. This is due to the trunking efficiency of multiplexing low rate users and the ability to schedule users by frequency, which provides resistance to frequency-selective fading.

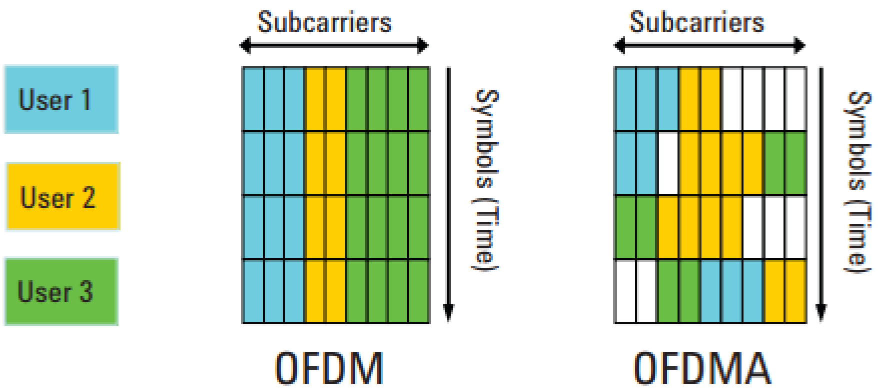

Figure 3.2 OFDM and OFDMA subcarrier allocation [23]

A different number of subcarriers can be allocated to various users, in order to be able to provide differential QoS, namely to control de binary data rate and error probability individually for each user, as it is shown in the figure above.

OFDMA can be seen as an alternative to combining OFDM with TDMA or TDSM (Time Division Statistical Multiplexing), namely communication through packets. Users with low data rate can emit continuously with low power, rather than use a high power carrier in bursts.

In order to keep the overhead to a reasonable dimension, the time/frequency resource grid is subdivided into logical basic resource units which are composed of sub-bands in frequency (a sub-band contains several subcarriers) and one or more OFDM symbols in time domain. The basic resource units are then mapped to the physical OFDMA frame as illustrated in the figure below. If the subcarriers and symbols suffer a permutation in frequency and optionally in time domain, generating a distributed scheme, the data is spread over the carrier frequency, making thus use of frequency diversity, which reduces the probability that a whole data block is affected by frequency selective fading or interference in a sub-band.

In contrast, when it comes to the contiguous permutation scheme a one-to-one mapping is realized between the logical resource units and the corresponding physical subcarriers, allowing thus the exploitation of multi-user diversity by frequency-selective scheduling. However, inter-cell interference must then be avoided by frequency partitioning or by using various coordination schemes. Both WiMAX and LTE specify these permutation schemes. Presently, only the distributed permutation scheme is used in commercial networks due to its lower requirements on channel feedback and lower scheduling complexity.

In OFDMA networks, cell edge users are constrained by interference. In order to resolve this problem, interference mitigation schemes are employed. The most basic schemes are simple frequency reuse schemes combined with frequency diversity permutations. In frequency reuse, the available spectrum is divided into parts which are afterwards assigned to cells/sectors such that the distances between cells with the same spectrum partition are maximized in order to keep the interference level below a certain threshold.

Conventional frequency reuse schemes segment the available spectrum in fixed-sized parts, this partitioning being independent of the measured interference at the mobile stations. Fractional frequency reuse schemes deal with this problem by assigning the whole frequency spectrum to users at the cell center, while maintaining a higher-order frequency reuse scheme for users at cell-edge. In LTE, this is achieved by different power distributions across the frequency spectrum.

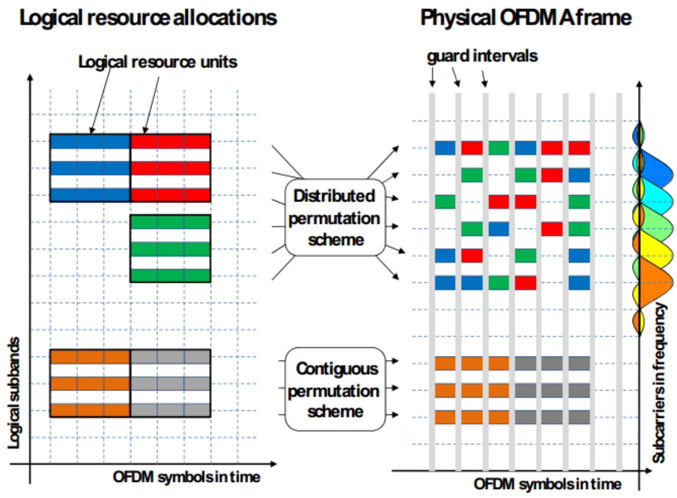

Figure 3.3 OFDMA resource allocation and frame permutation schemes [24]

Another interference mitigation technique is beamforming by making use of antenna arrays. Basically, beamforming means focusing the antenna lobe on the mobile device by a phase delaying introduced at the antenna constitutive elements. In the downlink, this leads to a narrower antenna pattern towards the mobile station, in this manner the interference in other directions being reduced. In the uplink, the array gain obtained from the addition of the signals at the antenna elements also increases performance. Furthermore, uplink interference from adjacent sites can be cancelled by placing nulls into the direction of the main interferers [24].

OFDMA offers a series of advantages:

- Flexibility in the deployment across various frequency bands with minimum changes to the radio interface
- Mitigation of neighbor-cell interference by using carrier permutation among the users of different cells
- Interferences within the cell are mitigated through allocations with cyclic permutations
- Offers frequency diversity through the spreading of the carriers in the entire spectrum of use
- Supports adaptive modulation (QPSK, 16-QAM, 64-QAM, 256-QAM) for each user
- Uses short bursts per user of approximately 100 symbols for better statistical multiplexing and a lower jitter

Among the disadvantages of OFDMA, we can mention:

- High sensitivity to frequency offsets and phase noise
- The complex electronics behind OFDM technology, including the FFT algorithm and forward error correction mechanism are always active, regardless of the transfer rate, which is inefficient from the point of view of power consumption

- Dealing with the co-channel interference from nearby cells is more complicated than in CDMA, requiring dynamic channel allocation and coordination among the adjacent eNBs
- High PAPR, having typical values of 10-12 dB (according to the number of subcarriers), requiring that power amplifiers in the transmitter chain to exhibit a linear output over a large range of frequencies, decreasing thus the power efficiency, and hence increasing the energy consumption of an OFDMA terminal. In order to counteract this effect, PAPR reduction schemes should be implemented. Moreover, increased chipset complexity due to the implementation of advanced receiver designs like MIMO techniques lead to higher energy consumption, compared to 3G or 2G systems.

3.2 Multiple access technology in the uplink: SC-FDMA

A major drawback of OFDM modulation, as well as any kind of multi-carrier transmission, is the large variations in the instantaneous power of the transmitted signal. Such power variations imply a reduced power-amplifier efficiency and higher power-amplifier cost. This is especially critical for the uplink, due to the high importance of low mobile-terminal power consumption and cost. This is the main reason why 3GPP LTE adopted in the uplink SC-FDMA, which keeps most of the benefits of FDMA, but with a reduced PAPR, making it suitable for uplink transmission by user terminals, because energy consumption is a major issue for mobile stations. SC-FDMA is often viewed as DFT (Discrete Fourier Transform) coded OFDM where time-domain data symbols are transformed to frequency-domain by a discrete Fourier transform before going through the standard OFDM modulation, such that each data symbol is transmitted in parallel over a group of subcarriers. The main advantage of OFDM, as for SC-FDMA, is the robustness against multipath signal propagation, making it suitable for broadband systems.

The uplink transmission scheme is based on SC-FDMA, more specifically DFTS-OFDM (Discrete Fourier Transform Spread OFDM), the uplink sub-carrier spacing being Δf = 15 kHz. There are two cyclic-prefix lengths defined: normal cyclic prefix and extended cyclic prefix corresponding to seven and six SC-FDMA symbols per slot. The cyclic prefix timings for uplink are the same as for the downlink.

Despite its name, SC-FDMA also transmits data over the air interface in a multitude of sub-carriers. The SC-FDMA and OFDM transceivers have a very similar structure, with the observation that the former has a new DFT block before the subcarrier mapping. Hence, SC-FDMA can be considered as an OFDM system with a DFT mapper. The structure of a SC-FDMA transceiver is illustrated in the figure below.

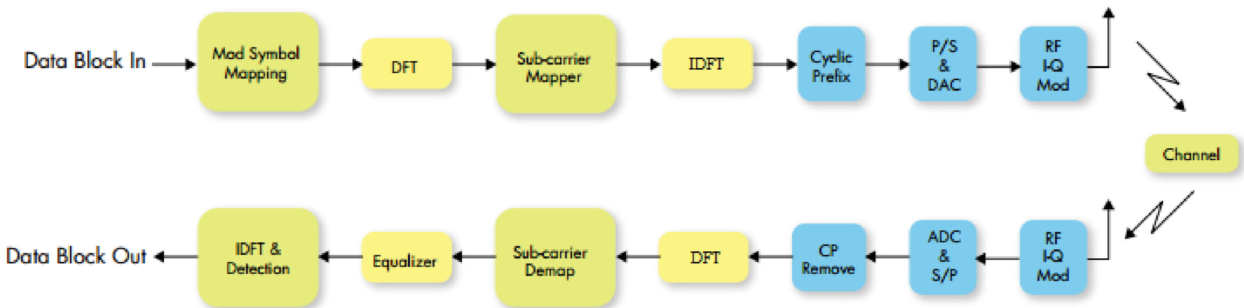

Figure 3.4 SC-FDMA transceiver [25]

After the mapping of input data bits into modulation symbols from a specific modulation alphabet (e.g 4 bits for 16-QAM modulation), the transmitter groups the modulation symbols into a block of M symbols, which is the applied to a size M DFT block that has the purpose to port the symbols in time domain into frequency domain. The output of the DFT is the applied to the input of a size N IDFT bloc, where usually N>M and the unused inputs of the IDFT are set to zero. This block is used to generate the time-domain samples of these subcarriers. Similar to OFDM, a cyclic prefix inserted for each transmitted block, allowing thus straightforward low-complexity frequency-domain equalization at the receiver side. Afterwards, a parallel to serial converter follows, the transmission chain being ended by digital-to-analog converter and RF subsystems.

If the DFT size M would be equal to the IDFT size N, the cascaded DFT and IDFT blocks will neutralize each other. However, if N>M and the unused inputs to the IDFT are set to zero, the IDFT block will have at the output a signal with ‘single-carrier’ properties, i.e. a signal with low power variations, and with a bandwidth that depends on M. More specifically, assuming a sampling rate F_s at the output of the IDFT, the nominal bandwidth of the transmitted signal will be $BW = (M/N) * F_s$. Thus, by varying the DFT block size M the instantaneous bandwidth of the transmitted signal can be modified, allowing for flexible-bandwidth assignment. Furthermore, by shifting the IDFT inputs to which the DFT outputs are mapped, the transmitted signal can be shifted in the frequency domain.

Each data symbol is DFT transformed prior to being mapped to subcarriers, this is why the SC-FDMA is called DFT-precoded OFDM. In OFDMA we use multiple subcarriers to transmit data, each of the subcarriers being modulated by a different data symbol, this symbol lasting relatively long. SC-FDMA is practically a multi-carrier technique, all subcarriers in the uplink being modulated by the same data, but the symbol lasts for a shorter period of time. Basically, multiple subcarriers carry each data symbol due to mapping of the symbols’ frequency domain samples to subcarriers. Because each data symbol is spread over multiple subcarriers, SC-FDMA offers frequency diversity gain in a frequency selective channel.

The DFT output of the input data symbols is mapped to a subset of subcarriers, a process which is called subcarrier mapping, which can be of two types: localized mapping and distributed mapping. In localized mapping, the DFT outputs are mapped to a subset of consecutive sub-carriers, restraining them in this manner to only a fraction of the system bandwidth. In distributed mapping, the DFT outputs are assigned to subcarriers over the entire bandwidth in a non-continuous manner, resulting in zero amplitude for the remaining subcarriers. A particular case of distributed SC-FDMA is called interleaved SC-FDMA, where the occupied subcarriers are equally spaced over the entire bandwidth. The distributed SC-FDMA is more robust to frequency selective fading and offers additional frequency diversity gain because the information is spread across the entire system bandwidth. Localized SC-FDMA must be used in combination with channel-dependent scheduling in order to offer multi-user diversity in frequency selective channel conditions.

It can be shown that the SC-FDMA baseband time domain samples after IDFT is the initial group of data symbols repeated with the symbol period in time domain, as it can be seen in the figure below.

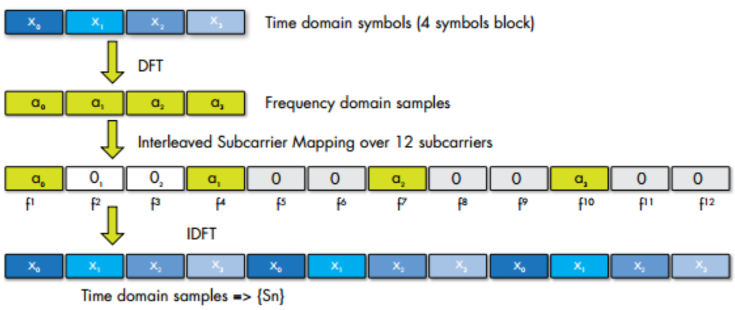

Figure 3.5 Time domain representation of interleaved SC-FDMA [25]

The main advantage of SC-FDMA is a low PAPR of the transmitted signal. PAPR has a direct relationship with the efficiency of the transmitter's amplifier, the maximum power efficiency being obtained when the amplifier works in the saturation point. A lower PAPR makes the use of the power amplifier near the saturation point possible. As a result, a greater efficiency is obtained. At a signal with a high PAPR, the working point of the power amplifier must be lowered in order to diminish the distortions of the signal, decreasing also the efficiency of the amplifier. The PAPR of SC-FDMA in comparison with the one corresponding to OFDMA is improved with about 3 dB, which means that power consumption in terms battery life can be extended by at least two times (3 dB means half the power) [26].

A graphical comparison between OFDMA and SC-FDMA is shown in the figure below. In this example, four subcarriers (M=4) over two symbol periods with the data represented by QPSK modulation are used. What is worth mentioning is that real LTE signals are allocated in units of 12 adjacent subcarriers.

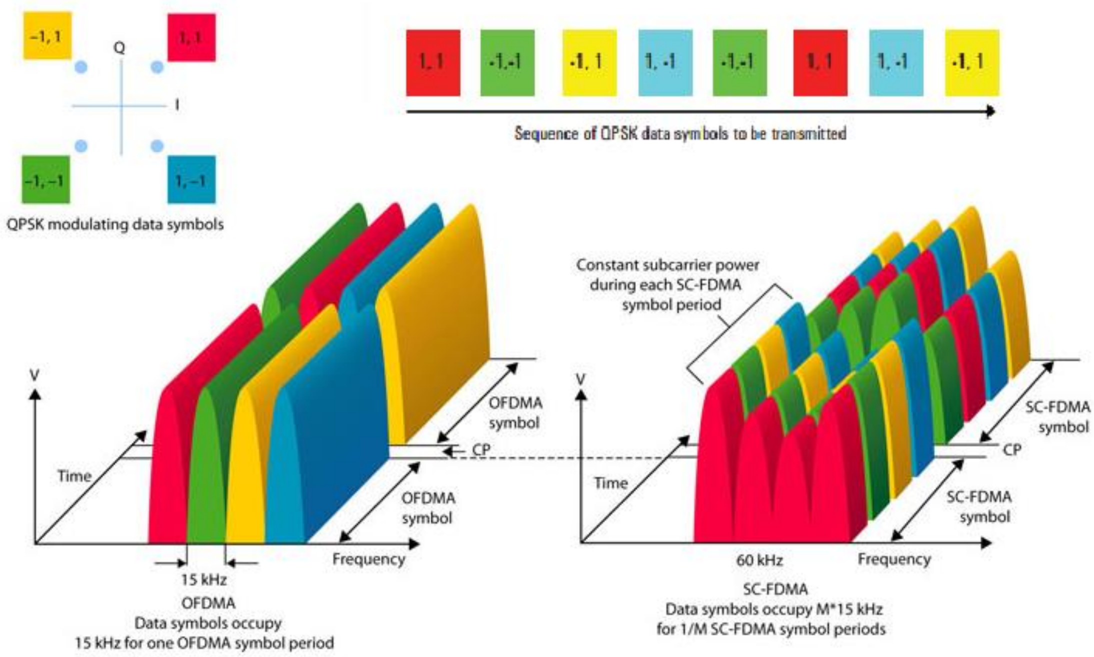

Figure 3.6 Comparison between OFDMA and SC-FDMA transmitting QPSK symbols [27]

Each one of the four adjacent 15 kHz subcarriers is modulated for the OFDMA symbol period of 66.7 µs by one QPSK data symbol. In this four subcarrier example, four QPSK data symbols are processed in parallel, only the phase of each subcarrier is modulated and as a consequence the subcarrier power remains constant between symbols. After each OFDMA symbol period, the cyclic prefix is inserted and the next four symbols are transmitted in a parallel manner. Graphically, the cyclic prefix is shown as a gap, but in reality it contains a copy of the end of the next symbol, which means that the transmission power is continuous but has a phase discontinuity at the symbol boundary. To obtain the transmitted signal, an IDFT is performed on each subcarrier to create M time domain signals, which are then vector-summed to create the final time-domain waveform used for transmission.

In the case of SC-FDMA the process begins with a special pre-coding process but afterwards is subjected to the same stages as in OFDMA. The most obvious distinction between the two schemes is that OFDMA transmits the four QPSK data symbols in parallel, one for each subcarrier, while SC-FDMA transmits the four QPSK data symbols in series at four times the rate, with each data symbol occupying a bandwidth that is M times larger than the subcarrier bandwidth (M x 15 kHz). Thus, the OFDMA signal is clearly multi-carrier with one data symbol for each subcarrier, with a symbol period that is relatively long, whereas the SC-FDMA signal appears to be more like a single-carrier transmission with each data symbol being represented by one wideband signal, but for a much shorter symbol period. Even if the OFDMA and SC-FDMA symbol lengths are the same at 66.7 μs; the SC-FDMA symbol contains M “sub-symbols” that represent the modulating data. The high PAPR of OFDMA is caused by the parallel transmission of multiple symbols. By transmitting the M data symbols in series at M times the rate, SC-FDMA occupies the same bandwidth as OFDMA but, the PAPR is identical with the one used for the original data symbols. In OFDMA, by adding together many narrowband QPSK waveforms will yield higher peaks than in the case of SC-FDMA [28].

3.3 Multiple Input Multiple Output

The way the available spectrum is used is changed through several approaches, including MIMO. Release 8 of the 3GPP specifications, which specifies the principles towards 4th generation systems (LTE), includes new operation requirements, where a base station and handset communicate by making use of two or more transmit and receive antennas and exploits advantageously the differences in radio transmission paths that occur between them. The goal is to increase both the cell’s overall capacity and the data rate that a single user experiences in the network.

MIMO radios exploit the differences that occur in the paths between the transmitter and the receiver. If in a conventional single-channel radio system one data “link” is created between the transmitter and the receiver, in a MIMO radio system multiple such links are created. This is done by creating a mathematical model of the paths from transmitters to receivers. The subsequent equations that arise must be solved as fast as the channel is changing. If the data links can be completely segregated, the channel capacity increases linearly as more transmitter-receiver links are added.

MIMO also represents a new challenge for network operators because unlike traditional cellular networks which provide the best service under line-of-sight conditions, MIMO flourishes under rich scattering conditions, where signals from different Tx antennas take multiple paths to reach the UE at different times. In order to achieve the promised throughputs in LTE systems, operators must fine-tune their networks’ multipath conditions for MIMO, targeting both rich scattering conditions and high SNR for each multipath signal. Good multipath conditions create the signal orthogonality that influences directly the amount of data that can be sent on a particular frequency band.

There are four ways to make use of the radio channel: SISO, SIMO (Single Input Multiple Output), MISO (Multiple Input Single Output) and MIMO.

SISO is the most basic radio channel access mode, making use of only one transmit antenna and one receive antenna. This access mode has been used since the beginning of radio communications and represents the baseline against which all the multiple antenna techniques are compared.

SIMO uses one transmitter and two or more receivers, being often referred to as receive diversity. This radio channel access mode is well suited for low SNR conditions in which a theoretical gain of 3 dB is obtainable when two receivers are used. There is no change in the data rate because only one data stream is transmitted. There are two types of SIMO that can be used:

- **Switched diversity SIMO**: it looks for the strongest signal and switches to the corresponding antenna.
- **Maximum ratio combining SIMO**: This form of SIMO takes both signals and sums them up, making the signals from both antennas to contribute to the overall signal.

MISO mode uses two or more transmitters and one receiver. It is often referred to as transmit diversity. The same data is sent on the transmitting antennas but coded such that at the receiver each transmitter is recognizable. Transmit diversity increases the robustness of the signal to fading and can increase performance in low SNR conditions. It does not increase the data rates, but it supports the same data rates with a greater power-efficiency. Transmit diversity can be improved with closed loop feedback from the receiver, which means that the receiver indicates to the transmitter the optimum equilibrium between phase and power that is to be used for each transmit antenna.

MIMO requires two or more transmitters and two or more receivers. It achieves the increase in spectral capacity by transmitting multiple data streams simultaneously on the same frequency, exploiting the different paths taken within the radio channel. In order for a system to be described as MIMO, it must have at least as many receivers as there are transmitted streams. True MIMO, with two transmitters and two receivers with independent data content, is also known as spatial multiplexing. Each receiver interprets the output of the channel as a combination of the outputs from the transmitters. The receivers use channel estimation techniques and matrix mathematics to separate the two data streams and demodulate the data. Ideally, in the case of 2×2 MIMO (2 Tx and 2 Rx) antenna configuration the throughput would be doubled but in practice, the doubling of data capacity is never achieved, but considerable increases can be easily observed.

MIMO can combine SIMO and MISO techniques in order to obtain greater SNR gains, by this ensuring improved coverage and data rates. However, when SNR is high, additional throughput gains are minimal, because the relationship between SNR and throughput has a logarithmic allure, obtaining only a little benefit from further increasing SNR. In order to obtain throughput gains where SNR is already very high, LTE uses spatial multiplexing.

Supplementary to good multipath conditions, spatial multiplexing relies on high values of SNR to produce large throughput gains. In spatial multiplexing, the total power of the transmission remains the same, even if multiple data streams are transmitted, the total SNR being distributed between these multiple data streams, each of them having a lower power level. Because the throughput increase is very small when the SNR is already high, each of the multiple data streams may be capable of transmitting nearly as much data as a single stream.

In multiple antenna techniques, the incoming data is split in order to be transmitted to multiple antennas. The incoming data may be split in two or more parallel streams, where each stream contains a different portion of the incoming data, or it may be duplicated, and each antenna transmits basically the same information. The only difference between each stream in this case is the mathematical manipulation of the transmitted data which does not change in any way the information included in it. The following terms are thus used:

- **Transmission layer**: one data stream transmitted through one antenna. The number of transmission layers equals the number of antennas
- **Spatial layer**: a data stream which includes unique information, not included at the other layers
- **Rank**: the number of different transmission layers transmitted in parallel or, in other words, the number of spatial layers.

The potential capacity and coverage enhancement is a function of the number of antennas and the rank in use. In the right side of the graph below, rank 1 scenarios with different number of antennas are shown. It describes scenarios where all antennas transmit the same information. As it can be seen, rank 1 antennas do not improve capacity but extend coverage as more transmission antennas are used. These are called transmit diversity techniques. On the left side of the graph, spatial multiplexing techniques are shown, where capacity improvement increases as the number of different transmission streams gets higher. However, the higher the rank, the higher the required SNR required for decoding this signals and as a result, the coverage range gets shorter.

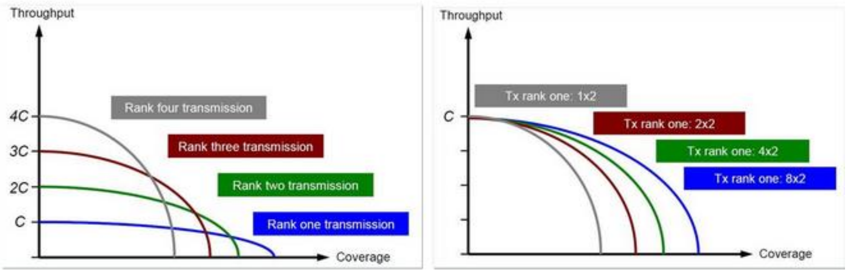

Figure 3.7 Throughput and coverage range versus rank [29]

Let us consider the case of a system made of two Tx and Rx antennas, where the transmitter transmits two streams, X_1 and X_2. Ideally, only the X_1 signal will be received at receiving antenna A_{R1}, and only X_2 signal will be transmitted at receiving antenna A_{R2}. However, X_1 is received at A_{R2} and X_2 is received at A_{R1} due to multi-path propagation. If the receiving antennas are not correlated, it is possible to exploit the signals arriving from multiple reflections and decode X_1 and X_2. Known reference signals are sent from the transmitter to the receiver. By decoding these signals and comparing them to a known pattern, the receiver is also able to characterize the channel and estimate the h values of the matrix, which represent the channel coefficients: h_{11} is the coefficient describing the channel between the antenna T_{X1} transmitting the X_1 signal and A_{R1}, whereas h_{12} stands for the channel between A_{R1} and T_{X2}. The equation below models the received signal taking into account the additive noise present in the channel and the sent signal:

$$\begin{bmatrix} Y1 \\ Y2 \end{bmatrix} = \begin{bmatrix} h11 & h12 \\ h21 & h22 \end{bmatrix} \begin{bmatrix} X1 \\ X2 \end{bmatrix} + n$$

LTE uses feedback mechanisms known as pre-coding and beamforming – both procedures of "closed-loop MIMO", where the handset requests changes to the cross-coupling of the transmitter outputs in order to improve the ability of the receiver to decode the received signal.

The terms codeword, layer, precoding and beamforming have been agreed upon specifically for LTE to refer to signals and their processing, being used in the following way:

- **Codeword**: represents the user data before it is formatted for transmission. One or two codewords, CW0 and CW1, can be used depending on the predominant channel conditions and case of use. In the most common case of SU-MIMO (Single User MIMO), two codewords are sent to a single UE, but in the case of the downlink MU-MIMO (Multi-User MIMO), each codeword is sent to only one UE.
- **Layer**: for MIMO, at least two layers (streams) must be used, and up to four are allowed. The number of layers is always less than or equal to the number of antennas.
- **Precoding**: sometimes, a tight feedback from the user to the transmitter is needed. Based on the momentary characteristics of the channel, the UE sends the transmitter a feedback which contains a "recommendation" for the activation of a specific precoding matrix at the transmitter. The precoding modifies the signal before being transmitted for improving the ability of the receiver to decode the received MIMO signals, coping with low SNR. This feedback is called PMI (Precoding Matrix Indicator). Since this process should track the channel's dynamics, antenna techniques that require PMI feedback can only serve slow moving users. The precoding choices are well-defined in a lookup table known as codebook, which is used to quantize the available options, limiting in this way the amount of information sent from the receiver to the transmitter. This means that the codebook index provides an approximation to the channel, implying some level of residual error. The UE sends a message to the eNB scheduler with the codebook index that matches as close as

possible the channel conditions at that particular moment. The scheduler has to respond rapidly, within milliseconds, because the radio channel is a fast changing one. The more channel information is provided by the UE, the more accurate the information will be, but also more resources will be used for signaling.

- **Beamforming** modifies the transmit signals to give the best CINR (Carrier to Interference and Noise Ratio) at the channel output

The theoretical throughput gains from MIMO depend on the number of Tx and Rx antennas, the radio propagation conditions, the ability of the transmitter to adapt in a timely manner to the changing conditions of the channel, and last, but not least, the SNR. Ideally, the paths in the radio channel are completely uncorrelated, almost as if separate. Such conditions are almost impossible to be obtained in free space. As a general rule, MIMO works best in high SNR conditions with minimal line of sight. Line of sight means high channel correlation and seriously reduces the potential for gains. Consequently, MIMO is particularly suited to indoor environments, which can exhibit a high degree of multi-path propagation and limited line of sight.

SU-MIMO and MU-MIMO are used in the uplink. In order to implement SU-MIMO, the UE requires two transmitters, which represents a tremendous challenge in terms of cost, size and battery consumption, all of these needing to be as small as possible for a mobile handset. For this reason and for the fact that the increased data rates in the uplink that SU-MIMO will offer are not as important as they are in the downlink due to asymmetrical traffic distribution, SU-MIMO is not a priority for development.

The UE has usually a single transmitter in its baseline configuration, being still capable of supporting a new form of MIMO, namely MU-MIMO. Unlike the receive function, MIMO does not require that the transmitters are in the same physical device or location. Thus, uplink MIMO can be implemented using two transmitters that belong to two different UEs, creating thus the potential for an increase in uplink capacity, even if individual users will see no increase in data rates. MU- MIMO involves the simultaneous transmission of codewords via layers from different UEs at the same time and frequency. The fact that the transmitters are physically separate has two consequences: firstly, precoding is not possible since the source data cannot be shared between the two UEs to create the necessary orthogonality of the two data streams, reducing in this way the potential gains that co-located transmitters may have had. Secondly, the separation of the transmitters increases the likelihood that the radio channels as seen by the eNB will be uncorrelated. The potential gains that are lost through lack of precoding will be more than compensated for by the gains that are more likely to be obtained from better channel de-correlation, making MU-MIMO a valuable technique for improving uplink capacity [30].

As mentioned earlier, the 3GPP Release 8 standard for LTE defines several multiple antenna techniques called, according to 3GPP terminology, antenna configuration or transmission modes. These transmission modes differ from each other by the benefit they provide, which can be translated to range or capacity, and by the conditions required by their operation.

In LTE, seven transmission modes are supported in order to improve both link- and system-level performance in a wide range of scenarios. These transmission modes allow up to four layers to be transmitted in the downlink.

1. Single antenna (SIMO); suitable for fast users
2. Transmit diversity (MISO); enhances gain, for fast users
3. OLSM (Open Loop Spatial Multiplexing), enhances capacity, for fast users
4. CLSM (Closed Loop Spatial Multiplexing), enhances capacity, for slow users
5. MU-MIMO; enhances capacity, for fast users
6. Closed loop single layer precoding (MISO); enhances gain, for slow users
7. Linear array beamforming (MISO); enhances gain, for fast users

The UE ability to provide detailed and timely information on its channel conditions determines the communication to be operated in either closed loop or open loop transmission mode. An open loop communication between the eNB and the UE is established when the UE is moving too fast to provide a thorough report on channel conditions in a timely manner for the eNB to select the precoding matrix. Even when the UE is moving relatively slowly, other factors as, for example the processing speed of the UE or uplink data capacity may result in open loop operations. In open loop communications, the eNB receives minimal information from the UE: a RI (Rank Indicator), which represents the number of layers that can be supported given the current channel conditions and the employed modulation scheme; and a CQI, which represents a summary of the channel conditions under the current transmission mode, roughly corresponding to SNR. The eNB uses the CQI to select the correct modulation and coding scheme for the corresponding channel conditions. The eNB selects its transmission mode and the amount of resources dedicated to the UE based on whether the CQI and RI reported by the UE match the expected values, and whether the signal is being received at an acceptable BER (Bit Error Rate).

In closed loop operations, the UE examines the channel conditions of each Tx antenna, including the multipath conditions, providing the eNB with an RI as well as PMI to assist the selection of the optimum precoding matrix for the current channel conditions. In closed-loop testing lost packets are retransmitted using incremental redundancy based on real-time packet acknowledgement feedback from the receiver. Finally, the UE provides a CQI given the RI and PMI, allowing the eNodeB to adapt to the transmission channel conditions in a time-effective manner. Closed loop operations are predominantly important for spatial multiplexing, where MIMO offers the most considerable throughput gains.

The first transmission mode is a simple single layer transmission using a single transmitting antenna. It is the only transmission mode provided in LTE specifications which is not a multiple antenna technique. It might be a SISO or SIMO configuration, where the UE may have multiple antennas for enhancing signal quality using algorithms like MRC (Maximum Ratio Combining).

Transmission mode 2 is a rank 1 technique where all data streams transmit the same information. This yields in a transmit diversity which provides additional robustness to the link but does not increase the throughput. Antennas should be with low spatial correlation. This can be achieved by cross-polarized antennas (antennas which are orthogonally polarized) or with large inter-antenna distances (of at least 5-6 wavelengths). This transmission mode does not require closed loop PMI feedback and therefore can support high speed users.

Transmission mode 3 is a spatial multiplexing technique with two transmitting antennas, where different information is transmitted on each of the transmission layers, the rank being therefore 2. It uses a technique called CDD (Cyclic Delay Diversity), where instead of providing a tight, closed-loop PMI feedback a predefined codebook of pre coder matrices is cycled across the frequency band, along with a layer permutation designed to give to each layer similar average channel quality. No feedback is required to select the precoders, this mode being thus suitable in scenarios where timely channel dependent feedback cannot be made available like, for instance, in high-speed scenarios.

Transmission mode 4 is a spatial multiplexing technique with up to 4 transmitting antennas, where different information is transmitted on each of the transmission layers. The rank is therefore between 2 to 4, depending on the number of antennas. A closed loop PMI feedback is used to specify the required precoding matrix. Therefore, this model is used only for slow-moving users. Antennas should be with low spatial correlation, achieved by orthogonally-polarized antennas or with intra-antenna distance of at least 5-6 wavelengths. Depending on the configured uplink feedback mode, the UE may feedback:

- Multiple preferred PMIs, where each PMI is valid for a particular sub-band (frequency selective precoding)
- One PMI that is valid for all sub-bands (wide-band precoding)

Transmission mode 5 serves simultaneously multiple users using the same frequency resources. The users are separated in the space domain by individual beams directed to each user. The number of users is limited by the number of spatial layers. Each user should have a dedicated spatial layer. Transmitting antennas used with this transmission mode are linear antenna arrays.

Transmission mode 6 may be regarded as a single layer (rank 1), closed loop spatial multiplexing technique. This creates a beamforming effect, which provides 1-2 dB improvement over Mode 2 transmit diversity technique. It can also be seen as a special case of SU-MIMO spatial multiplexing. In this mode the UE feeds channel information back to the eNB to indicate suitable precoding to apply for the beamforming operation. It is a closed loop technique, which requires tight feedback from the UE and therefore is suitable for only slow moving or stationary users.

Transmission mode 7 is a rank 1 technique based on beamforming using linear antenna arrays. This mode extends the cell coverage by concentrating the transmitted power in the direction of each served user. Increased SNR is achieved by phase adjustments of the signals transmitted on the different antennas with the aim of making the signals add-up constructively on the receiver side.

Beamforming gain is composed of two parts:

- **Array gain**: $10 \times \log(\text{Number of antennas})$
- **Diversity gain**: 0-5 dB depending on the modulation and multipath severity

This transmission mode implies an open loop technique, therefore it may be used for fast-moving subscribers [31].

4. Channel structure and bandwidths

One of the major objectives of the LTE system is to offer a flexible bandwidth support for deployment in very different spectrum arrangements. Consequently, the LTE physical layer supports bandwidths in 180 kHz increments starting from a minimum bandwidth of 1.08 MHz. In order to provision channel sensitive scheduling and to obtain low transmission latency for packet transmission, the scheduling and transmission interval is defined as a 1 ms subframe. Two cyclic prefix lengths are defined: CP and extended CP, in order to offer support for deployment in the case of small and large cells, respectively. The subcarrier spacing is of 15 kHz and is chosen in order to obtain a balance between the overhead introduced by the cyclic prefix and the robustness to Doppler shift. A smaller subcarrier spacing of 7.5 kHz is defined for MBSFN in order to accommodate large delay spreads with a reasonable-sized CP overhead. In the uplink, transmissions with contiguous resource block allocation are supported due to the use of SC-FDMA. Frequency diversity is obtained by using inter-subframe and intra-subframe hopping.

4.1 The frequency bands intended for LTE. Channel bandwidth

E-UTRA is designed to operate in various operating bands, as illustrated in the figure below.

E-UTRA Operating Band	Uplink (UL) operating band BS receive UE transmit F_{UL_low} – F_{UL_high}	Downlink (DL) operating band BS transmit UE receive F_{DL_low} – F_{DL_high}	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 – 894MHz	FDD
6[1]	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 – 1447.9 MHz	1475.9 MHz – 1495.9 MHz	FDD
12	699 MHz – 716 MHz	729 MHz – 746 MHz	FDD
13	777 MHz – 787 MHz	746 MHz – 756 MHz	FDD
14	788 MHz – 798 MHz	758 MHz – 768 MHz	FDD
15	Reserved	Reserved	FDD
16	Reserved	Reserved	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
20	832 MHz – 862 MHz	791 MHz – 821 MHz	
21	1447.9 MHz – 1462.9 MHz	1495.9 MHz – 1510.9 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
Note 1: Band 6 is not applicable.			

Table 4.1 E-UTRA frequency bands [32]

The LTE system supports a set of six channel bandwidths, as it can be seen in the table below:

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

Table 4.2 Transmission bandwidth configuration N_{RB} in E-UTRA channel bandwidths [32]

The transmission bandwidth configuration BW_{config} is 90% of the channel bandwidth $BW_{channel}$ for the range 3 - 20 MHz. For the channel bandwidth of 1.4 MHz, the transmission bandwidth is only 77% of the channel bandwidth, making LTE deployment in the small 1.4 MHz channel less spectrally efficient than in the 3 - 20MHz channel bandwidths. The actual transmission bandwidth is generally smaller than the channel bandwidth to provide the guard bands for the two edges. The spectral efficiency in the downlink is 5 bits/sec/Hz (3-4 times greater than in Release 6 HSDPA), whereas in the uplink is 2.5 bits/s/Hz (2-3 times that of Release 6 HSUPA). According to ITU-T, spectral efficiency is defined as the ratio between the amount of information transferred over a distance and the product of the frequency bandwidth, the geographic space and the time denied to other potential users. For a cellular mobile system, it is defined as SUE= (Traffic in Erlang) / (Amount of spectrum in MHz × Area in km^2).

The channel edges are defined as the lowest and highest frequencies of the carrier separated by the channel bandwidth, i.e. at F_C +/- ($BW_{Channel}$ /2), as it can be seen in the figure below.

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 E-UTRA carriers is defined as following:

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

where $BW_{Channel(1)}$ and $BW_{Channel(2)}$ are the channel bandwidths of the two respective E-UTRA carriers. The channel spacing can be adjusted to optimize performance in a particular deployment scenario.

The channel raster is 100 kHz for all bands, which means that the carrier central frequency must be an integer multiple of 100 kHz.

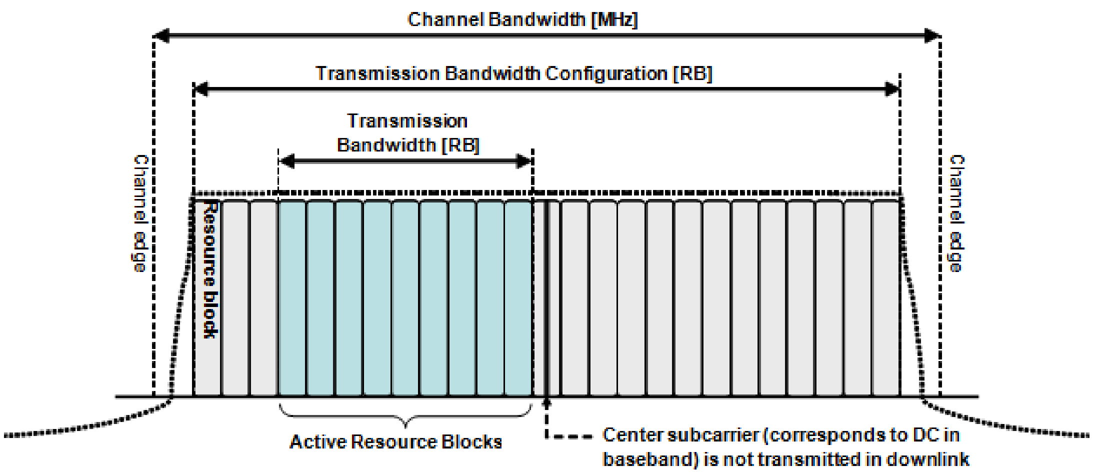

Figure 4.1 Channel Bandwidth and Transmission Bandwidth Configuration for one E-UTRA carrier [32]

4.2 Refarming

Spectrum is without any doubt the scarcest and most valuable of all the resources needed for mobile communications. In order to meet the increasing demand of mobile data, most operators will need to define a high priority on making the most effective use of their finite spectrum assets.

Unfortunately, there is not much spectrum available for LTE, and the price for LTE spectrum is very high, determining many operators to consider the potential of existing frequencies. In most countries, the 1800 MHz band is used for GSM, and it has a lot of free frequency bands that can be used for other mobile technologies like, for example, LTE, gaining in this way extra revenue by offering mobile broadband on a frequency traditionally used for GSM services.

Refarming the GSM band to LTE requires dedicating a portion of spectrum to LTE without affecting the quality and capacity of the GSM service by using more efficiently the spectrum allocated to GSM services, while maintaining or even improving network quality.

3GPP specifies the bandwidth for the 1800 MHz spectrum as 2 × 75 MHz for FDD, with uplink frequencies between 1710-1785 MHz, and downlink frequencies between 1805-1880 MHz. This is a considerable continuous bandwidth that includes frequencies supported in Europe, APAC, India, China, Africa, and the Middle East. 1800 MHz is technically neutral in many countries, which makes it easier to refarm enough spectrum for LTE and increases the potential of 1800 MHz becoming a worldwide roaming spectrum. License allocations on the 1800 MHz band are usually less fragmented than for example on the 900 MHz band, including typically 10-15 MHz of non-fragmented spectrum. By making use of the flexible carrier bandwidth of LTE, all or a portion of the 1800 MHz band can be reallocated for LTE use.

It should also be noted that 1800 MHz spectrum used for LTE, most suitable for wide-area coverage, has some coverage benefit over the new LTE 2600 MHz band, which is suitable for dense urban hotspot coverage: it provides roughly 3 dB better link budget in comparison to LTE 2600 MHz, leading to the doubling of the coverage area, device battery savings or enhanced indoor bit rates.

Installed antennas covering the 1800 MHz band, feeders, site space, power facilities and backhaul transport of GSM1800 can be reused through a variety of known installation practices. Alternative solutions cover both partial refarming to GSM-LTE dual mode or to LTE-only.

There are three possible solutions for LTE1800 refarming:

1. **Refarm the entire GSM1800 spectrum for LTE use**: operators need to swap their existing equipment that operates in 1800 MHz band with one for 900 MHz band, the spectrum refarming being quite straightforward. Usually, GSM1800 traffic is not high. The major issue is to swap the user terminals in order to make use of other GSM frequency bands. This should not pose a real problem, because dual- or triple-band GSM terminals are widely spread nowadays. In order to ensure smooth operation with the existing networks, a complete set of cell reselection mechanisms in idle state and handover types within each radio access technology and between access technologies must be supported. Both intra-frequency and inter-frequency handovers are specified for each technology. Load balancing is needed especially between GSM/EDGE and WCDMA/HSPA, while LTE 1800 MHz refarming typically reduces overall GSM capacity.
2. **Refarm part of the GSM1800 spectrum for LTE use:** in this scenario, LTE1800 is located between GSM1800 frequency bands. Thus, operators have to consider the interferences that will arise between LTE and GSM at the frequency edge. In this case, operators have to improve spectrum utilization in GSM1800 and reduce its bandwidth. They also have to improve spectrum utilization in GSM900 and port the traffic from GSM1800 to GSM900. DFCA (Dynamic Frequency and Channel Allocation) and OSC (Orthogonal Sub-Channel) are key techniques to enhance spectral efficiency of GSM and EDGE. DFCA provides denser reuse of GSM carriers in a reduced spectrum due to LTE refarming. OSC

enables Dual Half Rate concept for voice calls, which allows the transport of up to two times the number of half rate channels, i.e. 4 voice channels per GSM timeslot. In downlink orthogonal transmission by QPSK modulation is applied, in uplink the dual user MIMO concept is used. Over half of all GSM devices support OSC already today.

3. **Refarm part of the GSM1800 spectrum for LTE use**: in this scenario, LTE1800 is located only on one side of GSM1800 frequency band. Operators will have to consider the interference between LTE and GSM at the frequency edge as well as the interference between LTE and other neighboring frequency bands.

The LTE air interface is very scalable in terms of bandwidth, assuming small nominal guard band (10 %, with the exception of the 1.4 MHz case). When LTE1800 is located between GSM1800 frequency bands and the LTE1800 bandwidth is 1.4 MHz or 3 MHz, a 3GPP recommended guard band of 0.2 MHz is needed. If the LTE1800 bandwidth is 5, 10, 15, or 20 MHz, no GSM/LTE co-site guard band is required. In the case when LTE1800 is located on one side of GSM1800 frequency band and the LTE bandwidth is 1.4 MHz or 3 MHz, a guard band of 0.2 MHz is also needed, and interference between LTE and its neighboring frequency band has to be taken into account [33].

From the EPC perspective the 2G/3G/LTE deployment can be done either with combined or separated SGSN/MME elements, the main difference between the two being that the SGSN handles both user plane and control plane while the MME handles only the control plane traffic.

During the traffic migration from 2G/3G to LTE, an increase in the demand for the S/PGW capacity is expected to arise as LTE provides higher spectrum efficiency compared to 2G/3G. In LTE, there is an always-on IP connection between de user users and the EPC, while in 2G/3G the users are typically attached to GPRS without a packet data protocol context. As a consequence, an increased need for bearer and signaling capacity is to be considered in EPC.

The SGSN may be located in the existing site from where the traffic is routed to SGW. The MME location is independent on the user plane, which means that it can be centralized in the data center or combined with the SGSN on a regional level.

When it comes to the interoperability between LTE and 2G/3G, two solutions are available:

1. **CSFB (Circuit Switched Fallback)**: is a technology where voice and SMS (Short Message Service) services are delivered to LTE devices through the use of GSM or another circuit-switched network. It is needed because LTE is a packet-based all-IP network that cannot support circuit-switched calls. When an LTE device is used to make or receive a voice call or SMS, the device "falls back" to the 3G or 2G network in order to complete the call or to deliver the SMS message.
2. **SR-VCC (Single Radio Voice Call Continuity)**: IMS is needed and the cost is very high.

Packet-switched service interoperability requires load sharing between LTE and 3G. This can be implemented in two stages:

1. When the LTE system has much spare capacity, UMTS/LTE dual-mode terminals will first reside in an LTE network where packet-switched traffic is carried.
2. When the LTE system reaches its capacity threshold, a load-balance between LTE1800 and HSPA is performed. Premium LTE users remain in the LTE network for optimum service experience, all their traffic being carried on the LTE network. Other LTE users reside in an HSPA network, load balancing optimization offering them the possibility to reselect the LTE network in order to access latency-sensitive or throughput-sensitive services such as real-time gaming and FTP downloading [34].

4.3 LTE UE categories

The LTE UE classes ensure that eNB can communicate correctly with the UE. By transmitting the LTE UE category information to the eNB, the performances of the UE can be determined by the eNB in order to establish a proper communication between the two entities.

As the LTE class defines the overall performance and the capabilities of the UE, it is thus possible for the eNB to establish a communication using the known capabilities of the UE, avoiding thus the situation of communication beyond the performances of the UE.

The LTE system supports five UE categories with different radio access capabilities, different LTE UE categories having different performances and a wide range of supported parameters. LTE category 1, for example does not support MIMO, but LTE UE category five supports 4x4 MIMO. Thus, category 5 devices have the possibility of having better performance than the other categories, due to better uplink modulation and four receiver antennas. However, it will be very hard to have a mobile phone belonging to category 5 because of great power consumption, thermal dissipation problems and the technical difficulties to fit a 4 x 4 MIMO antenna in a mobile phone.

UE Category	Maximum number of DL-SCH transport block bits received within a TTI	Maximum number of bits of a DL-SCH transport block received within a TTI	Total number of soft channel bits	Maximum number of DL MIMO layers	Maximum number of bits of an UL-SCH transport block transmitted within a TTI	Support for 64QAM in UL	Total layer 2 buffer size [bytes]	Peakrate DL/UL [Mbps]
Category 1	10296	10296	250368	1	5160	No	150 000	10/5
Category 2	51024	51024	1237248	2	25456	No	700 000	50/25
Category 3	102048	75376	1237248	2	51024	No	1 400 000	100/50
Category 4	150752	75376	1827072	2	51024	No	1 900 000	150/50
Category 5	299552	149776	3667200	4	75376	Yes	3 500 000	300/75

Table 4.3 DL and UL physical layer parameter values set by the field UE-Category [35]

The UE categories are differentiated through the maximum supported data rates for downlink and uplink and the number of supported data streams. The modulation scheme used in the downlink is 64 QAM for all classes, whereas for the uplink the first four classes support 16 QAM and class 5 supports 64 QAM.

The total layer 2 buffer size is defined as the sum of the number of bytes that the UE is capable of storing in the radio link control transmission windows, reception and reordering windows for all radio bearers.

It should also be noted that in the first release of the LTE system, MIMO spatial multiplexing is not supported in the uplink.

The parameter soft channel bits indicates the total number of soft channel bits available for HARQ processing and it is used for deciding whether a HARQ retransmission is in “chase combining mode” (retransmitting the same redundancy version) or in “incremental redundancy mode” (retransmitting a different redundancy version).

The UE category is transmitted by the UE to the E-UTRAN during call setup by making use of the RRC protocol. In addition to the UE class, various UE capabilities are defined separately through the feature group indicators, which are transmitted by the UE to the network as a part of the call setup procedure and are used to inform the network of the UE capabilities concerning certain LTE features. These include, for example: the capability of the UE to support inter-frequency handover, periodic measurements for self-optimized networks, inter-RAT measurements, intra-subframe frequency hopping in the uplink, simultaneous transmission of uplink control information and semi-persistent scheduling. By knowing the reported UE class and capabilities, the E-UTRAN can index the different features that can be supported by the user [36].

4.4 Frame and slot structure

3GPP TS 36.211 defines the 3G LTE frame and subframe structure for the E-UTRA. Through these, the LTE network can maintain synchronization and is able to manage the different types of information needed to be carried between the eNB and the UE.

The LTE air interface has two major components: physical signals and physical channels. Physical signals are generated in Layer 1 and are used for various purposes, including: system synchronization, cell identification (allowing thus the UE to identify and synchronize with the network), and radio channel estimation. The main purpose of physical channels is to carry data from higher layers including control, scheduling, and user payload. Physical channels carry the user and system information.

Even if the LTE downlink and uplink use different multiple access schemes, they have a common frame structure, which defines the frame, slot, and symbol in the time domain. Two radio frame structures are defined for LTE, as shown below, differing between the TDD and the FDD modes because there are different requirements on separating the transmitted data:

- **Type 1:** used for the LTE FDD mode systems.
- **Type 2:** used for the LTE TDD systems.

4.4.1 Frame structure type 1

Frame structure type 1, as illustrated in the figure below, is applicable to both full duplex and half duplex FDD. In half-duplex FDD operation, the UE cannot transmit and receive at the same time. However, there are no such restrictions in full-duplex FDD. The duration of one LTE radio frame is 10 ms (T_{frame}=307200×T_S). One frame is divided into 10 subframes of 1 ms each ($T_{subframe}$=30720×T_S), each subframe being divided into two consecutive slots of 0.5 ms each (T_{slot}=15350×T_S). Each slot contains either six or seven OFDM symbols, depending on the CP length. Uplink and downlink transmissions are separated in the frequency domain.

All the time durations are defined in terms of the sample period T_S=1/f_S, where f_S =30.72 Msamples/sec. The useful symbol time is the inverse of the subcarrier spacing (which is usually 15 Khz) 1/15 kHz= 66.6 μs. In the case of a normal CP, the first symbol in the slot has a CP length of T_{CP}=160×T_S= 5.2 μs, whereas the following symbols have a CP length of T_{CP}=144×T_S=4.69 μs. The reason why different CP lengths of the first symbol are available is to make the slot length in terms of time units divisible by 15360. Thus seven symbols can be placed in the 0.5-ms slot as each symbol occupies 66.6 + 4.69 = 71.29 microseconds. When extended CP is used, which lasts for $T_{CP\text{-}e}$=512×T_S=16.67 μs and is the same for all symbols, the total symbol time is 66.6 + 16.67 = 83.27 microseconds. Since the length of one slot is fixed and cannot be changed, six symbols can then be placed in the 0.5-ms slot, the number of symbols that can be placed within a slot being decreased. The CP is longer than the typical delay spread of the channel, which is usually of the order of a few microseconds. The normal CP is used in urban cells and high data rate applications while the extended CP is used in cases like multi-cell broadcast and in very large cells (e.g. rural areas).Frames are useful to send system information. Subframes facilitate resource allocation and slots are useful for synchronization. Frequency hopping is possible at the subframe and slot levels.

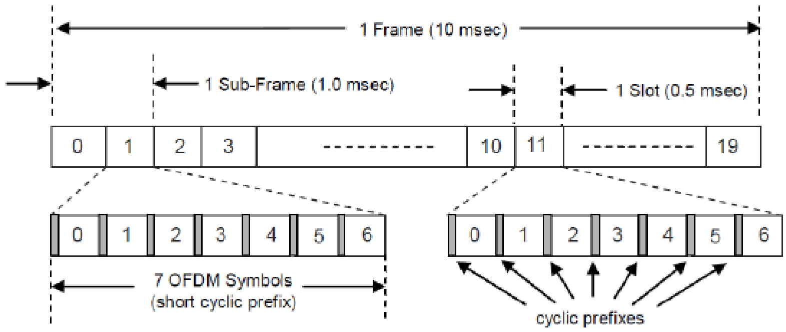

Figure 4.2 Frame structure type 1 [37]

The normal CP represents an overhead of 7.5% of the entire slot payload. One way to reduce the overhead generated CP insertion is to reduce the sub-carrier spacing Δf (to 7.5 kHz), which corresponds to an increase in the symbol time. But this will increase the sensitivity of the OFDM transmission to frequency instability resulting from fast channel variations (for example high Doppler spread) as well as different types of frequency errors due to the non-ideality of electronics.

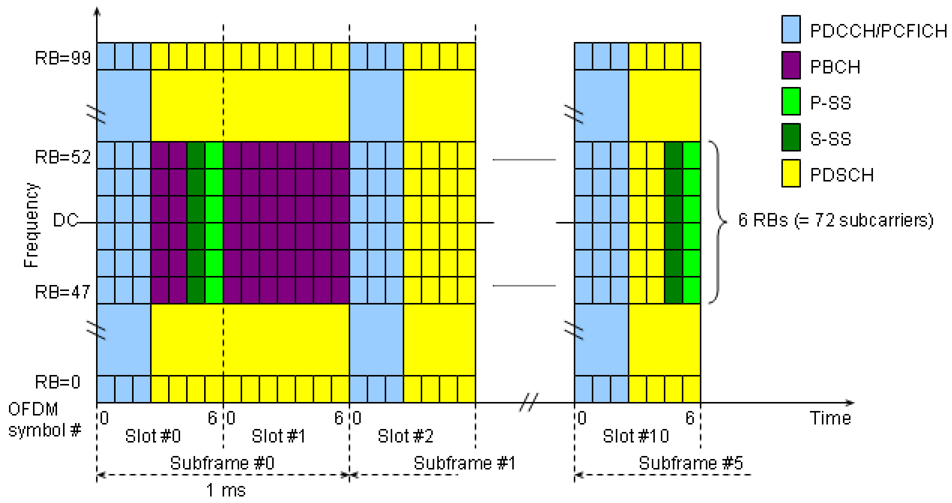

Figure 4.3 FDD frame structure with normal CP and 20 MHz system bandwidth [37]

The LTE signal can be represented in a two dimensional system as shown in the figure above. The horizontal axis is for the time domain, while the vertical one is for the frequency domain. The fundamental unit for the horizontal axis is the symbol, whereas for the vertical axis is the subcarrier.

As already mentioned, the LTE band consists of multiple equally spaced channels called subcarriers. They are placed 15 kHz apart, regardless of the system bandwidth. Thus, if the bandwidth of the LTE channel increases, the number of accommodated channels also increases: for a 5 MHz LTE band we have 300 subcarriers, for a 10 MHz LTE band we have 600 subcarriers.

The smallest bi-dimensional unit in the physical layer is the RE (Resource Element) which comprises one symbol in time domain and 1 subcarrier in frequency domain. The smallest unit that can be scheduled for transmission is called the RB (Resource Block) and consists of a slot in time-domain and 12 subcarriers in frequency domain, corresponding to 12×15 kHz = 180 kHz in the frequency domain. The number of subcarriers within a PRB (Physical Resource Block) is 12 or 24 for the case of 15 or 7.5 kHz subcarrier spacing respectively (this spacing is used for MBSFN transmissions with extended CP). This is the most important unit in LTE because radio resources are allocated in units of RBs. If the normal CP is used, a RB will contain 12 subcarriers over seven symbols. If the extended CP is used, the RB contains only 12 subcarriers over six symbols. The minimum number of RB within a slot is six for the 1.08 MHz transmission bandwidth, while the maximum number of RB within a slot is 110 for the 19.8 MHz transmission bandwidth.

The UE is specified the allocation for the first slot of a subframe, the second slot of the subframe being implicitly allocated. If the eNB specifies one RB as the resource allocation for the UE, the UE actually uses two RBs, one RB in each of the two slots of a subframe. Upon activation of frequency hopping, the actual RBs that carry the UE data can be different in the two slots.

What is to be noted is that not all of the sub-carriers are modulated: the DC sub-carrier is not used as well as the sub-carriers on either side of the channel band.

4.4.2 Frame structure type 2

Frame structure type 2 is applicable to TDD. Each radio frame of length 10 ms consists of two half-frames of 5 ms. Each half-frame consists of five subframes of 1 ms each. Subframes consist of either an uplink or downlink transmission or a special subframe containing the DwPTS (Downlink Pilot Timelsot) and UpPTS (Uplink Pilot Timeslot), separated by a transmission gap GP (Guard Period). The special subframe lasts for 1 ms and is defined to allow switching from downlink to uplink and conversely. Each subframe, with the exception of the special subframe, consists of two slots of 0.5 ms each.

Uplink-downlink configurations with both 5 ms and 10 ms downlink-to-uplink switch-point periodicity are supported. In the case of 5 ms downlink-to-uplink switch-point periodicity, the special subframe exists in both half-frames. A smaller switch-point periodicity of 5 ms yields in lower latency relative to 10 ms periodicity at the expense of low efficiency due to an additional special subframe, the main reason for this being that in the case of 10 ms downlink-to-uplink switch-point periodicity, the special subframe exists in the first half-frame only.

The allocation of the subframes for the uplink, downlink, and special subframes is determined by one of seven different configurations. Subframes 0 and 5 and DwPTS are always reserved downlink transmissions, subframe 1 is always a special subframe. Subframe 2, UpPTS and the subframe immediately following the special subframe are always reserved for uplink transmission. The composition of the other subframes varies depending on the frame configuration.

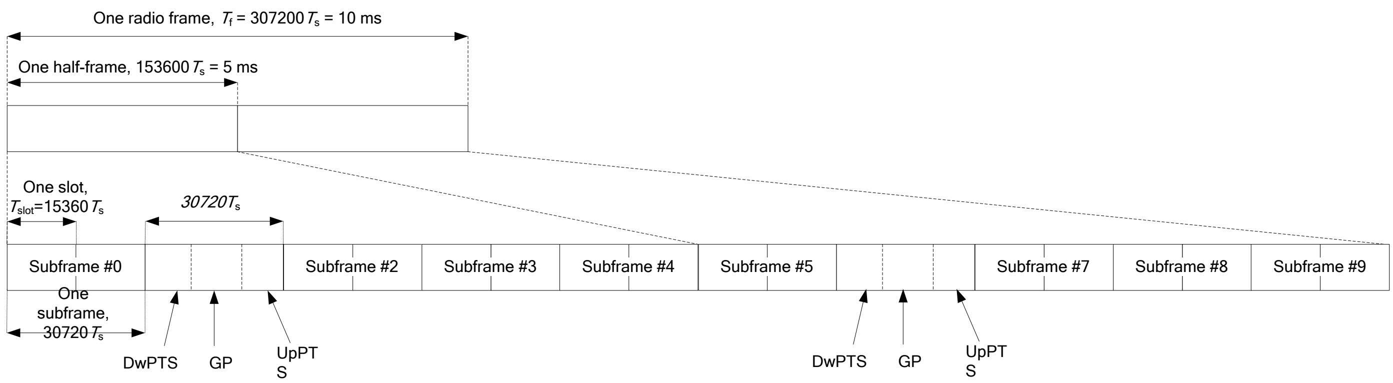

Figure 4.4 Frame structure type 2 (for 5 ms switch-point periodicity) [38]

The DwPTS field carries synchronization and user data as well as the downlink control channel for transmitting scheduling and control information. The UpPTS field is used for the transmission of the PRACH and the SRS (Sounding Reference Signal) [39].

4.5 Reference signals

4.5.1 Downlink reference signals

In both the downlink and the uplink there are RS (Reference Signals), which are employed by the receiver to estimate the amplitude and phase flatness (combination of errors in the transmitted signal and additional imperfections that are due to the radio channel) of the received signal. Demodulation would be unreliable due to phase and amplitude shifts in the received signal, especially for high-order modulation schemes such as 16 QAM or 64 QAM, without the use of the RS. In the case of this high order modulation schemes, even a small error in the received signal amplitude or phase can cause demodulation errors.

Thus, in order to allow for coherent demodulation at the UE, reference symbols (also known as pilot symbols) are inserted in the OFDM time-frequency grid to allow for channel estimation. Downlink reference symbols are inserted within the first and third last OFDM symbol of each slot (corresponding to the fifth and fourth OFDM symbols of the slot in case of normal CP and extended CP respectively) with a frequency domain spacing of six sub-carriers, as shown below. Furthermore, there is a frequency domain offset of three sub-carriers between the first and second reference symbols. As a consequence, there are four reference symbols within each RB. The UE will make an interpolation based on multiple reference symbols in order to estimate the channel. In the case of two Tx antennas, RS are inserted from each antenna in such a way that the reference signals from the second antenna are offset in the frequency domain by three subcarriers. In order to allow the UE to accurately estimate the channel coefficients, there is no information transmitted on the other antenna at the same time-frequency location of reference signals.

The reference symbols have complex values which are determined according to the position of the symbol and are cell-dependent. LTE specifications refer to this as a two-dimensional reference-signal sequence which indicates the LTE cell identity. A number of 510 reference signal sequences exist, which correspond to 510 different cell identities. The reference signals are obtained from the product between a two-dimensional pseudo-random sequence and a two-dimensional orthogonal sequence. There are 170 different pseudo-random sequences that correspond to 170 cell-identity groups, and three orthogonal sequences, each of them corresponding to a specific cell identity within the cell identity group.

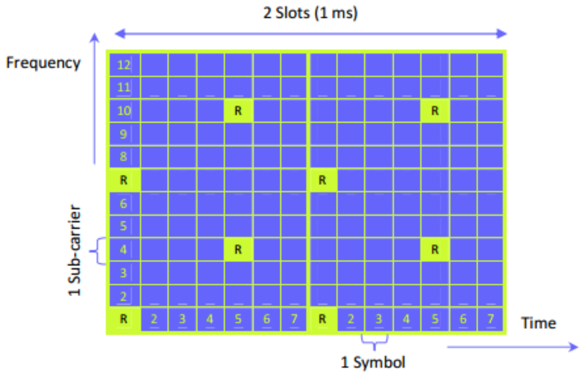

Figure 4.5 Location of reference symbols within a RB (one antenna system, normal CP) [40]

Two types of downlink RS are defined: cell-specific RS and UE-specific RS.

The cell-specific reference signals are used for downlink measurements, like channel quality estimation, MIMO rank calculation, MIMO precoding matrix selection and measurements for handover, as well as for demodulation of non-MBSFN transmissions, including downlink control channels and PDSCH transmissions not using UE-specific reference signals. For these RS, up to a maximum of four antenna ports are supported.

The UE-specific reference signals are provisioned to support rank-1 beam-forming on the downlink. For these RS, a single antenna port is supported.

4.5.2 Uplink reference signals

Two types of reference signals exist for LTE uplink. The first of them is DM-RS (Demodulation Reference Signals) which is employed for channel estimation for PUSCH or PUCCH demodulation. These signals are time multiplexed with the uplink data and are transmitted on the third or fourth SC-FDMA symbol of an uplink slot for extended or normal CP, respectively, using the same bandwidth as the data.

The second reference signal used in the uplink is SRS (Sounding Reference Signal) which is used to measure the uplink channel quality for channel-sensitive scheduling. The DM-RS cannot be used for frequency selective channels since they are assigned over the allocated bandwidth to a UE. SRS is a wideband reference signal usually transmitted in the last SC-FDMA symbol of a 1 ms subframe. User data transmission is not allowed in this block, resulting in about 7% reduction in uplink capacity. The SRS is an optional feature and it has a high degree of configurability in order to control overhead. Thus, it can be turned off in a cell [40].

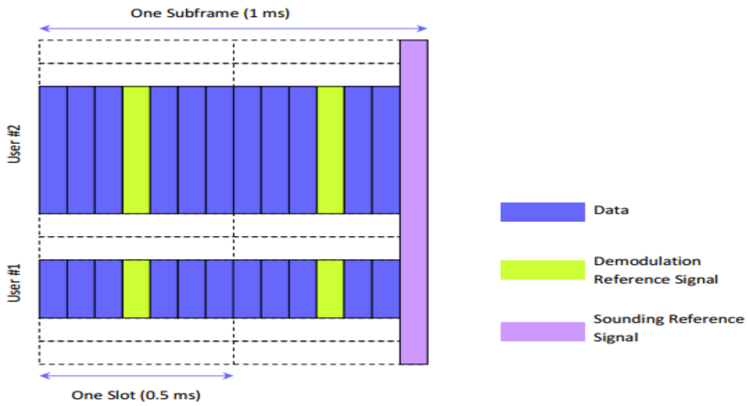

Figure 4.6 Uplink DM-RS and SRS (with normal CP) [40]

5. Radio resource management

RRM (Radio Resource Management) is an eNB application level function whose primary goal is the make an efficient use of the available radio resources. RRM manages the assignment, re-assignment and release of radio resources considering various aspects. It covers the control, assignment, sharing and distribution of radio resources among the users of the RAN (Radio Access Network) by dynamically allocating the available resources. The purpose of RRM is to ensure that the radio resources are used as efficiently as possible, taking advantage of the available adaptation techniques, and to serve the users according to their configured QoS (Quality of Service) requirements. The objective of RRM is to satisfy the service requirements with the minimum cost to the system, ensuring optimized use of spectrum.

In the case of LTE, RRM includes a variety of algorithms that provide services such as power control, resource allocation, mobility control, and QoS management to ensure that the radio resources are used in the best manner.

RRM may be regarded as a central application at the eNB level responsible for inter-working between different protocols, namely RRC, S1AP, and X2AP, so that messages can be transferred in a proper manner to different nodes at the Uu, S1, and X2 interfaces.

5.1 Functions of Radio Resource Management

RAC (Radio Admission Control)

The main purpose of RAC functional module is to admit or reject the establishment requests for new radio bearers. In order to achieve these desiderates, RAC considers the overall resource availability in E-UTRAN, the QoS requirements, the levels of priority, the QoS provided to in-progress sessions and the QoS requirement of the new radio bearer request.

The major objective of RAC is to ensure high radio resource deployment by accepting radio bearer requests when radio resources are available, while simultaneously ensuring the correct QoS for in-progress sessions, by rejecting radio bearer requests when they cannot be deployed.

RAC interacts with RBC (Radio Bearer Control) module to accomplish its functions.

RAC is implemented within the eNB.

RBC (Radio Bearer Control)

The RBC functional module involves the establishment, maintenance and release of radio bearers.

When RBC establishes a radio bearer for a service, it considers the overall resource availability in the network, the QoS requirements of the sessions that are already set-up and the QoS required for the new service.

RBC also deals with the maintenance of radio bearers of in-progress sessions when a change of the radio resource context arises (mobility, for example).

The release of radio resources associated with radio bearers at session termination, handover and other contexts also involves the RBC.

RBC is implemented within the eNB.

CMC (Connection Mobility Control)

The CMC functional module deals with the management of radio resources related with idle or connected mode. In idle mode, this module defines criteria and algorithms for cell selection, reselection (implemented through setting various parameters, like thresholds and hysteresis values that define the best cell and control when the UE should select a new cell) and location registration that the UE uses when selecting or camping on the best cell. Moreover, the eNB broadcasts parameters that are used to configure UE measurement and reporting procedures.

In connected mode, the mobility of radio connections has to be supported by the module without disruption of services.

Handover decisions may be based on UE and eNB measurements, but other inputs may be taken into consideration, such as neighbor cell load, traffic distribution, transport and hardware resources and operator specific policies.

Inter-RAT RRM is one of the sub-modules of this module and is responsible for managing the resources in inter-RAT mobility, namely handovers.

CMC is implemented within the eNB.

DRA (Dynamic Resource Allocation)

DRA, also known as PS (Packet Scheduling) allocates and de-allocates resources, such as buffer, processing resources and RB to user and control plane packets.

DRA encompasses also several sub-tasks: the selection of radio bearers whose packets are to be scheduled and management of the required resources (e.g. the power levels or the specific RBs used). DRA takes into consideration the QoS requirements associated with the radio bearers, the CQI for UEs, buffer status, interference level, etc. DRA may also take into account restrictions or preferences on some of the available RBs or RB sets due to inter-cell interference coordination considerations.

DRA is implemented within the eNB.

ICIC (Inter Cell Interference Coordination)

ICIC manages the radio resources in such a way that inter-cell interference is kept under control. The ICIC mechanism includes a frequency domain component and a time domain component.

ICIC is intrinsically a multi-cell RRM function that takes into account information (the resource usage status and traffic load) from multiple cells.

The preferred ICIC method may differ from uplink and downlink.

The frequency domain ICIC manages radio resources, notably the radio RBs.

The time domain ICIC uses ABSs (Almost Blank Subframes) to protect resources affected by strong inter-cell interference.

ICIC is implemented within the eNB [41].

LB (Load Balancing)

LB handles the uneven distribution of the traffic load over multiple cells. The major tasks of LB are:

- To influence the load distribution such that radio resources remain highly utilized
- To keep the QoS of in-progress sessions to the maximum possible extent
- To maintain the call-drop probability to a sufficiently small level

LB algorithms may result in handover or cell reselection procedures with the purpose to redistribute the traffic from highly-loaded cells to underutilized cells.

IRRRM (Inter-Rat Radio Resource Management)

IRRRM deals with the management of radio resources concerned with inter-RAT mobility, especially inter-RAT handover. In this particular case, the handover decision will also take into account the involved RATs resource situation as well as UE capabilities and policies of the operator.

IRRM may also include options for inter-RAT load balancing for idle and connected mode UEs.

SPID (Subscriber Profile ID)

RRM assigns SPID parameters received via the S1 interface to a configuration that is locally-defined in order to apply specific RRM strategies (e.g., to define RRC_IDLE mode priorities and control inter-RAT/inter-frequency handover in RRC_CONNECTED mode).

SPID is an index that refers to user information (such as mobility profile, service usage profile). The SPID information is UE-specific and applies to all of its radio bearers [42].

5.2 Scheduling and link adaptation in LTE

Reception conditions in mobile wireless networks vary greatly in frequency and time. In order to deal with these situations and guarantee the best possible QoS according to the radio channel conditions, LTE uses AMC (Adaptive Modulation and Coding) as the link adaptation technique that adjusts dynamically the transmission parameters (employed modulation scheme and code rate) according to instantaneous channel conditions in order to achieve a defined BLER (Block Error Rate) of below 10%, in order to keep retransmissions within an acceptable range. This is achieved by adapting the modulation scheme under use (QPSK, 16 QAM and 64 QAM) and the FEC (Forward Error Correction) coding rate.

The channel quality in the downlink is measured by the UE using the reference symbols, yielding after the measurement is complete CQIs (Channel Quality Indicators) that are afterwards sent to the eNB. A mapping between each CQI value and specific modulation schemes and code rates is performed, which are then used by the eNB for the downlink transmission. Thus, link adaptation in this context refers to rate adaptation or selection of the most suitable MCS depending on CQI values. The CQI is a measure of the dominant channel conditions and it is used in the scheduler and link adaptation techniques. It is a quantized value (usually with 30 levels of quantization) of the measured SINR (Signal to Interference and Noise Ratio) at the receiver, associated with an integer index within a MCS table. In general, link adaptation can also involve transmission power control.

Scheduling represents the process of dynamically allocating the physical resources among the UEs based on the scheduling algorithm.

Both scheduling and link adaptation require the CQI as input. Link adaptation is logically placed at the output of the scheduler in order to determine which users are scheduled and what RBs are allocated to them, and the output of both scheduler and link adaptation, which consists of the UE IDs of the scheduled users, the allocated resources and the MCS to be used for transmission, are sent to the UEs via PDCCH [43].

Various algorithms may be used by the scheduler in order to agree upon which users are to be scheduled and which resources will be allocated to them. Different aspects may be taken into account, such as spectral efficiency and fairness. Some of the most used algorithms are described below.

RR (Round Robin) scheduler

It represents the simplest technique of scheduling, in which users who have data to be transmitted are allowed to take turns, disregarding the CQI. The fairness of this scheduler arises from the fact that time and frequency resources are distributed evenly among the users. Because the users are scheduled without taking into account the instantaneous channel quality, this scheduler provides low spectral efficiency.

FDM (Frequency Division Multiplexing) scheduler

It represents a special case of the RR scheduler, where the users that have data to be transmitted are allocated an equal share in terms of frequency resources. In terms of time and frequency resources that are allocated to the users the fairness of this scheduler is the same as the one of the RR scheduler, but offers lower system performance than the latter. When CQI is also considered when distributing frequency resources, a channel quality dependent FDM scheduler is deployed.

MAX C/I (Maximum Rate) scheduler

This type of scheduler always schedules one user for a TTI by selecting the user who has the best instantaneous channel quality, thus the highest possible data rate. This scheduler is spectrally efficient, maximizing thus system performance, but it is fairness that it lacks.

PFTF (Proportional Fair in Time and Frequency) scheduler

This scheduler selects a certain number of users using as a decision criterion the ratio between the instantaneous channel quality and the average channel quality during the last averaging window period, which can be defined within the scheduler. Thus the users who have the best channel quality with respect to their average channel quality get scheduled. In terms of fairness and spectral efficiency, it lies between MAX C/I and RR schedulers.

After the users have been scheduled and the RBs have been allocated among them, the link adaptation determines the MCS that will be used for transmitting the data of each user. Usually, the MCS that offers the highest data rate while not exceeding a certain BLER set as a target is selected.

The data rate is determined by the chosen MCS, upon which the error rate also depends, together with the prevailing channel quality. The most robust transmission is achieved with QPSK, which maps two bits to each modulation symbol, resulting thus four stages. The large distance among the modulation points encountered in QPSK allows the correct decision at the receiver to be made with higher probability, even in a very noisy channel. Higher modulation schemes, like 16 QAM or 64 QAM allocate 4 or 6 bits respectively per modulation symbol, allowing thus a higher data rate and bandwidth efficiency, at the cost of a better SINR requirement at the receiver for error-free demodulation.

UEs send the received channel quality to the eNB by transmitting a CQI value, which can refer to the wideband receive quality or to the receive quality of each sub-band relative to the wideband average in the case of sub-band CQI reports. From this measurement information, a MCS scheme is derived, which is done by choosing 1 MCS out of 16 to achieve the target BLER of below 10%.

Due to inaccuracies determined by the quantization process, the propagation with a certain delay of the information and the averaging of the SINR to reduce the transmission overhead, a CQI adjustment at the eNB is performed. This is done by adjusting the CQI values by a certain margin, called LAM (Link Adaptation Margin), as shown below:

$$\left[CQI_{eff}\right]=\left[CQI\right]-\left[LAM\right]$$

The size of the matrices above depends on the number of users in the cell, number of RB and number of transmission streams in the case of MIMO. CQI_{eff} is the effective value that will be used by the scheduler and link adaptation procedures. Thus, the LAM can increase or decrease the original CQI. A lower CQI means a lower data rate, in other words, a more robust MCS.

There are two types of link adaptation:

- **Fixed link adaptation**: adapts the CQI for all users, for all RBs and for all the transmission streams by the same fixed value (the LAM in the equation above can be regarded as a matrix with equal and constant elements)
- **Differential link adaptation**: the LAMs change according to some algorithm. In this case, the LAM in the equation above can be seen as a matrix whose elements might change independently of each other). The algorithms that perform the update of LAM may act as a control loop that considers the current set of LAMs and feedback (usually the ACK and NACK feedback for the recent transmissions is used) from the system and outputs the new set LAMs [44].

The data that is modulated with different MCSs to subcarriers needs to be protected against inherent transmission errors. To deal with this problem, LTE defines a turbo de-/encoder with trellis termination of a native code rate of one-third (the code rate is the ratio between the source data rate and the resulting protected data rate; a code rate of one-third encodes 1 bit into 3 bits). This represents basically channel coding, through which the turbo coder adds redundancy bits to the data, some bit errors being thus possible to correct.

Additionally, one parameter being controlled is the UL transmit power. UL power control is implemented to deal with the near-far effect, which occurs when a user is close to the base station (near) and another user is further away from the base station (far), having thus a higher path loss leading to a lower receive power of the signal from the more distant user. Moreover, the two signals sent by the two transmitters represent noise for one another. Thus, the greater the power of the transmission of the first user, the smaller the SNR of the more distant one. The solution is the dynamic power adjustment of the transmitters: the closer transmitters use less power so that the SNR at the receiver from all other transmitters is the same. This can have a major impact on the battery life of the UE depending on the distance from the base station [45].

5.3 Hybrid Automatic Repeated Request

HARQ is a technique that makes use of both FEC to correct a number of errors and conventional ARQ to detect any further errors and requests for data retransmission. After FEC is used to correct a subset of errors, the receiver employs an error detecting code, most often a CRC one to detect if the packet still has some errors. If some errors still exist, the data is discarded and a NACK (Negative Acknowledgement) is sent to the transmitter to notify it to retransmit the data. In the case that all errors have been corrected, an ACK (Acknowledgement) is sent to confirm that the data was received error-free. The process is repeated until the maximum number of allowed retransmission attempts is reached.

LTE uses HARQ with soft combining to deal with the retransmissions. In this process, the erroneous packets are buffered at the receiver because some information is still present in the received signal, even if the decoding process could not be performed correctly. The buffered packets are afterwards combined with their retransmissions and passed to the decoder for FEC, followed by error detection.

The retransmissions may not necessarily contain exactly the same coded bits as long as they contain the same information bits. HARQ with soft combining is categorized as chase combining or incremental redundancy depending on whether the retransmissions contain the exact coding bits or not [43].

5.4 LTE QoS Concepts

QoS refers to the ability to provide a certain service like a voice call, a mobile game or a video call at a required quality level. Certain services require higher priority than other guaranteed resources (for example, the voice calls should be prioritized over file transfer). It is expected, of course, that the quality level will align with the customer's expectation, for example: performing a voice call without distortions, breaks or other disturbances or playing a mobile game and getting an immediate response to any move. In addition to that, there are premium subscribers who want to have a better user experience and are eager to pay more for high bandwidth and better network access on their devices.

The LTE network, as many other communication systems, delivers a variety of data types belonging to different services using limited air interfaces and routing resources of the network. Since LTE is an IP based technology, there are no switched circuits assigned for active sessions and therefore, alternative mechanisms need to be used in order to guarantee the required resources and QoS. These are the main factors that fueled the need for a set of conventions and rules for:

- Classification of services and related minimal QoS requirements
- Prioritization of services
- QoS enforcement rules and methods

QoS is handled by several principles:

- **Differentiation**: differentiate between services (voice, public Internet, peer-to-peer file sharing, video-streaming, mobile-TV) and possibly, differentiate between customers (business vs standard, post- vs pre-paid, roamers and privileged – e.g. police). For example, differentiate between the voice call of a premium customer and a file transfer initiated by an ordinary customer.
- **Prioritization**: based on the differentiation made earlier, provide to each service customer combination the required priority, while allocating air interfaces and routing resources at each node of the network.
- **Admission** control: since the system has limited resources, there might be situations where in order to guarantee a minimum QoS to high priority service subscribers, some lower priority service subscribers will have to be blocked, even before the session has been started or in some extreme situations even dropping the lower priority session.

In order to allow service customer differentiation, the LTE network includes a mechanism that attaches a QoS tag to each packet. These tags are determined based on the type of service each packet belongs to and also based on the type of customer that uses this service. Each packet is characterized by two QoS tags, as it can be seen in the figure below.

ARP=j	QCI #k	To UE # n	Service ID # m	Data

Figure 5.1 QoS tags

The first QoS tag is ARP (Allocation and Retention Priority), which defines the relative importance of a resource request. This is used as a basis for admission control policy in order to decide whether a new session could be admitted or if a lower priority session should be dropped.

The second tag is QCI (Quality Class Indicator) determines the treatment that IP packets of user data (and IMS signaling) should receive while in session. It defines minimum QoS parameters, like delay and packet error rate, that the packet of a certain service should retain.

As it can be seen, QCI is actually a figure that defines a set of QoS parameters related to a specific type of service. The QCI levels are divided in two major categories. The first one is GBR (Guaranteed Bit Rate) QCIs, which dictate that the packets of this category will always get the guaranteed bit rate, even at the expense of other non-guaranteed bit rate packets. The other category is Non-GBR in which packets of this category might experience congestion related packet-loss.

The other parameters defined by QCI, which are also shown in the table below, are:

- **Priority**: provides the scheduling and routing priority for each packet at each node
- **Delay**: marks the upper bound, with 98% confidence, for the delay a packet can experience between the UE and the PDN-GW
- **PELR** (Packet Error Loss Rate), which provides the upper bound for the percentage of packets that might be lost. The non-GBR services can experience additional packet loss due to congestion

<table>
<tr><th>QCI</th><th>Resource type</th><th>Priority</th><th>Delay (ms)</th><th>PELR</th><th>Example services</th></tr>
<tr><td>1</td><td rowspan="4">GBR</td><td>2</td><td>100</td><td>0.01</td><td>conversational voice</td></tr>
<tr><td>2</td><td>4</td><td>150</td><td>0.001</td><td>conversational video (live streaming)</td></tr>
<tr><td>3</td><td>3</td><td>50</td><td>0.001</td><td>real time gaming</td></tr>
<tr><td>4</td><td>5</td><td>300</td><td>0.000001</td><td>non-conversational video (buffered streaming)</td></tr>
<tr><td>5</td><td rowspan="5">Non-GBR</td><td>1</td><td>100</td><td>0.000001</td><td>IMS signaling</td></tr>
<tr><td>6</td><td>6</td><td>300</td><td>0.000001</td><td>Video (buffered streaming); TCP-based (www, e-mail, chat, ftp, progressive video, etc.)</td></tr>
<tr><td>7</td><td>7</td><td>100</td><td>0.001</td><td>voice, video (live streaming), interactive gaming</td></tr>
<tr><td>8</td><td>8</td><td rowspan="2">300</td><td rowspan="2">0.000001</td><td rowspan="2">Video (buffered streaming); TCP-based (www, e-mail, chat, ftp, progressive video, etc.)</td></tr>
<tr><td>9</td><td>9</td></tr>
</table>

Table 5.1 Classes of QCI

QoS is implemented between the UE and PDN-GW. In order to manage the delivery of the packets in a differentiated manner, the packets are grouped into bearers. The bearers identify packet flows that receive a common QoS treatment. A bearer may include packets of multiple services which require the same QoS treatment (PELR, delay, priority) and is defined by a combination of QCI and the destination IP address. There are two types of bearers:

- **Default bearer**, which is set up when the terminal attaches to the network. It provides the basic connectivity and is kept as long as the terminal retains the IP address. The default bearer is a non-GBR bearer.
- **Dedicated bearer**, which is set up for QoS sensitive applications. The PCRF determines which packet flows are mapped on the dedicated bearers [46].

Usually, LTE networks that have VoLTE implemented have two default bearers and one dedicated bearer. The first default bearer is used for signaling messages related to IMS network, employing QCI 5. The dedicated bearer is used for VoLTE VoIP traffic, being linked to the first default bearer and using a QCI 1. The second default bearer is used for all other smartphone traffic (video, chat, email, browser, etc.) considering that QCI 9 is used here.

In LTE, QoS is applied on the radio bearer (on the LTE-Uu link between the UE and eNB), S1 bearer (between the eNB and S-GW) and S5/S8 bearer (between S-GW and PDN-GW). All these three bearers are collectively known as EPS bearer, whereas the radio bearer and the S1 bearer are collectively known as e-RAB (e - Radio Access Bearer), as shown in the figure below.

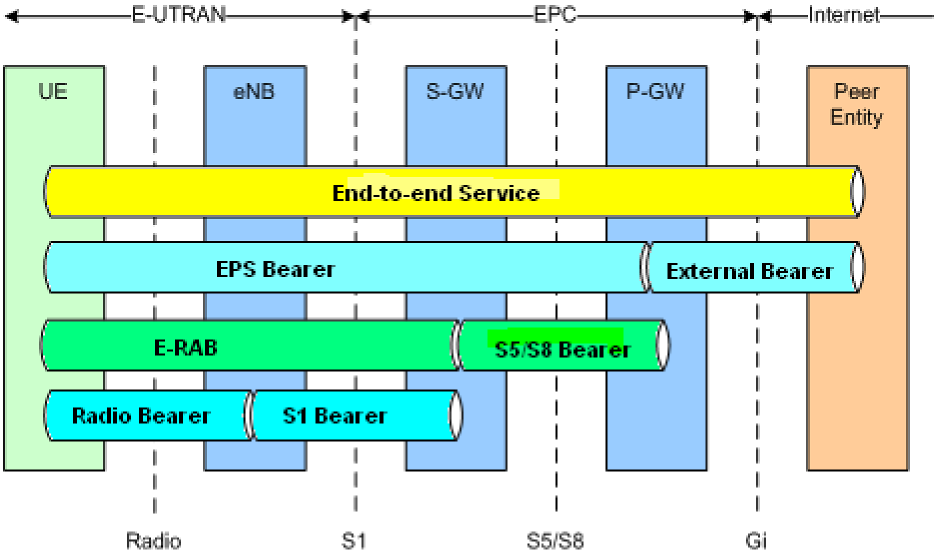

Figure 5.2 LTE bearer service architecture [47]

From the figure above, we can observe that an e-RAB transports the packets of an EPS bearer between the UE and EPC. The radio bearer transports the packets of an EPS bearer between an UE and an eNB, whereas the S1 bearer transports the packets of an e-RAB between the eNB and the S-GW. The S5/S8 bearer transports the packets of an EPS bearer between a S-GW and a PDN-GW.

EPS is a connection-oriented transmission network and, as a consequence, it requires the establishment of a virtual connection between two endpoints represented by the UE and PDN-GW, prior to any traffic being sent between them. This virtual connection is called an EPS bearer, emphasizing that the connection provides a transport service with specific QoS attributes.

An EPS bearer is characterized by:

- The two endpoints implicated in the communication
- A QCI that describes the type of service that employs the virtual connection
- A flow specification that describes the GBR and MBR (Maximum Bit Rate) of the aggregate traffic flow that passes through the virtual connection
- A filter specification that defines the traffic flows, mentioning the IP addresses, protocols, port numbers for which the transport service is provided between the two endpoints

The “EPS bearer activation” procedure is used to establish the EPS bearer, and can be triggered by both endpoints. As an example, this procedure might be triggered by the PDN-GW when it determines that a new VoIP call is requested and a new EPS bearer should be established in order to support this call [48].

6. Voice in LTE

LTE is an IP based system optimized for data delivery. However, voice services are still a major part of the operator's income. In order to provide voice services and assure the communication with telephony networks, the following are required to support voice service:

- The capability to provide and interpret standard signaling messages used in telephony systems, such as: request for a call, "ringing" notification, end of a call, etc.
- The ability to provide and receive digitized voice signals in standard formats used in telephony networks like, for example, PCM (Pulse Coded Modulation)
- Having the mechanism for maintaining the required QoS for voice packets while they are processed by the network

Until LTE, these functions were performed by the CS (Circuit Switched) core.

GSM networks started using FR (Full Rate) speech codec and advanced to EFR (Enhanced Full Rate). The AMR (Adaptive Multi Rate) codec appeared in Release 98 for GSM. AMR data rates are between 4.75 kbps and 12.2 kbps, the highest AMR rate being equal to the EFR. AMR uses a sampling frequency of 8 kHz, which means that the bandwidth of the audio signal is between 300 – 3400 Hz. The AMR codec is used for WCDMA, for the voice service over HSPA, and it can also be used for LTE. Most of the mobile voice traffic is carried with GSM EFR or AMR coding.

The AMR-WB (AMR Wideband) codec was added to 3GPP Release 5. It uses a sampling frequency of 16 kHz, implying an audio bandwidth of 50 – 7000 Hz and much better audio quality than AMR, even if its typical rate of 12.65 kbps is comparable to the AMR's 12.2 kbps. This is the 3GPP recommended speech codec for VoLTE (Voice over LTE) and is also known as HD Voice.

To the AMR rate of 12.2 kbps corresponds a packet size of 31 B, whereas the IP header is 4-60 B. Thus, IP header compression is a top priority for an efficient VoIP implementation, being required both in the UE and eNB.

The always-on IP connectivity requires keep alive messages when the UE does not have a phone call in progress, the frequency of these messages depending on the VoIP solution.

Voice capacity represents the maximum number of users that can be supported within a cell without exceeding an outage of 5%. A user is said to be in an outage if at least 2% of its VoIP packets are lost during the call, either due to errors or to them being discarded.

As a rule, lower AMR rates provide higher capacity than higher AMR rates. The AMR codec data rate is an operator-defined parameter that offers a tradeoff between capacity and voice quality. Thus, depending on the data rate of the AMR codec (AMR 5.9 kbps, AMR 7.95 kbps and AMR 12.2 kbps are the most common codecs used in current networks), the cell capacity ranges between 210-470 users in downlink and 210 – 410 users in uplink, corresponding to 42-94 users per MHz per cell in downlink and 42-82 users per MHz per cell in downlink, considering a 5 MHz LTE bandwidth. LTE VoIP 12.2 kbps can provide up to fifteen times the efficiency of GSM EFR. The high efficiency translates in more voice traffic squeezed within a narrower spectrum [49].

Most cellular systems, prior to LTE included CS core that besides offering a dedicated guaranteed circuit to each voice session, also communicated with external telephony networks using the standard signaling protocols and digitized voice formats.

The 3GPP HSPA version 6 for example included two cores: a PS (Packet Switched) core optimized for packet data delivery and a CS core which handles all tasks related to voice sessions, including communication with external telephony networks.

The LTE network however, includes only a PS core and does not have a CS core for handling voice sessions. As an alternative, 3GPP specified the use of IMS as the platform for handling voice sessions. According to the LTE standards, IMS is the entity that is supposed to handle standard telephony signalization, translate between VoIP and standard telephony formats and assist with the process of maintaining the required QoS (including mainly delay, error rate and bandwidth requirements) for the voice call. The full IMS voice service is called MMTel (Multi-media Telephony).

IMS in its full scale is a feature-rich complicated and expensive subsystem. Commercial systems are already available and some are already deployed, but mass adoption is still slow, mainly because of the complicated integration which requires architectural changes in existing and operating networks, the lack of experience of IMS supporting large-scale networks and the cost of the full-scaled system, which is very high.

So, if a full IMS service is not available yet for most carriers, a few alternatives are available for delivering voice services for the LTE subscribers. Three solutions are available today for allowing LTE carriers to offer voice services. Two solutions, namely CSFB (Circuit Switched Fallback) and VoLGA (Voice over LTE via Generic Access) are based on the use of the legacy 2G/3G CS subsystems. CSFB instructs handsets to ‘fallback’ to a 3G or 2G network whenever voice services are required. VoLGA intends to packetize the existing voice and SMS services and deliver them over LTE. The third alternative is a standard named VoLTE (Voice over LTE) which is based on a minimum set of IMS network entities and functionalities. VoLTE was created to allow LTE carriers to provide voice service without the need to invest in a complete IMS system and make radical changes in the network.

6.1 Circuit-Switched Fallback

The CSFB mechanism, which has been specified in 3GPP TS 23.272, uses the legacy 2G/3G core network and radio access for providing voice service to LTE subscribers equipped with dual-mode handsets. With CSFB approach, whenever a request for a voice call arrives, the LTE user is handed over to a legacy CS network, assuming it provides an overlapping coverage. At the end of the call, at the legacy network, the device may, but not necessarily re-associate and register again in the LTE network.

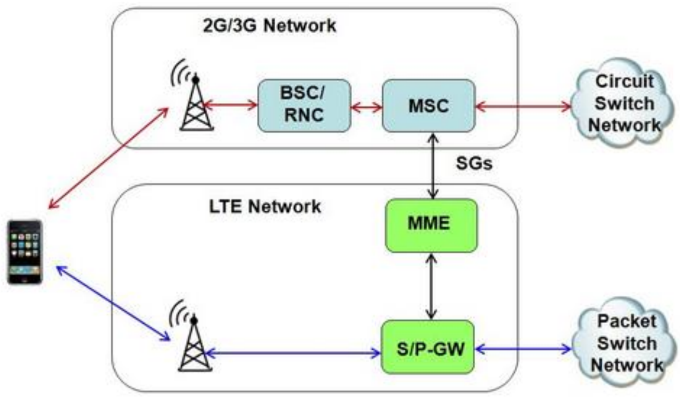

Figure 6.1 CSFB [50]

To support CSFB, a new interface – SGs, is introduced, connecting the MME with the MSC (Mobile Switching Centre). When the LTE handset with CSFB is turned on, it registers in both legacy and LTE networks, performing initially a CS and PS attach to the LTE network, also informing the network that it is CSFB capable for incoming and outgoing calls. To enable fast handover to the legacy network when required, the legacy network needs to know the location of the LTE user. Therefore, the MME which tracks the location of the user within the LTE network continuously provides the location update information to the legacy MSC using the new SGs interface. If successful, the MME signals the UE that it has been registered in the CS network and that incoming calls will be signaled to it.

Assuming the UE is initially served by the LTE network and has an active IP connectivity, when the UE decides to make a voice call (mobile originated calls), it sends a service request message to the MME, which checks whether the UE is capable of handling CSFB and notifies the eNB to transfer the UE to the legacy network. Before handing over the UE, the eNB may request the UE to perform measurements of the neighboring 2G/3G cells or it may use a list of preconfigured cells. When the destination cell has been identified, the eNB triggers a cell change to the 2G/3G network by sending a RRC message to the UE.

Any IP data transmission taking place at the time the voice call is initialized will either be handed over to the legacy network or it will be suspended until the UE returns to the LTE network. If the IP data transmission is also handed over, then it might operate in the lower speeds of the legacy network, which can also reject the handover of the IP session if it is not able to process it. If the fallback network is UMTS, the E-UTRAN will perform a PS handover such that the data session will be continued during the voice call. If the fallback network is GSM, a PS handover occurs only if DTM (Dual Transfer Mode) is supported by both the UE and the network, enabling thus voice and data calls to be handled simultaneously. Otherwise, the data session is suspended for the duration of the voice call, being thereafter resumed when the UE enters the LTE network.

When an incoming call arrives to the LTE user (mobile terminated call), the call request first arrives to the MSC which the UE has previously registered to. When the MSC gets the call request, it sends a paging message to the related MME via the SGs interface. This message is forwarded to the UE which is still connected to the LTE network. If the user accepts the call, it sends a service request to the MME, which then notifies the eNB to transfer the UE to the legacy network and the eNB decides the best network for the UE to perform the handover with.

Delivering and receiving SMS messages does not require CSFB. The MSC simply forwards the SMS received from the SMSC (SMS Service Center) to the MME by using the SGs interface, the MME being thereafter responsible to deliver it to the UE via MME to mobile device signaling messages that are transparent to the eNB. The same procedure is applied for mobile originated SMS messages, but in the reverse order [51].

6.2 Single Radio Voice Call Continuity

SRVCC (Single Radio Voice Call Continuity) is not an alternative for voice delivery but rather a voice-call handover process. SRVCC is a method for ensuring fast and reliable handover of an LTE user while it is in an active IMS-based voice session while it leaves the LTE coverage area to a legacy network coverage area, keeping thus call continuity.

As an SRVCC-capable mobile engaged in a voice call determines that it is moving away from LTE coverage, it notifies the LTE network. The LTE network determines that the voice call needs to be moved to the legacy circuit domain. It notifies the MSC server of the need to switch the voice call from the packet to the circuit domain and initiates a handover of the LTE voice bearer to the circuit network. The challenge with SRVCC is to make a handover while the UE is connected to a single radio at a given time.

To allow SRVCC continuity, the UE and both networks (LTE and the target legacy network) should support the service. In addition, a special interface called Sv has been introduced between the MME and the MSC.

For supporting SRVCC, the IMS network should also include an application server called SCC AS (Service Centralization and Continuity Application Server), which handles the signaling required for the process.

When the LTE signal strength starts diminishing in the usage area, the UE signals back to the eNB about the change in signal strength and the SRVCC handover is initiated. The eNB then calculates the best network available to handle the served UE and sends the request for handover to the MME. In the handover request, the eNB also specifies that the handover is also SRVCC based. A new voice call request is then sent to the IMS using a special number STN-SR (Special Transfer Number for SRVCC) which is generated for each UE and is stored in the HSS. This number is sent to the MME by the HSS when the UE first contacts the network. Receiving an STN-SR number indicates the SCC AS that the corresponding call needs to be routed to a different network and it starts the redirection process to the legacy endpoint. After the resource operation is complete, the MME confirms the handover request in the provided area by the eNB, which then transmits this confirmation to the UE and also provides the required information about the target network. In the final steps, the UE is detected in the legacy network and the call is re-established with the UE.

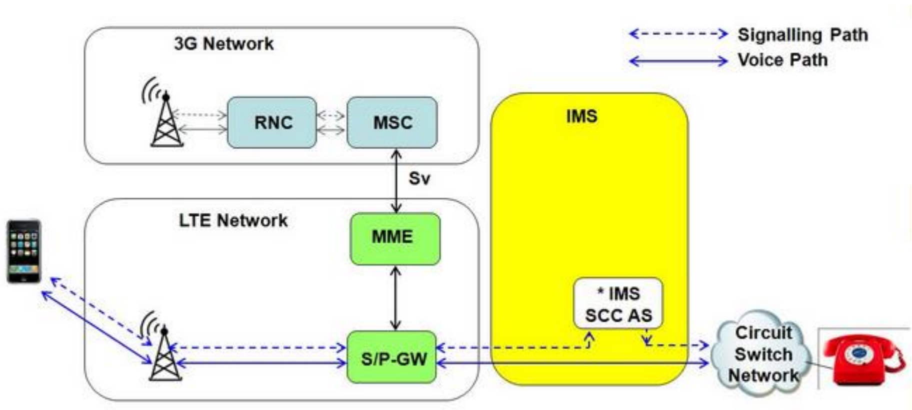

Figure 6.2 SRVCC [50]

6.3 Voice over LTE

The LTE mobile communication standard is optimized for data transfer and its designed is based on a flat, all-IP PS network. Thus, it does not include a CS path that is currently used for voice and SMS services. However, these last two services provided by mobile carriers represent a very important part of their revenues, determining the adoption of voice offered through the LTE network. With VoLTE (Voice over LTE), formerly known as One Voice, which is described in GSMA VoLTE IR.92 specification, based on a series of 3GPP-adopted standards, consumers will be able to experience “telco-grade” voice services that have a quality at least equal, if not greater, than legacy CS voice services. It is this concept that essentially differentiates VoLTE from other over-the-top voice services like Skype, because otherwise the two have basically the same working principle: digitized voice-band audio that is encapsulated in IP packets, transmitted and upon receipt, it is demodulated into voice-band audio.

The IP protocol is a ubiquitous one when it comes to digital data transmissions, including voice. However, generic IP involves best-effort delivery, which cannot guarantee latency, which is desirable to be as low as possible for voice traffic. Because there is no solid proof that the networks of the future will always have enough available capacity to serve users on a best-effort basis, VoLTE was adopted.

In order for voice to run over an LTE network, VoLTE is deployed on the costly IMS platform rather than an IP network (the Internet for example), offering thus a level of control that cannot be achieved otherwise. MMTel, which is deployed on the IMS core, is the technology that offers the telephony service in both LTE and fixed networks. VoLTE needs to be supported by both LTE RAN and EPC, this being possible through a software upgrade.

The primary motivation for choosing and developing VoLTE is the guaranteed QoS it can offer. QoS describes the delivery of a minimum level of latency, availability and data integrity, being a concept that can be applied on a per-user basis or on a per-application service.

In order to achieve certain levels of QoS, each link in the end-to-end chain needs to be controllable by the network operator, which is almost an impossible task when dealing with the best-effort delivery and the uncertainties that may arise in the case of an IP-based network. The deployment of VoLTE by using the IMS subsystem allows operators to increase the quality of their services by committing to QoS levels.

There are four RAN features that are required for supporting VoLTE, as shown below. These features will enable carriers to clearly differentiate between VoLTE and VoIP.

- **SPS (Semi-Persistent Scheduling)**

In LTE, every PRB on the DL and UL has to be granted in an explicit manner. The tradeoff is that the resulting overhead is not very efficient for traffic that requires a repeated allocation of small packets, as in the case of LTE.

When it comes to dynamic scheduling, in the DL resources are being allocated when data is available, whereas in the UL, the UE dynamically requests the grant for transmission each time data arrives in the UE's uplink buffer. While this type of scheduling is well-suited for burst-specific, infrequent and bandwidth-demanding data transmissions (like web surfing, video streaming) it is not the best option for real-time streaming applications like voice calls, where data is sent in short bursts at regular intervals (every VoIP packet is sent or received every 20 ms when speech is actually performed), with a very low data stream rate, generating thus considerable overhead for the scheduling messages because very little data is sent for each scheduling message.

SPS offers a solution to this problem: instead of scheduling each DL/UL transmission, a transmission pattern is defined instead of using an opportunity-based approach, reducing thus significantly the scheduling assignment overhead.

During silence periods in the speech, wireless voice codecs do not transmit any data, sending only SDIs (Silence Description Information). During these periods of silence, the persistent scheduling is stopped, this being the main reason why it is called semi-persistent scheduling.

- **TTI bundling**

LTE uses TTIs of 1 ms in order to address the end-to-end latency problem, waiting for the HARQ process to acknowledge transmissions every 1 ms. The problem is that at the periphery of a cell, a UE might not have sufficient time at its disposal to reliably deliver an entire voice packet in one TTI. The solution is to merge (or bundle) multiple TTIs together, without waiting for HARQ feedback for each TTI. In this manner, a VoIP packet is sent as a single packet data unit during a bundle of consecutive TTIs and the HARQ acknowledgement is expected only after the last TTI of the bundle has been transmitted.

- **DRx (Discontinuous Reception)**

DRx (also known as sleep mode) is a process of turning off the radio receiver of an UE when incoming messages are not expected to be received. For DRx to be operational, the UE must be carefully coordinated with the mobile network for the grouping of messages. Since VoLTE traffic is predictable (a VoIP packet at every 20 ms), the UE will wake up during scheduled periods to look for its messages, reducing thus the power consumption, extending in this way battery life.

- **RoHC (Robust Header Compression)**

In streaming applications, the overhead of IPv4 is 40 B, while for IPv6 it is 60 B. Using the AMR-WB codec at a rate of 14.4 kbps translates in a payload data for a VoIP packet of about 50 B/20 ms frame, which means that in certain cases the overhead can be greater than the payload.

Thus, RoHC is employed to reduce the 40-60 B of overhead into only 1-3 bytes, making thus better use of the scarce bandwidth of the wireless systems, improving efficiencies on the air interface [52].

7. Key performance indicators and measurements for LTE Radio Network Optimization

In this chapter, an outline of important performance parameters that are needed to troubleshoot and optimize the network is presented, along with a series of KPIs (Key Performance Indicators) that are used to verify the performance of the most important network functions.

7.1 Radio quality measurements

3GPP defined in TS 36.214 a series of radio quality measurements, which can be divided in two categories: E-UTRAN measurements that are provided by the eNB and UE measurements that are provided by the mobile handset.

These radio quality measurements are performed both in UL and DL. The only uplink measurement performed by the UE and sent through a RRC measurement report is the UE TX power, which represents the power used by the mobile handset to physically send the UL signal towards the eNB.

The eNB measures the following parameters: RIP (Received Interference Power), thermal noise power, TA (Timing Advance). Moreover, an air interface tester can additionally provide the following information: channel baseband power (which measures the change in power amplitude of a particular physical channel), I/Q constellation diagrams (are used to check the quality of the modulated symbols of the received radio signals by comparing the pattern of the received symbols with the ideal constellation points), using a metric called error vector magnitude which indicates how far the points are from the ideal locations).

After the measurements are performed they will be sent to the Operation and Maintenance Center via a proprietary protocol.

7.1.1 Measurements performed by the UE

RSRP (Reference Signal Received Power)

This indicator is used to measure LTE cell coverage in the DL. RSRP is used in cell selection/reselection procedures in RRC_idle mode. The UE will send RRC measurement reports that report RSRP values that fall within the range -140 to -40 dBm, with a resolution of 1 dBm.

The primary purpose of RSRP is to determine the best cell on the DL radio interface in order to select it as the serving cell for initial random access or for an intra-LTE handover.

RSRP measurements can be performed in both RRC_idle and RRC_connected modes

There is also a correlation between RSRP and the user plane QoS. For an outdoor cell, the RSRP measurement results can fall within three ranges:

- RSRP > -75 dBm. In this case, excellent QoS is to be expected if the cell has not too many subscribers demanding bandwidth
- -95 dBm < RSRP < -75 dBm. In this case, a slight degradation of the QoS can be expected
- RSRP < -95 dBm. In this case, the QoS becomes unacceptable and throughput declines to zero between -108 dBm to -100 dBm. In this radio conditions, call drops are to be expected.

However, indoor cells, due to their limited coverage, can deal with radio conditions that are worse than those of the outdoor environment. Consequently, acceptable QoS can be encountered for RSRP values of around -115 dBm.

RSRQ (Reference Signal Received Quality)

In cellular networks, when a UE moves from cell to cell, performing cell selection or reselection and handover, it measures the signal strength and quality of the neighboring cells through two parameters: RSRP and RSRQ.

The RSRQ measurement represents the SNR of the reference radio signals and it can only be performed in RRC_connected mode. When the RSRP is not sufficient to make a reliable handover or cell selection/reselection procedure decision, the RSRQ measurement provides additional information. In the LTE handover procedure RSRP, RSRQ or both can be used.

$$\mathrm{RSRQ[dB]} = 10\lg\frac{\mathrm{RSRP}}{\mathrm{RSSI}}$$

RSSI (Received Signal Strength Indicator) is a parameter which provides information about the total received wideband power, measured for all signals, including all interferers and thermal noise.

The RSRQ takes value between -19.5 and -3 dB, with a resolution of 0.5 dB.

When comparing RSRP and RSRQ measurement results performed at the same geographic location, it is possible to determine if coverage or interference problems are present at that specific location. Thus, if the RSRQ is low while RSRP is stable, this indicates increased DL interference (possible causes: external interferer, interfering neighbor cell). If both RSRP and RSRQ have small values, this means an increased path loss for the signal, indicating thus an area with weak coverage due to obstacles, for example.

Three quality ranges can be defined for RSRQ. If RSRQ > -9 dB, the best subscriber experience is offered. If RSRQ lies between -9 and -12 dB, a slight degradation of QoS might be observed. If RSRQ < -13 dB, the user experience worsens with a high risk of call-drop arising and low throughput.

PH (Power Headroom)

Is expressed in decibels and is defined as the difference between the UE maximum transmit power and the estimated power for the PUSCH channel transmission. Basically, it is the power reserve that can be added to the UL transmission if the UE moves toward the edge of the cell or it requires a service with a higher bit rate.

PH reports are sent on the MAC layer and not in RRC measurement reports. The eNB uses this reported value to estimate how much uplink bandwidth a UE can use for a specific subframe. The more RB the UE employs, the higher the UE Tx power, which should not exceed the maximum power defined in the specification. Therefore, a UE cannot use additional RBs, which translates into bandwidth, if it does not have enough PH.

The reporting range of PH is between -23 to +40 dB.

7.1.2 Measurements performed by the eNB

RIP (Received Interface Power)

RIP represents the uplink noise floor for a set of UL RBs. Usually, the noise in the UL is generated by the UL signals received from all UEs in a specific frequency range of these RBs on a single Rx antenna.

The reporting range for RIP is between -126 to -75 dBm.

The number of active subscribers determines the load in a cell. A possible cause for exceptional values of RIP is represented by high-frequency signal sources that emit outside the LTE RAN.

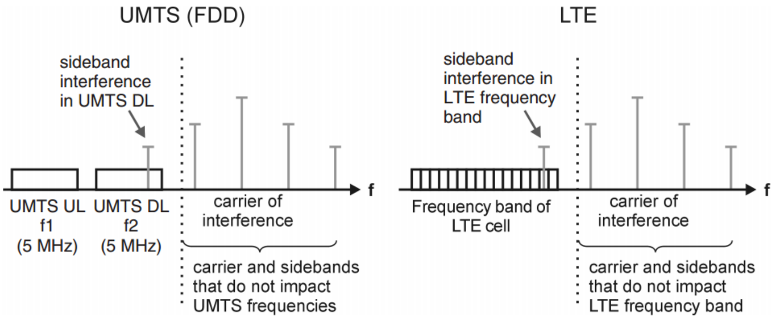

Figure 7.1 The impact of external interference in UMTS and LTE cells [53]

As it can be seen in the figure above, the effects of interference caused by external sources have a smaller impact on an LTE cell than on an UMTS one. In UMTS, if a single sideband of an interfering high-frequency source superimposes on the UL or DL frequency band of a cell with a certain power, the entire bandwidth of the cell is affected. In the worst case scenario, the cell is rendered unusable for transmission of UMTS radio signals, because all the connections are distributed over the entire bandwidth. In LTE, due to the diversity in both frequency and time domain offered by the scheduling grid, the impact of the interfering sideband is reduced to a minimum. Thus, the interfering carrier will affect only a few subframes of the entire frequency band, minimizing the impact on the entire cell. We may conclude that LTE is more robust to interference than UMTS FDD.

TA (Timing Advance)

TA is required to synchronize the transmission and reception of UL radio signals between the UE and eNB in the time domain of the air interface. It represents the estimated time for an UL subframe to travel from the UE's Tx antenna to the eNB's Rx antenna. TA mitigates the problem that would be generated by an unaccounted delay, that would determine the timeslots of the signal from a certain UE and that of the next signal from another UE received at the eNB to overlap, causing ISI.

TA takes integer value between 0 and 1282. The step size of TA is 16 T_S, where T_S is the sampling period of the LTE signal: $T_S = 1/(15000*2048) = 32.552083$ ns. Thus, one step TA in the time domain equals TA = $16T_S = 0.52$ µs.

Considering that the radio waves travel at the speed of light c=300000 km/s, the distance for a single TA step is: $r = c \times T_S = 156$ m.

The maximum distance from a UE to the eNB is obtained by multiplying the maximum value of TA, which is 1282, with the distance for a single TA step, which is roughly 156 m, yielding a distance of about 200 km.

Usually, 6 bits are allocated for TA, which translates into 2^6=64 index values, covering thus a distance of 64 × 156 m = 9984 m, with a TA granularity of 2 Hz (a TA command is sent every 500 ms)

Thermal noise power

Thermal noise power represents the UL noise for the entire bandwidth of the receiving cell, without any signals received from the UE, being basically UL without LTE traffic [54].

7.2 User plane KPIs

These types of KPIs are usually measured on a per call basis, but it does not necessarily imply that these should be stored always in this manner.

To get an estimate of the user plane load, it suffices to collect the information about the data that has been transferred for a longer period of time, between 15-60 minutes. Using this information, an average throughput can be computed, but this procedure is not usually applied. Instead, a better approach to measure the user experience of throughput is considered: it implies to collect throughput measurement samples during each call that is active in a particular cell and store and count the obtained results in a bin histogram table (the histogram of a discrete variable that can acquire m different values is called a “m-bin histogram”). The examination of the bin distribution of throughput measurement samples can be used to evaluate of the subscriber’s throughput quality experience, even if the above-mentioned distribution reflects the nature of the traffic rather than the experience of the user. Considering this, it appears that the throughput bin histogram approach is universally valid. It makes sense to evaluate the user experience when it comes to throughput-sensitive real-time and non-real time services (video streaming for example), but for an overall measurement of the user experience and the quality of the cell it is not applicable.

Another method to gain an insight into the user’s throughput experience is to measure the throughput of a particular connection and store the minimum, mean and maximum results obtained in the measurement in a call detail record.

The most accurate way to obtain throughput measurement information is to compute a set of measurements that have been sampled with a high sampling frequency during an ongoing connection. The obtained results are stored together with their corresponding timestamps in a database, which will be accessed in order to represent the results as a graph along the time axis of a measurement diagram or as a table. The greatest benefit from this kind of representation is that the throughput measurement results are represented graphically in parallel with other measurements or with the occurrence of signaling events that influence the data transmission’s speed.

All of the above principles can be applied to other user plane performance measurements.

7.2.1 IP throughput

It is defined as the data volume of IP frames that are transmitted within a defined time interval in the UL or DL direction. There is a series of different versions of IP (like IPv4, IPv6, etc.), so a suite of protocols should be supported by this measurement. However, for the throughput measurement it is not important which type of IP is used. What is really important is that the protocol frames on the IP layer to be correctly decoded.

The throughput measurement of user plane IP can be performed in various points of the EPC and E-UTRAN, yielding different results even within the time-span of the same call. The most logical way is to measure the IP throughput on the radio interface of the UE in order to evaluate the user experience in the most accurate way. Moreover, it makes sense to measure the IP throughput over the GTP tunnels of the S1-U and S5 and on the SGi interfaces. It is very difficult to make a correspondence between the SGi and a particular mobile subscriber because there is no tunnel associated with the subscriber.

It is easy to detect the sender and receiver of an IP packet using the source and destination address in the IP header. However, it is not an easy task to determine which one of these addresses belongs to the subscriber, information that would allow separate measurements for the UL and DL direction. In order to differentiate between the UL and DL frames, it is necessary to track temporary IP addresses during the attach procedure or to track the GTP tunnel endpoint identifiers assignment in the tunnel assignment procedure of GTP-C and S1AP, and associate this information with the throughput measurements.

The length field of the IP header provides the total size of the IP frame, which includes both header and data. By using this field, the volume of IP data can be measured.

7.2.2 Application throughput

The application throughput for non-real-time services is obtained by dividing by time the data volume measured on the seventh layer protocol of TCP/IP user plane protocol stacks. The measurement procedures are to a certain extent similar to the one presented for IP throughput, considering the fact that the different layer 7 applications can be differentiated by taking into account the UDP (User Datagram Protocol) and TCP port numbers.

Two aspects have to be dealt with in order to accurately measure application throughput: determine the correct volume of application data that has been transferred and determine the appropriate start and stop trigger points for the measurement.

To view the TCP and UDP flows in order to determine different application layers according to port numbers, it is necessary to look at layer 6 into the IP flow. Here it will be observed that UDP transmits immediately application data, while TCP needs a start-up procedure represented by the three-way handshake (SYN, SYN-ACK, ACK) that is not used to transmit any application layer data, although the port number of the application is included in this handshake. Thus, the application throughput on UDP starts immediately with the application data of the first UDP frame, where the application’s specific UDP port numbers are included. For TCP, the start trigger of the throughput measurement should be the first application layer message.

The biggest hindrance on measuring the application layer KPIs is represented by IPsec, which is usually used for mobile VPN connections, and which does not offer the possibility to decode TCP/UDP frames due to the ciphering that is performed within IPsec.

7.2.3 TCP startup KPIs

In the case of non-real-time services (file transfer, web-browsing, e-mail), ETSI TS 102.250 defines a set of KPIs concerning service startup failure ratio, service startup time, IP service startup failure ratio and IP service startup time.

The service setup time refers to the start trigger point of the first application frame that is sent. IP service setup considers the trigger points for the start and stop of delay measurements in the TCP layer. The trigger points of service setup and IP service setup can in fact be identical.

The service startup failure ratio is simple to be computed by using the protocol trigger points in a TCP environment: each TCP SYN message is considered as a TCP service attempt and the first ACK message of the TCP flow is considered as a success event.

7.2.4 Packet jitter

It represents the average of the deviations from the network mean latency, representing an important QoS parameter for real-time services that make use of UTP transport due to their delay sensitivity. The jitter has no impact on the user plane QoS for non-real-time services like web-browsing, e-mail or file transfer [54].

8. Simulation models and performance metrics

This chapter represents an introduction to the simulator which has been used, the system models, scenarios and a selection of the most important and basic simulation parameters. The scenario-specific parameters and their corresponding values will be given at the appropriate moment, in connection with the obtained results.

8.1 General aspects of the simulator

All simulations were conducted on a Matlab based simulator, developed by Technische Universität Wien and are freely available at the following link: http://www.nt.tuwien.ac.at/about-us/staff/josep-colom-ikuno/lte-simulators/ under an open, free for non-commercial academic use license. The latest versions of the LTE System Level simulator (v1.6r885) and of the LTE Link Level simulator (v1.7r1089) have been used.

In the process of development and standardization of LTE, as well as in the implementation stage of equipment manufacturers, simulations are needed to be performed in order to test and optimize algorithms and procedures. These simulations have to be carried on both the physical layer (link-level) and on the network (system-level) context.

Link – level simulations allow investigating and fine-tuning parameters such as MIMO gains, AMC feedback, modeling of channel encoding and decoding or the modeling of the physical-layer for system-level (in system level simulations the physical layer is abstracted from link-level results). System – level simulations focus more on network-related issues, like scheduling, interference management and mobility handling, investigating in this way network performance. In link-level simulations it is not possible to emphasize issues like cell planning, scheduling or interference. Simulating the totality of the radio links between the UEs and eNBs is not a practical approach to perform system level simulations because it is very computationally demanding. This is the main reason why in system level simulations the physical layer that is being employed is an abstraction that considers simplified models that capture the essential characteristics with high accuracy, by maintaining a low degree of complexity.

The link-level simulator is made up of three major blocks: transmitter, channel model and receiver. The transmitter and receiver are linked through the channel model, which is used to transmit the downlink data. Signaling and uplink feedback is assumed to be error-free, which is very realistic because signaling is subjected to a stronger protection than data by using lower coding rates and lower-order modulation schemes.

In the downlink, information such as coding, HARQ, scheduling and pre-coding parameters are transmitted, whereas in the uplink CQIs, PMIs and RIs are being sent, forming the so-called CSI (Channel State Information) feedback. This feedback is used by the scheduling algorithm to assign resources to users in order to optimize the performance of the system (in terms of throughput for example).

In the case of the transmitter, based on the feedback of UE, a scheduling algorithm assigns RBs to UEs and chooses the appropriate MCS, the MIMO transmission mode (OLSM, CLSM) and the number of spatial layers for all served users.

The Vienna LTE link-level simulator supports block-fading channels (in this case the channel is constant during the duration of one subframe, namely 1 ms), as well as fast-fading channels (in this case time-correlated channel impulse responses are generated for each sample of the transmit signal).

The following channel models are supported by both simulators: AWGN, flat Rayleigh fading, ITU Pedestrian B, ITU Vehicular A, Winner Phase II+ (which is the most sophisticated among these channel models and supports arbitrary 3D antenna patterns, allowing thus the investigation on the impact of antenna tilting – be it electrical or mechanical -on the system achievable performance).

The simulation is performed by defining a ROI (Region of Interest) in which the eNBs and UEs are distributed and a simulation length expressed in TTIs. The DL-SCH is simulated only within the ROI.

The system level simulator consists of two parts: a link measurement model and a link performance model. The link measurement model determines the link quality based on the reports given by the UE, performing afterwards link adaptation and resource allocation. In order to abstract the measured link quality, the SINR per subcarrier is utilized as a metric. The output of the link measurement model (the receiver SINR, modulation and coding parameters) is used by the link performance model to predict the BLER of the link. The CQI feedback report is obtained by considering the subcarriers' SINRs and the target BLER. The CQI reports are generated by a SINR-to-CQI correspondence, as shown in the figure below, and are transmitted to the eNB via a feedback channel. At the transmitter, the most suitable MCS is selected based on the reported CQI to achieve the target BLER.

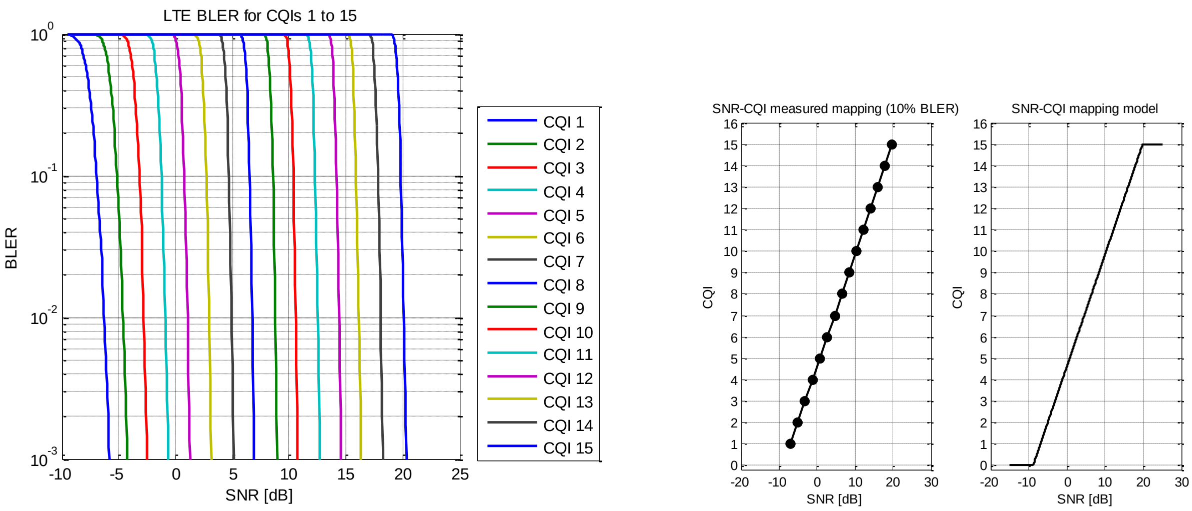

Figure 8.1 BLER to CQI mapping (left) and SNR to CQI mapping (right)

In order to consider the spatial and time correlation of the channel present in a wireless cellular system, macroscopic pathloss, shadow fading and small-scale fading are considered.

The macroscopic pathloss between an eNB sector and an UE models both the propagation pathloss caused by the distance and the antenna gain.

Shadow fading can be regarded as the irregularities of the geographical characteristics of the terrain that are represented by the obstacles in the propagation path between the UE and the eNB. It introduces an additional attenuation with respect to the average pathloss yielded by the macroscopic pathloss model. It is usually approximated by a log-normal distribution of mean 0 dB and standard deviation 10 dB.

The losses caused by the macroscopic pathloss and the shadow fading are dependent on the position and are time-invariant, whereas small-scale fading is modeled as a time-dependent process.

For each of the modeled MIMO transmission modes, a Zero Forcing receiver has been developed. A Zero Forcing equalizer represents a form of linear equalization algorithm used in communication systems which applies the inverse of the frequency response of the channel to the received signal in order to restore the signal after its propagation through the channel. The name zero forcing comes from the fact that ISI would be reduced to zero in a noiseless channel [55].

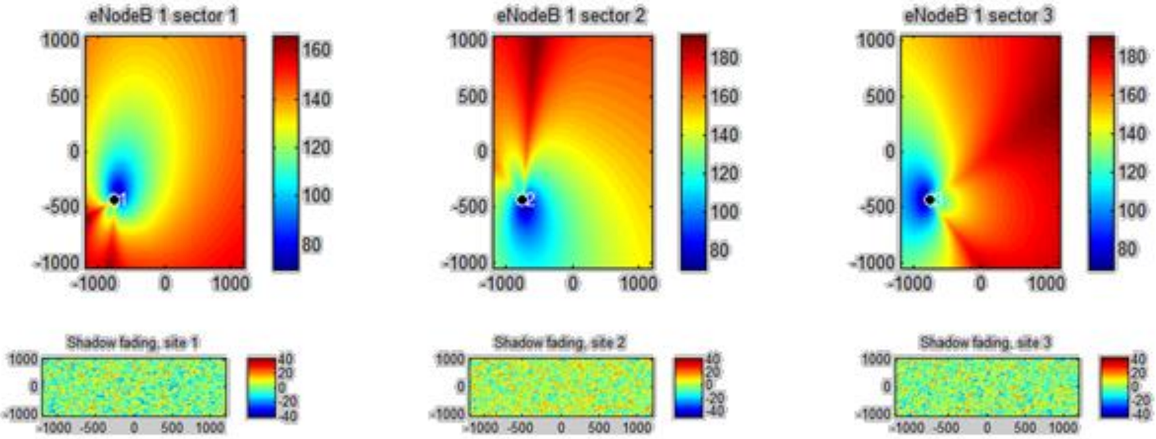

Figure 8.2 Macroscopic pathloss of the three sectors of an eNB (up) and the corresponding space correlated shadow fading map of the same site (down)

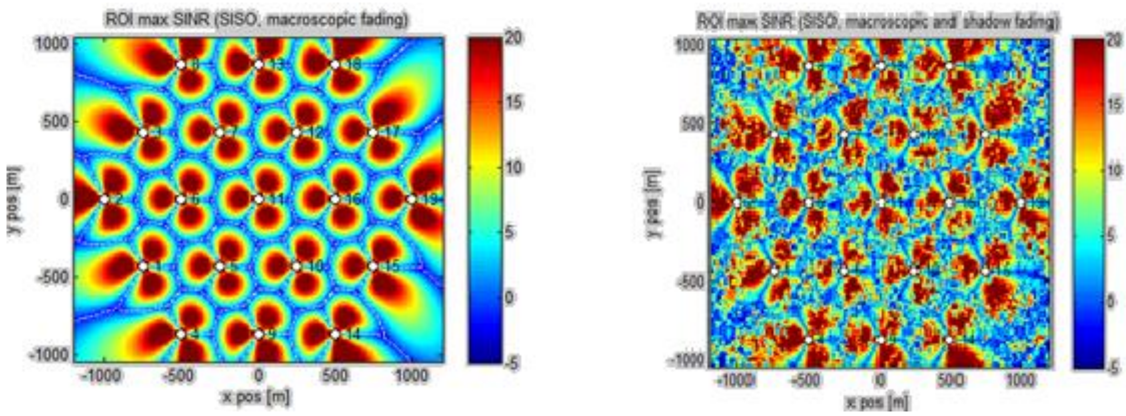

Figure 8.3 Sector SINR, computed with distance dependent macroscopic pathloss only (left) and with lognormal distributed space-correlated shadow fading superimposed (right)

The simulation outputs traces that contain link throughput and error ratios for each UE, as well as aggregated cell statistics, from which statistical distributions of throughputs and errors are extracted.

8.2 Simulation parameters and scenarios

For the system model, a hexagonal grid layout of cells, with three regularly shaped sectors each has been employed. Each site is equipped with three directional antennas, spaced 120° apart. The basic simulation parameters listed in the table below:

Carrier frequency	1.8 GHz
MIMO ($N_{TX} \times N_{RX}$)	2 × 2
MIMO mode	CLSM
Simulation time	1000 TTIs
Inter eNB distance	500 m
Number of eNB rings	2
Number of eNBs	57 (2 rings, 19 sites). Only 21 inner cells taken into account for the results
Considered UEs	center 7 sites(21 cells)
Minimum coupling loss	70 dB
Macroscopic pathloss model	TS36942 urban
eNB transmit power	43 W
Channel model	Winner II+
Length of the trace	10s
Shadow fading	none
Thermal noise spectral density	-174 dB/Hz
Antenna type	742212
Antenna gain pattern	Kathrein TS Antenna
Maximum antenna gain	15 dBi
Altitude of site	0 m
Height of site	20 m
Receiver height	1.5 m
Antenna mechanical downtilt	0°
Antenna electrical downtilt	7°
Antenna frequency	1.8 GHz
Feedback channel delay	3 TTIs
Feedback:	AMC: CQI; MIMO: PMI and RI
UE distribution	constant number of UEs per cell
Additional penetration loss	outdoor (0 dB)
Receiver model	Zero Forcing
Traffic model	Full buffer

Table 8.1 Basic configuration parameters

The minimum coupling loss describes the minimum loss in signal, expressed in dB, between the eNB and UE or UE and UE in the worst case and is defined as the minimum distance loss including antenna gains measured between antenna connectors. Recommended values are 70 dB for urban areas, 80 dB for rural areas.

The macroscopic pathloss model that has been used is TS36492 for the urban environment, described by the following relationship:

$$L = 40 \times (1 - 4 \times 10^{-3} \times Dhb) \times \log_{10}(R) - 18 \times \log_{10}(Dhb) + 21 \times \log_{10}(f) + 80 \text{ dB},$$

where R is the eNB-UE separation in km, f the carrier frequency in MHz and Dhb is the base station antenna height in meters, measured from the average rooftop level. This environment was chosen because there are both indoor and outdoor users who are covered by outdoor base stations, adding in this way more heterogeneity.

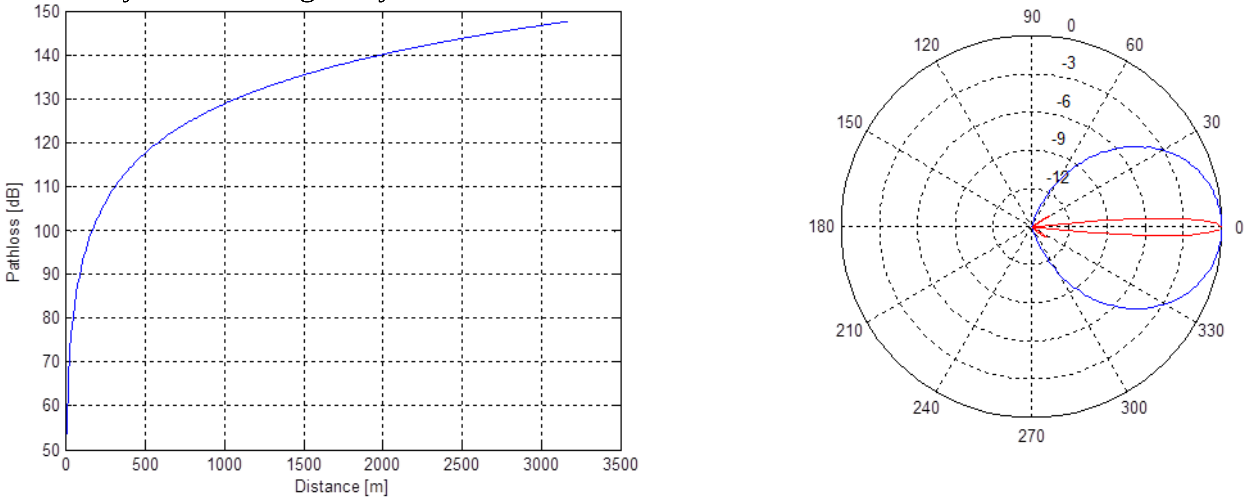

Figure 8.4 Macroscopic pathloss, using TS 36.492 urban area model (left) and 742212 antenna radiation pattern, blue: hor, red: vert [dBi], no electrical tilt (right)

All of the simulated cases were based on full buffer traffic model. In this model, each user has an infinite amount of data in the buffer to transmit or receive, depending whether it is UL or DL. Although this is not a real-life assumption, full buffer traffic model serves as a good base-line traffic model since scheduling and user throughputs are independent of the amount of data in the transmit buffer, thus making it easier to analyze the trends and user behaviors.

Shadow fading was omitted in order to be able to characterize the cell boundary by means of a SINR threshold Γ_{thr}.

Fairness rates how equally a resource (in this case throughput) is distributed among N users, and is defined as:

$$J(x) = \frac{(\sum_{i=1}^{N} x_i)^2}{N \times \sum_{i=1}^{N} x^2}$$

where x is a vector of length N, containing the throughput of each of the N users.

The following scenarios have been simulated:

	Bandwidth (MHz)	MIMO ($N_{TX} \times N_{RX}$)	Number of UE/cell	UE speed (km/h)	Scheduler	MIMO type
Scenario 1	20	2 × 2	1	0	Proportional Fair	CLSM
Scenario 2	10	2 × 2	1	0	Proportional Fair	CLSM
Scenario 3	20	2 × 2	5	0	Proportional Fair	CLSM
Scenario 4	20	2 × 2	5	0	Round Robin	CLSM
Scenario 5	20	2 × 2	5	0	Best CQI	CLSM
Scenario 6	20	2 × 2	5	3.6	Best CQI	CLSM
Scenario 7	20	2 × 2	5	50.4	Best CQI	CLSM
Scenario 8	20	2 × 2	5	50.4	Best CQI	OLSM

Table 8.2 Simulation scenarios

In all scenarios, a carrier frequency of 1.8 GHz was used. A constant number of UEs per cell was considered, with a constant speed as shown in the table above with randomly and uniformly distributed direction. In most of the cases the UE were stationary in order to obtain results that could be accurately compared with the measurements performed in the laboratory. The last two scenarios were conducted with a lower and a higher speed of the UE, while keeping all the other parameters unchanged, in order to emphasize the impact of the UE speed on various KPIs. The CLSM technique was preferred in most of the cases because it is thoroughly deployed in cellular networks nowadays and the UEs was small enough in order to allow the UE to send the CSI reports in a timely and relevant manner. The OLSM technique was chosen in the case of the greatest UE speed in order to allow a comparison to be made between the results yielded by the two techniques.

All simulations were performed on a system with a dual-core processor with the frequency of 2.5 GHz and 4 GB of RAM. Matlab 2011b was used with parallelization in order to allow for parallel runs of the simulator via parfor loops.

The following results have been obtained for the downlink:

Cell statistics	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8	Measurement unit
Fairness index	0.745082	0.825598	0.876853	0.782945	0.303231	0.294910	0.196853	0.216837	NA
Peak UE throughput (95%)	77.72	40.12	17.20	16.00	67.63	65.74	36.04	33.18	Mb/s
Average UE throughput	38.43	22.98	11.36	8.53	14.99	14.87	7.11	6.58	Mb/s
Edge UE throughput (5%)	13.68	8.91	5.07	2.99	0.00	0.00	0.00	0.00	Mb/s
Average cell throughput	38.43	22.98	56.79	42.67	74.94	74.33	36.24	33.19	Mb/s
Mean RB occupancy	99.84%	99.89	100.00%	99.90%	99.70%	99.70%	99.70%	99.70%	NA

Table 8.3 Simulation cell statistics

Simulation statistics	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7	Scenario 8	Measurement unit
Number of UEs	21	21	105	105	105	105	107	106	NA
Average UE throughput	38.43	22.98	11.36	8.53	14.99	14.87	7.11	6.58	Mb/s
Average UE spectral efficiency	2.29	2.74	3.35	2.54	3.62	3.49	0.88	0.82	bit/cu
Average RBs/TTI/UE	99.84	49.94	20.00	19.98	20.00	20.00	19.64	20.19	RBs
Rank indicator (RI) distribution: rank 1	99.83%	98.41%	98.16%	98.63%	98.02%	95.68%	94.03%	86.13%	NA
rank 2	0.17%	1.59%	1.84%	1.37%	1.98%	4.32%	5.97%	13.87%	NA
Simulation time	36.00	20.5	70	60.5	103	107	102	102	minutes

Table 8.4 Simulation statistics

Three KPIs, which are derived from the throughput ECDF (Empirical Cumulative Distribution Function) and do not take into account the effects of scheduling consist of: mean throughput, edge throughput and peak throughput. The terms “edge” and “peak” refer to the 5% and 95% points of the UE throughput ECDF, respectively. These can be interpreted as the performance of a UE at the cell edge and at cell center, respectively.

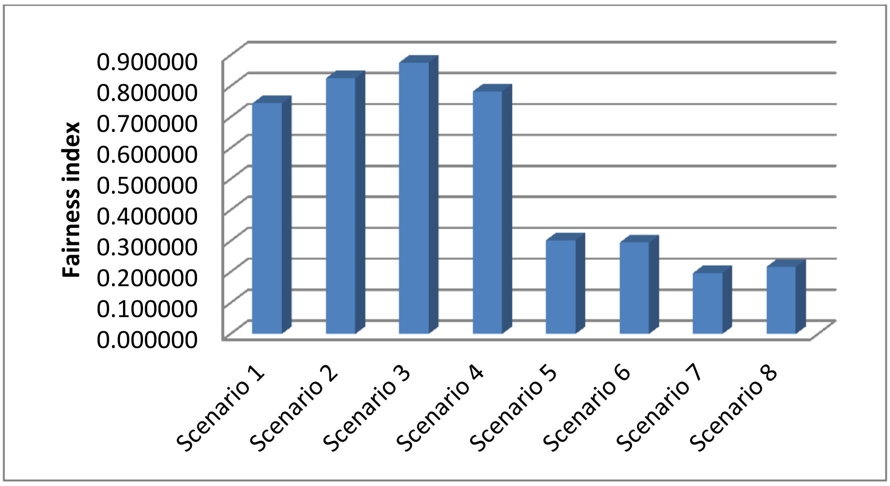

Figure 8.5 Fairness Index

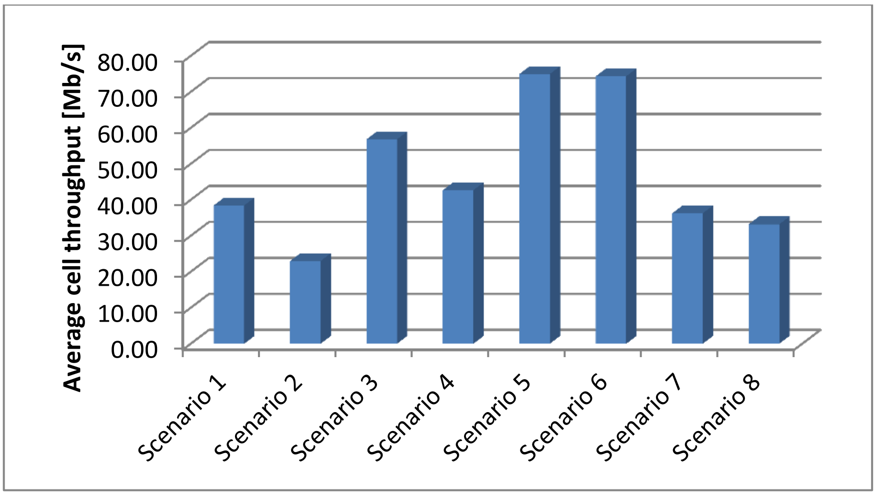

Figure 8.6 Average cell throughput

A performance comparison of different scheduling strategies is shown in Figure 8.5 in terms of fairness (Jain’s fairness index). The considered schedulers pursue different goals for resource allocation. The best CQI scheduler tries to maximize total throughput and completely ignores fairness by assigning resources to the users with best channel conditions. This is reflected in the simulation results in figure 8.5 and 8.6, showing the highest average cell throughput and the lowest fairness for the best CQI scheduler. Lower throughputs were obtained where the same best CQI scheduler was used, but when the UE was not any more stationary, but moving with speeds of 3.6 km/h and 50.4 km/h, highlighting that UE speed influences drastically the throughput. Round robin scheduling does not consider the UE feedback and cyclically assigns the same amount of resources to each user. In this way, ignoring the UE feedback should theoretically result in the worst throughput performance of all three schedulers considered here, this being also the case in our scenarios. Comparing Scenario 3, 4 and 5, where the same bandwidth, number of UE/cell and UE speed are considered and three different types of schedulers are employed, the results predicted by theory are confirmed. The proportional fair scheduler emphasizes multiuser diversity by scheduling the user that has the best current channel conditions relative to its own average. Proportional fair scheduler registers the highest fairness index and the second largest throughput, thereby resulting in a good tradeoff between throughput and fairness.

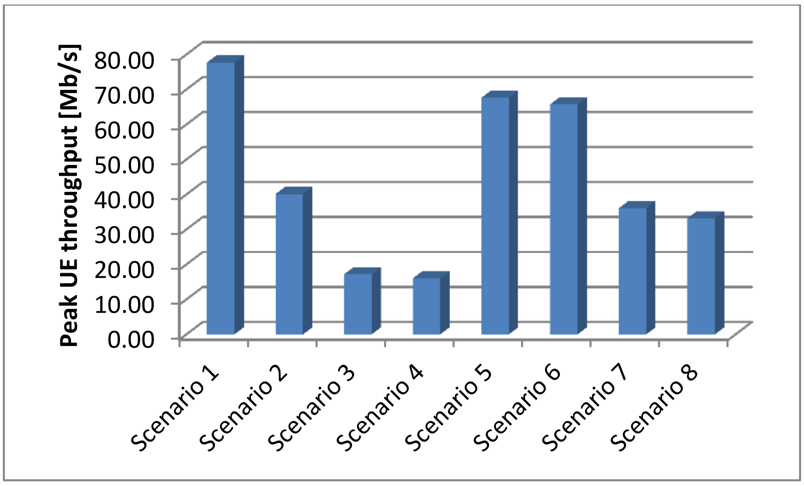

Figure 8.7 Peak UE throughput

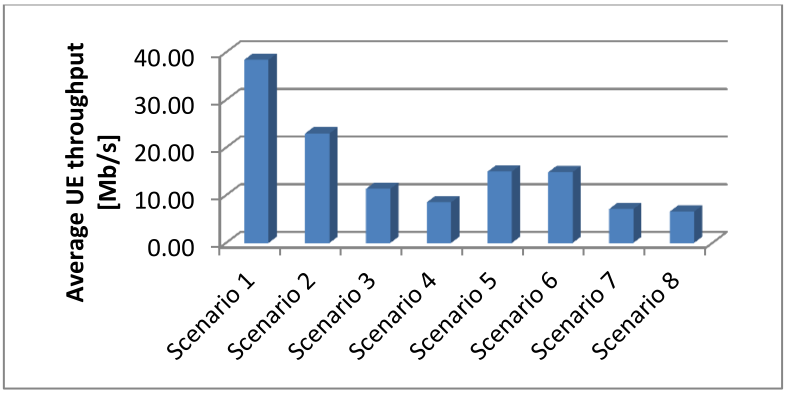

Figure 8.8 Average UE throughput

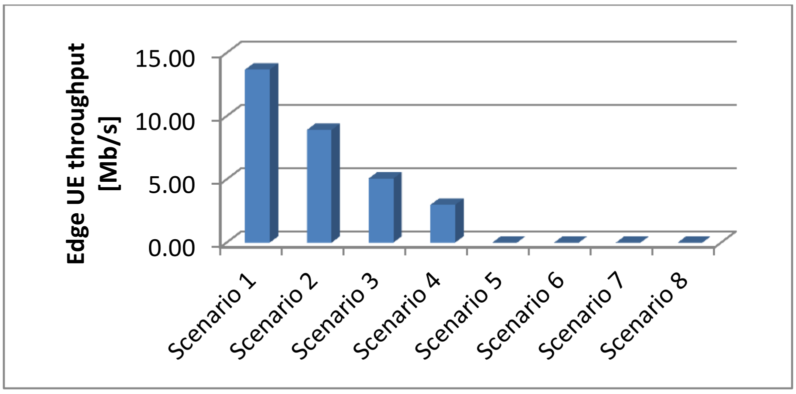

Figure 8.9 Edge UE throughput

When we consider the throughput the UE experiences at different regions in the cell, very interesting results are obtained. When it comes to the throughput registered in the center of the cell (peak UE throughput), the highest value is obtained when a BW=20 MHz is used and a single stationary UE/cell is present, offering a throughput in the downlink of 77.72 Mbps, using the proportional fair scheduler. Using the same scheduler and parameters, but with 5 UEs/cell, the throughput per UE decreases almost 5 times, reaching 17.2 Mb/s, remaining practically unchanged when the round robin scheduler is used. Employing the best CQI scheduler yields a peak DL throughput of 67.63 Mb/s, which decreases steadily when the UE speed increases, reaching almost half the throughput experienced by a stationary UE at a speed of 50 km/h. The average and edge UE throughput follow a similar behavior. What is interesting to remark is that OLSM registers slightly slower speeds than CLSM, and this is because OLSM does not require UE feedback, like CLSM does, which translates into a decrease in throughput. However, OLSM is preferable to CLSM when it comes to fast users and a feedback from a UE could not be issued and used in a timely manner. Another interesting aspect worth mentioning is that when the best CQI scheduler is employed, regardless of the throughput experienced by a UE at the edge of a cell is 0 Mbps, which is obviously not desirable in a real-life scenario. This is caused by the fact that at cell-edge, radio conditions are poor and this scheduler allocates radio resources to users with the best radio conditions, omitting thus marginal users.

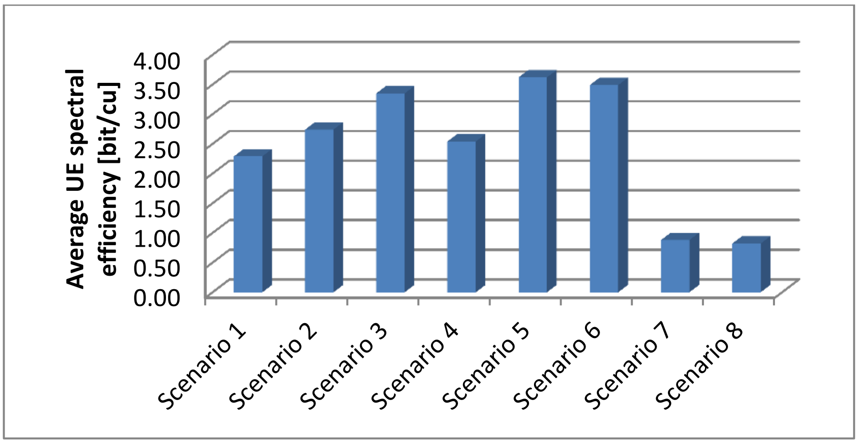

Figure 8.10 Average UE spectral efficiency

Spectral efficiency refers to the information rate that can be transmitted over a given bandwidth in a specific communication system. It measures how efficiently a limited frequency spectrum is utilized by a physical layer protocol, and sometimes by the media access control. It is usually measured in bit/s/Hz, or alternatively, it can be measured in bit/symbol, which is equivalent with bit/cu (bit per channel use), implying that the net bit rate (useful information rate excluding error-correcting codes) is divided by the symbol rate (modulation rate). The greatest spectral efficiency is achieved with the best CQI scheduler, reaching 3.62 bit/cu. This is because the best CQI scheduler allocates resources to the UE with the best radio conditions, maximizing in this way the information that can be sent in a certain bandwidth without additional redundancy that would have to be introduced in order to protect the data from being altered, maximizing thus the throughput, and thus the spectral efficiency. The proportional fair scheduler achieves the second largest spectral efficiency among the considered schedulers, achieving 3.35 bit/cu, a little less than best CQI because it allocates resources to the users who have the best channel quality with respect to their average channel quality. In a system with the RR scheduler, once a user is served by the eNB, it is not served again until all the other users in the system have been served. Thus, the RR scheduler has the same average spectral efficiency as a random scheduler, which schedules all users with the same probability and does not take into account the channel states of the users, achieving a spectral efficiency of 2.54 bit/cu, the smallest spectral efficiency of the three considered schedulers.

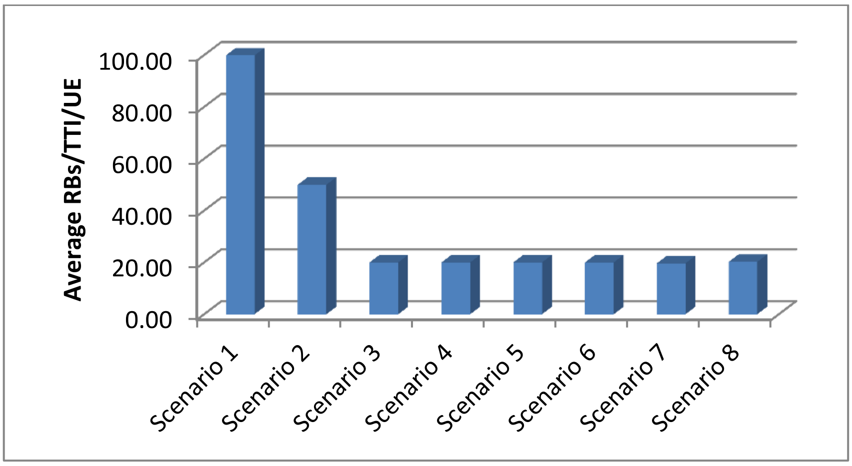

Figure 8.11 Average RBs/TTI/UE

When it comes to transmission bandwidth configuration, as the bandwidth increases the number of RBs increases. Thus, to the channel bandwidths of 1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz, 20MHz correspond 6, 15, 25, 50, 75, 100 RBs. The figure above illustrates exactly this situation: when a single UE/cell is present and a 20 MHz bandwidth is used, the UE is allocated 99.84 RBs. When the bandwidth decreases at 10 MHz, but the same conditions are kept, 49.94 RBs are allocated. When 5 UE/cell are considered and a bandwidth of 20 MHz is used, the 100 RBs are distributed evenly among the users, thus 20 RBs/UE are allocated.

In the figure below, the positions of eNBs and UEs are represented. Shaded via the selection of the appropriate cells on the cell list: approximate cell area corresponding to the center eNodeBs. The UEs attached to the selected eNodeBs are shown in black. The rest of the UEs are shaded.

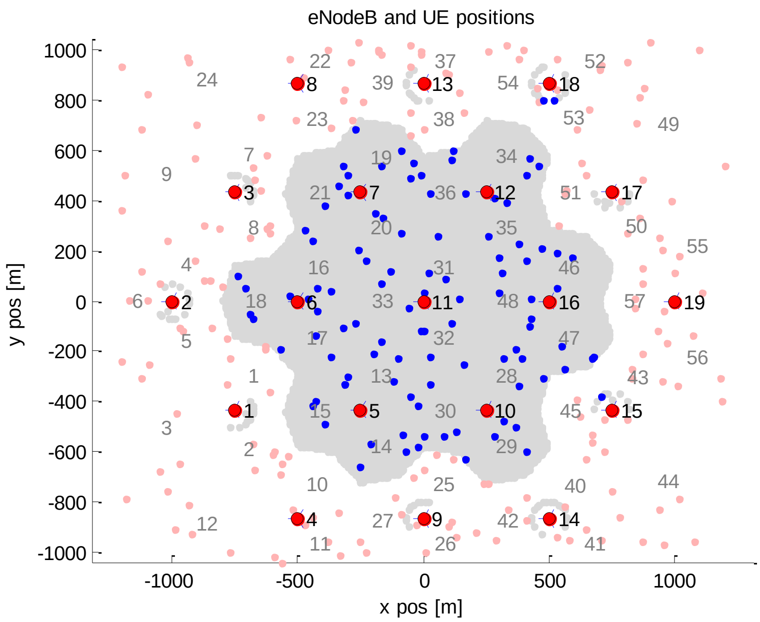

Figure 8.12 eNB and UE positions in the ROI

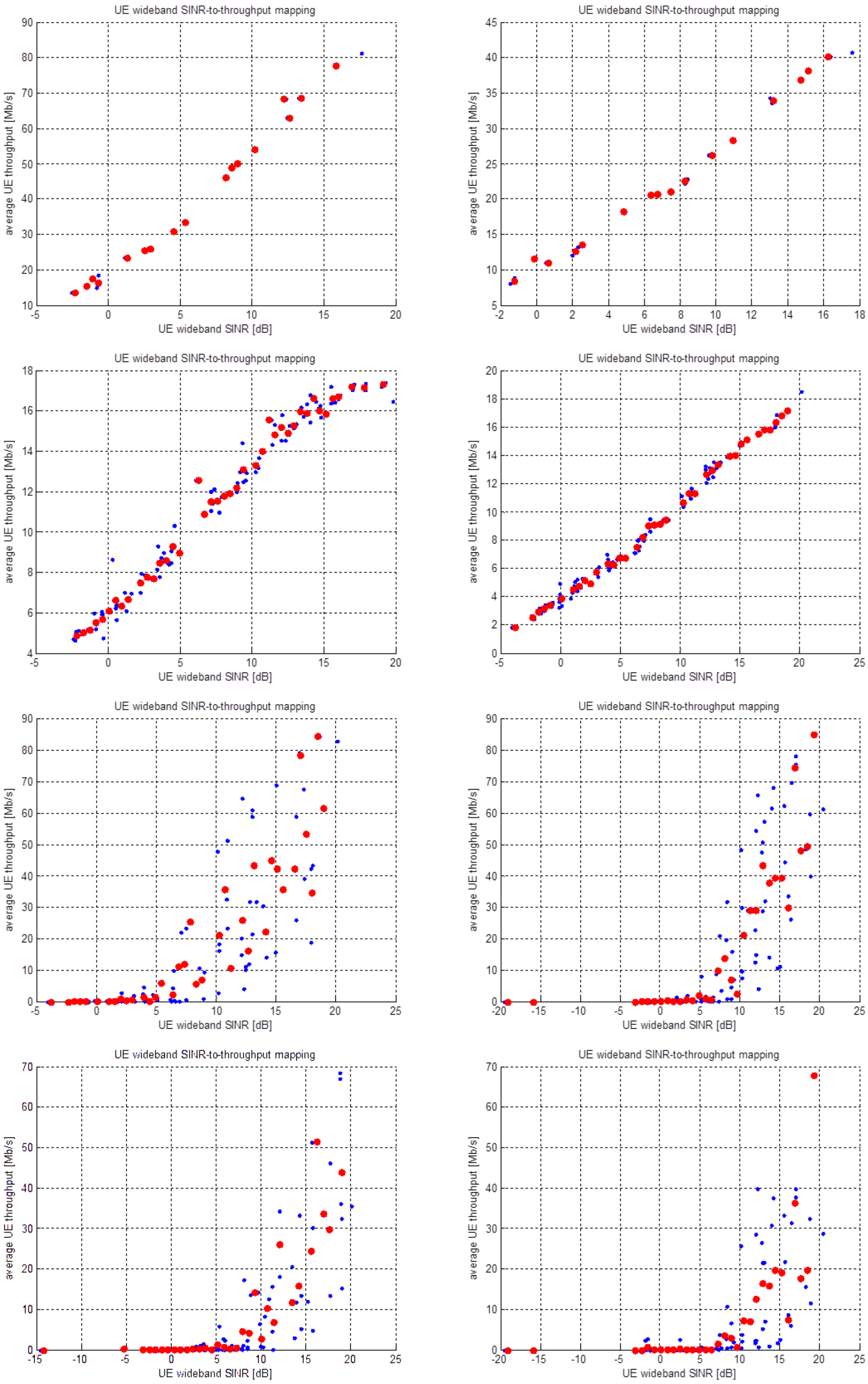

Figure 8.13 UE wideband SNR-to-throughput mapping scenario 1 (upper left) to scenario 8 (lower right)

The scatterplots above show for each UE the mapping between the wideband SINR and the throughput. Since many points could overlap, the binned mean values of the throughput over the wideband SINR is represented in red. It can be observed that an increase with 5 dB in the wideband SINR translates in a throughput increase of about 20 Mbps in the case of 1 UE/cell, and around 4-5

Mbps when 5 UE/cell are considered and when proportional fair and round robin schedulers are used. It can also be observed that these schedulers yield a more compact distribution of throughput vs SINR than best CQI, where the results are much more scattered and have a plateau at 0 Mbps. The SINR level at which the throughput is nonzero represents the SINR threshold Γ_{thr}, which defines the cell boundary. In most of the cases, this is between 5-7 dB. When the UE speed increases, this SINR threshold also increases, which means that in order to achieve the same throughput as a stationary user, a mobile user has to have better radio conditions.

Simulations were also performed in LTE Link Level Simulator in order to test the performance of an LTE transmission on an uncorrelated ITU Pedestrian B channel at 3 km/h and flat Rayleigh channel for several transmission modes. BLER and throughput were evaluated, choosing a value of 7 for the CQI and running the simulation for 5000 TTIs (5 seconds).

In wireless channels, fading may be caused by multipath propagation, generating the so-called multipath fading, or due to obstacle shadowing affecting the wave propagation, which is called shadow fading. Fading is usually mathematically modeled as a time-varying random change in the amplitude and phase of the transmitted signal. Flat fading occurs when the coherence bandwidth (the maximum bandwidth over which two frequencies of a signal are likely to experience comparable or correlated amplitude fading) of the channel is larger than the bandwidth of the signal, all frequency components of the signal experiencing the same magnitude of fading. The Rayleigh fading is a statistical model that assumes that the amplitude of a signal passing through a communications channel will fade according to a Rayleigh distribution. This type of fading is usually used to model the effect of dense urbanization on radio signals.

The basic settings used in the simulator are listed in the table below.

Parameter	Value
Number of UEs	1
Bandwidth	1.4 Mhz
Retransmissions	0 and 3
Channel type	Flat Rayleigh and uncorrelated PedB
Simulation length	5000 TTIs
Transmit modes	SISO, TxD (2x1 and 4x2) and OLSM (4x2)

Table 8.5 Basic settings used for the 1.4 MHz bandwidth simulation

The following results have been obtained, as it can be seen in the figures below.

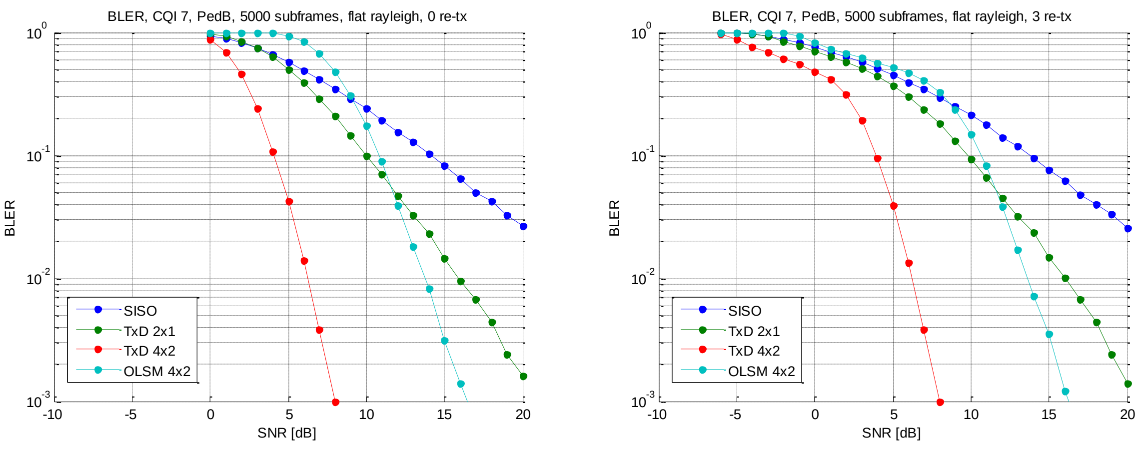

Figure 8.14 BLER, flat Rayleigh channel, no HARQ (left) and 3 retransmissions (right)

In the figure above, the BLER is viewed comparatively for no HARQ and three retransmissions respectively. In the case of transmit diversity 4×2, the best results are obtained because a certain BLER is achieved for the lowest SNR among the ones considered. It can be observed that the two techniques yield the same results for positive SNR, achieving a BLER of 10% for SNR=4 dB, an increase of 2 dB in the SNR resulting in a BLER of 1%. The second variant of transmit diversity, with two Tx antennas and only one Rx antenna, obtains the same results as the previous technique, but at SNR levels higher with 5 – 10 dB than TxD 4×2. OLSM 4×2 closely resembles TxD 2×1, whereas SISO is by far the least effective technique in terms of BLER, requiring a SINR of about 30 dB to achieve a BLER of 1%, a level with about 15 dB greater that TxD 2×1 and with about 25 dB higher than TxD 4×2. When three retransmissions are considered, the same results are obtained for BLER lower than 10%, the decrease to this threshold being more abrupt than in the first case for all techniques.

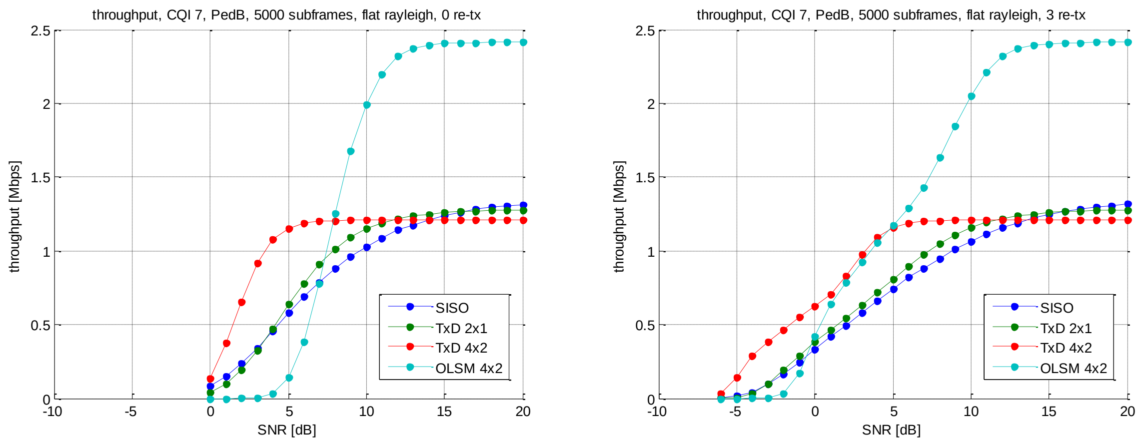

Figure 8.15 Throughput, flat Rayleigh channel, no HARQ (left) and 3 retransmissions (right)

When it comes to throughput in the case when no retransmissions are considered, OLSM 4×2 achieves a throughput two times greater than any other technique for a SNR greater than 12-13 dB, plateauing at a value of 2.4 Mbps. TxD 4×2 is the first technique to reach its maximum throughput of 1.2 Mbps at a SNR of 5 dB, whereas the two remaining techniques attain the same maximum throughput, but at a SNR of 12 dB, remaining constant at 1.3 Mbps at levels greater than 13 dB for the SNR. When three retransmissions are considered, the maximum throughput of each technique remains unaltered, as well as the SNR level at which now increase is observed. However, up to this threshold, which is 5 dB for TxD 4×2 and 12 dB for all other techniques, the increase is less steep than in the first case. It can be concluded that the number of retransmissions has a positive impact on the throughput only when radio conditions are poor, allowing to obtain higher throughputs for the same low levels of SNR than in the case where HARQ is not implemented.

The ITU Pedestrian B channel is described by a tapped delay line model, with a maximum of six taps, as shown in Figure 8.16.

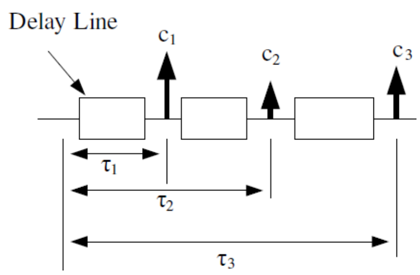

Figure 8.16 Tapped delay line model, with three taps

Each tap is characterized by a delay, denoted by τ_i, and an amplitude coefficient, denoted by c_i. By using this model, the out signal can be expressed as a linear combination of attenuated and shifted copies of the original signal. The first tap corresponds to the line of sight component, which has the greatest amplitude and the smallest delay, the other taps are due to various obstacles.

ITU recommends that the pedestrian models, which are designed to model either indoor or outdoor pedestrian environments, to be used to represent multipath conditions in micro-cells. These environments are characterized by low transmission powers, low antenna heights and low mobility (3-4 km/h).

The Pedestrian B channel model uses the parameters shown in the table below. The model is specified based on the number of taps that reach the receiver, the time delay relative to the line-of-sight tap, the average power relative to the strongest tap (which is usually the line-of-sight tap) and the Doppler spectrum of each tap.

Tap No.	Channel B		
	Relative Delay (ns)	Average Power (dB)	Doppler Spectrum
1	0	0.0	Classic
2	200	-0.9	Classic
3	800	-4.9	Classic
4	1 200	-8.0	Classic
5	2 300	-7.8	Classic
6	3 700	-23.9	Classic

Table 8.6 Pedestrian B channel parameters [56]

In the figure below, an ITU PedB channel is used to model the wireless environment, with a subject velocity of 3 km/h. The main difference between the cases when HARQ is not implemented in comparison with the one when three retransmissions are used is that the decrease from a BLER of 100% to a BLER of 10% is less steep in the former case, but the SNR levels at which certain BLER values are obtained remain unchanged. The most notable difference between the results obtained here and the ones obtained in the flat Rayleigh channel is that the SNR at which a BLER of 1% is achieved is roughly 5 dB lower in the former case than in the latter. Another observation worth mentioning is that SISO yields better results here, achieving a BLER of 1% for a SNR of 15 dB, which is roughly 10 dB lower than in the case of flat Rayleigh channel.

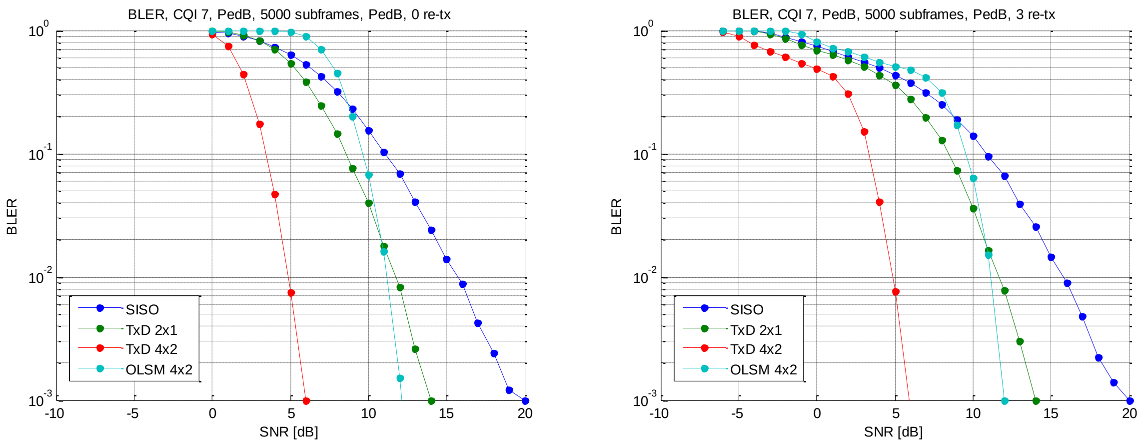

Figure 8.17 BLER, PedB channel, no HARQ (left) and 3 retransmissions (right)

In the figure below, we can observe the throughput evolution with respect to the SNR level, the behavior encountered here being practically identical with the one obtained in the case of a flat Rayleigh channel.

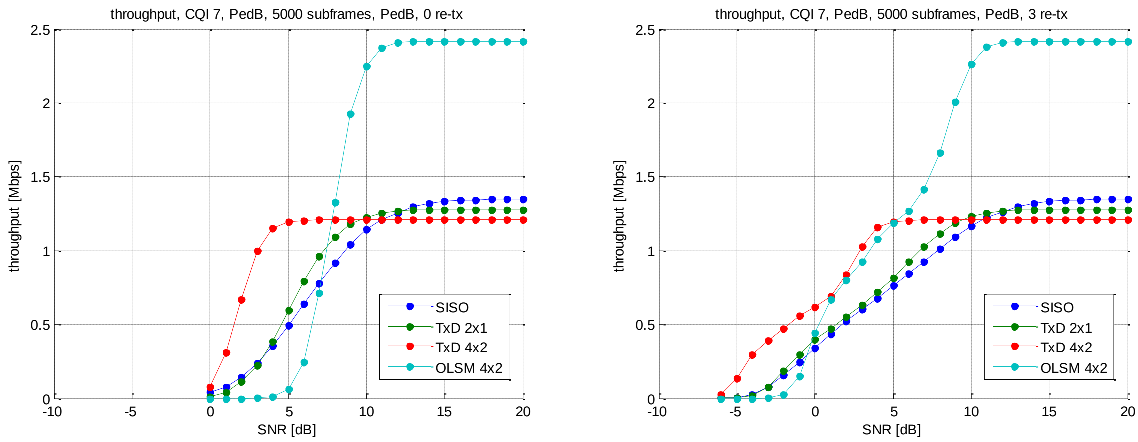

Figure 8.18 Throughput, PedB channel, no HARQ (left) and 3 retransmissions (right)

Another simulation in LTE Link Level Simulator has been carried out for the 20 MHz bandwidth and a carrier frequency of 1800 MHz in order to analyze the behavior of different types of transmit modes when a RR scheduler that allocates resources to stationary UEs is used. A zero forcing receiver has been considered, together with an uncorrelated Pedestrian B channel. The basic settings for this simulation are listed in the table below.

Parameter	Value
Number of UEs	1
Bandwidth	20 MHz
Retransmissions	0 and 3
Receiver	Zero Forcing
Scheduler type	Round Robin
UE speed	0 km/h
Scheduler assignment	static
Channel Type	uncorrelated Pedestrian B
Simulation length	1000 TTIs
Transmit modes	SISO, TxD (2x1 and 4x2) and OLSM (4x2)

Table 8.7 Basic settings used for the 20 MHz bandwidth simulation

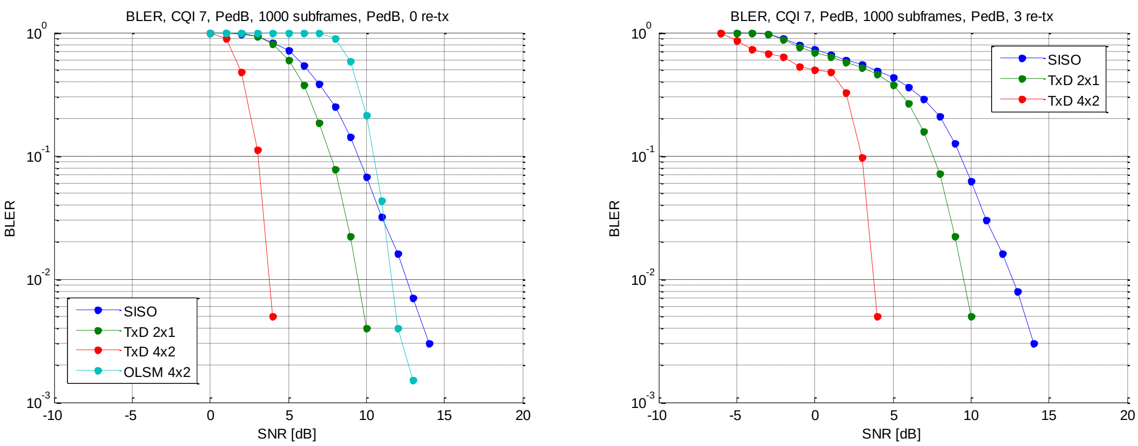

Figure 8.19 BLER, PedB channel, no HARQ (left) and 3 retransmissions (right)

In the figure above, the BLER-to-SNR dependency is shown. Again, the transmit diversity TxD 4×2 obtains the best results in terms of achieving a certain BLER rate at the lowest SNR level among the four considered cases. Even if the results are very similar between the case where no retransmissions are considered, in comparison with the case when three are taken into account, the HARQ scenario obtains slightly better results. Moreover, in the case of HARQ with three retransmissions it is possible to obtain, for example, at a SNR level of 0 dB a BLER of 49.9%, whereas in the case when no HARQ is considered, for the same SNR level the BLER is 99.5%. This observation strengthens the affirmation that HARQ facilitates the communication in poor radio conditions, a BLER of 99.5% being obtained in this case for a SNR of -6 dB. In the other transmission modes a similar behavior is observed, with practical identical results for SNR levels greater than 5 dB and with better BLER rates for SNR values between -6 and 5 dB than in the first case.

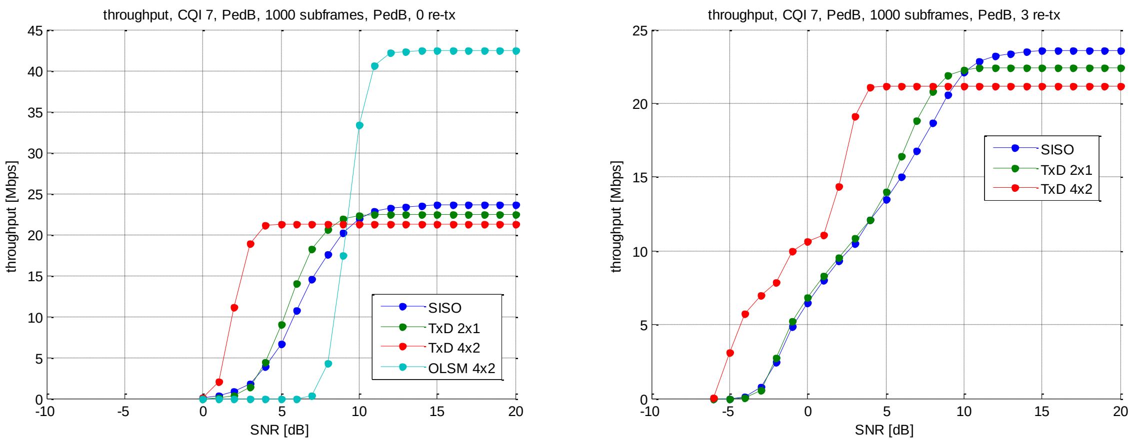

Figure 8.20 Throughput, PedB channel, no HARQ (left) and 3 retransmissions (right)

In terms of throughput, the allure of the results is very much alike, with the observation that at the same levels of SNR, the throughput registered when three retransmissions are employed is with about 7 Mbps greater than the one experienced when no HARQ is used. Thus, in the first case (no HARQ), at a SNR of 0 dB, no matter the transmit mode, the throughput experienced by the UE is 0 Mbps, whereas in the second case, when HARQ is implemented, at the same SNR level, SISO and TxD 2×1 register a 6.9 Mbps throughput, while TxD registers the highest throughput, namely 10.62 Mbps. This mode reaches its maximum throughput of 21.2 Mbps at a SNR level of 5 dB and afterwards an increase in the SNR level does not provide any more throughput gains. At SNR levels greater than 12 dB, all the other transmit modes also reach their maximum throughput and stagnate at this value. In this case, SISO and TxD 2×1 yield similar maximum throughputs, 23.56 Mbps and 22.37 Mbps respectively, whereas OLSM 4×2 registers 42.39 Mbps, which represents roughly a 50% increase.

8.3 Laboratory measurements

A series of measurements were performed in the laboratory in order to obtain a series of results to be compared with the ones obtained in the Matlab simulations. Three scenarios have been considered: first, the measurements were performed in a Faraday cage in order to isolate the receiver from external interferers and obtain results as close as possible to the ones obtained through simulation. After this, measurements were conducted, where the inter-antenna distance of the transmitter was very small and larger than 1.5 m. The throughput was measured with a FTP-based client in the laboratory and using speedtest.ro. The results were afterwards averaged. In order to maintain the uniformity with the simulated scenarios, a 2 × 2 MIMO configuration was used. The obtained results are shown below.

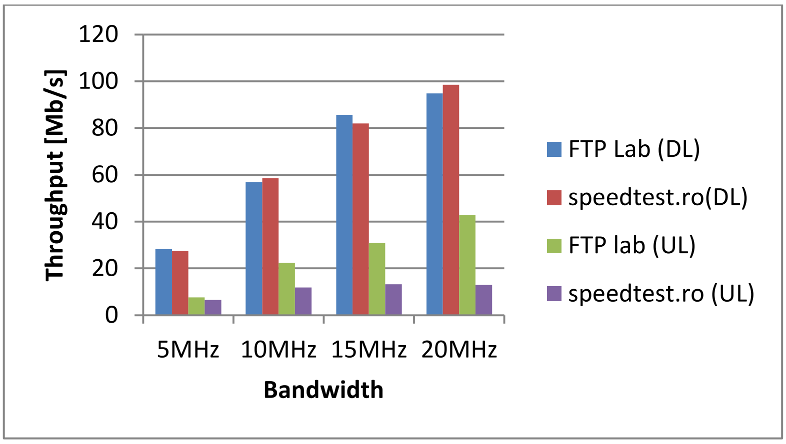

Figure 8.21 Throughput in Faraday cage

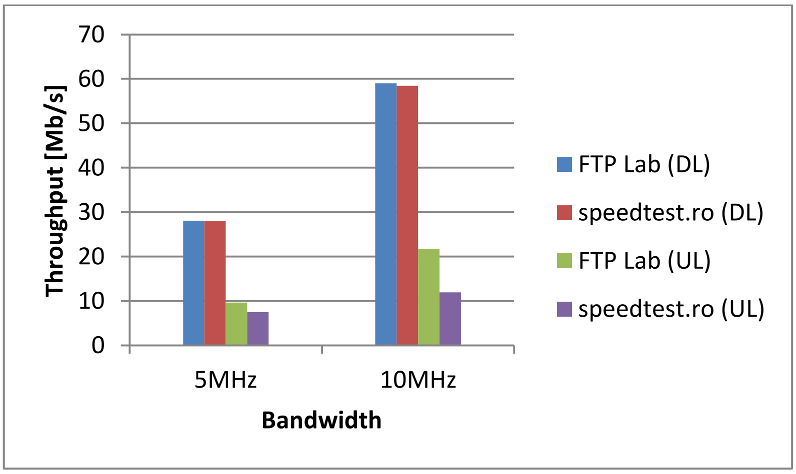

Figure 8.22 Throughput when antennas are more than 1.5 m apart

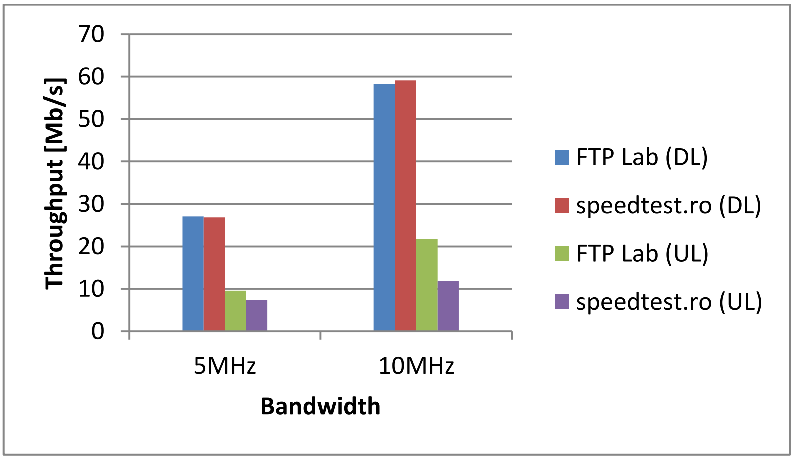

Figure 8.23 Throughput when antennas are close to each other

When it comes to the measurements performed in the Faraday cage, speeds of about 57 Mbps for a 10 MHz bandwidth were obtained, whereas for a bandwidth of 20 MHz a throughput of roughly 97 Mbps was obtained, which is about 17 Mbps more than in the results yielded by the simulations. In the UL, for a Bw = 10 MHz speeds of up to 22.42 Mbps were achieved, whereas a doubling in the bandwidth translated into, roughly, a doubling in throughput, achieving 42.82 Mbps.

When the antennas were more than 1.5 m apart, which at the working frequency of 1.8 GHz means 9λ (λ=0.1666 m), the spatial correlation between the antennas was small, and thus the obtained throughput was about 1 Mbps higher in the DL than in the case of closely-placed antennas. Thus, for a Bw = 5 MHz a throughput of 28 Mbps in DL was obtained, and for a Bw = 10 MHz, 69 Mbps was the average throughput. In the UL, regardless of the distance between the antennas, the throughput was practically identical, reaching about 9.6 Mbps for a 5 MHz bandwidth, and around 21.75 Mbps for a 10 MHz bandwidth.

Latency results could not be obtained in the simulator, but this was not the case for the laboratory environment. In this sense, the response at a ping test was measured in the laboratory at about 22.5 ms, whereas speedtest.ro registered a higher latency, of about 46 ms, which is caused by the increased distance between the laboratory where the measurements were performed and the location of the server of speedtest.ro.

Conclusion

In this thesis work, an effective study, analysis and evaluation of the LTE downlink and uplink performance with different MIMO techniques in comparison with the traditional SISO system have been carried out. The performance is evaluated with respect to two representative metrics, namely throughput and BLER, considering the use of different MIMO techniques (SISO, TxD 2×1, TxD 4×2 and OLSM 4×2) in two different channel models, namely flat fading Rayleigh channel and ITU Pedestrian B channel.

The case of no HARQ and a maximum of three retransmissions were considered in order to evaluate the change in performance for the above mentioned MIMO techniques. It can be concluded the number of retransmissions does not affect the maximum throughput obtainable through each techniques, the most notable observation being that for poor radio conditions, characterized by low values of the SNR, when using HARQ, the users experience a higher throughput than in the case when no retransmissions are employed. The performance of the rich multipath environment modeled by ITU Pedestrian B channel is better than the performance offered by the flat-fading Rayleigh channel, the same values of the BLER being obtained for SNR levels than are lower with about 5 dB for transmit diversity, SISO and spatial multiplexing respectively.

Analysis of the results obtained in the Matlab simulator reveal that the performance of MIMO, with both transmit diversity and OLSM - which is suitable for high velocity UEs, is better than SISO in both channel models. Spatial multiplexing is ideal for achieving very high peak rates, while transmit diversity is a valuable scheme to minimize the rate of bit error occurrence thereby improving signal quality. A set of measurements were performed in the laboratory for a channel bandwidth of 10 MHz and 20 MHz, which confirmed the results obtained through simulation.

The simulations performed in the LTE System Level simulator highlighted that UE speed has a major impact on the experienced throughput, an increase with about 50 km/h causing the throughput to decrease with about 50%. Moreover, different schedulers were considered, in order to emphasize that they are not equally fair and that different schedulers suit different user behaviors. Thus, the highest values for the fairness index were obtained by the RR and PF schedulers, whereas in terms of throughput, the best CQI scheduler yielded the greatest throughput, RR and PF schedulers obtaining much smaller, but similar throughputs.

The vision of LTE is therefore nothing less than an actual possibility and a true reality as this evaluation has demonstrated that the design goals and targets of LTE can be met with a high degree of reliability and certainty. This performance evaluation also provides useful information on LTE downlink planning, design and optimization for deployment.

References

[1] Farooq Khan. LTE for 4G Mobile Broadband – Air Interface Technologies and Performance. Cambridge University Press, first edition, 2009, pag. 1-2
[2] Ralf Kreher and Karsten Gaenger. LTE Signaling, Troubleshooting and Optimization. Wiley, second edition, 2011, pag. 1
[3] Farooq Khan. LTE for 4G Mobile Broadband – Air Interface Technologies and Performance. Cambridge University Press, first edition, 2009, pag 3
[4] https://sites.google.com/site/lteencyclopedia/home accessed on 25.05.2013
[5] Agilent. 3GPP Long Term Evolution: System Overview, Product Development and Test Challenges, page 4
[6] http://www.3gpp.org/LTE accessed on 25.05.2013
[7] Farooq Khan. LTE for 4G Mobile Broadband – Air Interface Technologies and Performance. Cambridge University Press, first edition, 2009, pag. 5
[8] Ralf Kreher and Karsten Gaenger. LTE Signaling, Troubleshooting and Optimization. Wiley, second edition, 2011, pag. 11-13
[9] Ralf Kreher and Karsten Gaenger. LTE Signaling, Troubleshooting and Optimization. Wiley, second edition, 2011, pag. 10
[10] Farooq Khan. LTE for 4G Mobile Broadband – Air Interface Technologies and Performance. Cambridge University Press, first edition, 2009, pag. 6
[11] Ralf Kreher and Karsten Gaenger. LTE Signaling, Troubleshooting and Optimization. Wiley, second edition, 2011, pag. 14
[12] Ralf Kreher and Karsten Gaenger. LTE Signaling, Troubleshooting and Optimization. Wiley, second edition, 2011, pag. 13-17
[13] http://www.artizanetworks.com/lte_tut_lay_2_str.html accessed on 27.05.2013
[14] Farooq Khan. LTE for 4G Mobile Broadband – Air Interface Technologies and Performance. Cambridge University Press, first edition, 2009, pag. 10
[15] http://www.artizanetworks.com/lte_tut_lay_2_str.html accessed on 27.05.2013
[16] Farooq Khan. LTE for 4G Mobile Broadband – Air Interface Technologies and Performance. Cambridge University Press, first edition, 2009, pag. 11-12
[17] http://www.eetimes.com/design/communications-design/4204835/An-overview-of-the-LTE-physical-layer-Part-III accessed on 27.05.2013
[18] http://www.artizanetworks.com/lte_tut_lay_2_str.html accessed on 27.05.2013
[19] A. R. S. Bahai, B. R. Saltzberg, M. Ergen, "Multi-carrier Digital Communications: Theory and Applications of OFDM", 2^{nd} Ed., Springer, New York, Oct. 2004
[20] G. L. Stüber et. al., "Broadband MIMO-OFDM Wireless Communications", Proc. of the IEEE, Vol. 92, No. 2, Feb. 2004
[21] http://www.radio-electronics.com/info/rf-technology-design/ofdm/ofdm-basics-tutorial.php accessed on 28.05.2013
[22] http://www.ixiacom.com/pdfs/library/white_papers/SC-FDMA-INDD.pdf accessed on 28.05.2013 accessed on 28.05.2013
[23] Agilent. 3GPP Long Term Evolution: System Overview, Product Development and Test Challenges, page 16
[24] http://www.sigcomm.org/sites/default/files/ccr/papers/2010/October/1880153-1880165.pdf accessed on 28.05.2013
[25] http://www.ixiacom.com/pdfs/library/white_papers/SC-FDMA-INDD.pdf accessed on 28.05.2013 accessed on 28.05.2013
[26] Hyung G. Myung, Junsung Lim, and David J. Goodman, "Single Carrier FDMA for Uplink Wireless Transmission." IEEE Vehicular Technology, Sept 2006

[27] http://mobiledevdesign.com/tutorials/WIreless_Everywhere_Not_Quite_Yet/index2.html accessed on 29.05.2013
[28] Agilent. 3GPP Long Term Evolution: System Overview, Product Development and Test Challenges, page 18
[29] MIMO Transmission Scheme for LTE and HSPA networks, June 2009, 3G Americas
[30] MIMO in LTE Operation and Measurement – Excerpts on LTE test. Agilent Technologies
[31] http://www.exploregate.com/Video.aspx?video_id=4#.UacNfUDQlTL accessed on 30.05.2013
[32] 3GPP TS 36.104 V9.13.0 (2012-09)
[33] http://www.zte.com.cn/cn/events/lteworld2012/solutions/201205/t20120516_364210.html accessed on 01.06.2013
[34] Nokia Siemens Networks. LTE 1800 MHz. Introducing LTE with maximum reuse of GSM assets
[35] 3GPP TS 36.306 V9.8.0 (2013-03)
[36] Amitabha Ghosh; Rapeepat Ratasuk; "Essentials of LTE and LTE-A", Cambridge University Press, 2011, pag. 19
[37] http://www.sharetechnote.com/html/FrameStructure_DL.html accessed on 03.06.2013
[38] 3GPP TS 36.211 V8.9.0 (2009-12)
[39] http://lteuniversity.com/get_trained/expert_opinion1/b/lauroortigoza/archive/2012/08/07/frame-structures-in-lte-tdd-and-lte-fdd.aspx accessed on 03.06.2013
[40] LTE in a Nutshell: The Physical Layer. Telesystem Innovations, 2010
[41] http://lteuniversity.blogspot.ro/2012/08/rrm-functions.html accessed on 04.06.2013
[42] Whitepaper “LTE E-UTRAN and its Access Side Protocols”. Radsys
[43] Dahlman, E. Parkvall, S. Sköld, J. Beming, P., 3G Evolution: HSPA and LTE for Mobile Broadband, Elsevier, Second edition, 2008
[44] Chamila Asanka Ariyaratne. Link Adaptation Improvements for Long Term Evolution. Blekinge Institute of Technology, Nov 2009
[45] Ralf Kreher and Karsten Gaenger. LTE Signaling, Troubleshooting and Optimization. Wiley, second edition, 2011, pag. 67
[46] http://www.exploregate.com/Video.aspx?video_id=110#.Uc10cDQweN1 accessed on 28.06.2013
[47] http://lte-epc.blogspot.ro/2009/12/lte-bearer-service-architecture.html accessed on 28.06.2013
[48] http://3gpp.wikispaces.com/What+is+the+EPS+Bearer%3F accessed on 28.06.2013
[49] Harri Holma, Antti Toskala. LTE for UMTS - OFDMA and SC-FDMA Based Radio Access. John Wiley & Sons, LTd., 2009, pag. 259-272
[50] http://www.exploregate.com/Video.aspx?video_id=58#.UbGZTfnQlTJ accessed on 07.06.2013
[51] Strategic white paper. Options for Providing Voice over LTE and Their Impact on the GSM/UMTS Network. Alcatel Lucent
[52] http://www.ecnmag.com/articles/2012/09/volte-what-makes-voice-over-ip-%E2%80%9Ccarrier-grade accessed on 09.06.2013
[53] Ralf Kreher and Karsten Gaenger. LTE Signaling, Troubleshooting and Optimization. Wiley, second edition, 2011, pag. 236
[54] Ralf Kreher and Karsten Gaenger. LTE Signaling, Troubleshooting and Optimization. Wiley, second edition, 2011, pag. 242-266
[55] Christian Mehlführer, Josep Colom Ikuno, Michal Šimko, Stefan Schwarz, Martin Wrulich and Markus Rupp. The Vienna LTE simulators - Enabling reproducibility in wireless communications research, SpringerOpen journal
[56] ITU Document, *Rec.ITU-R M.1225*, ITU-R, 1997