Chapter 1

LTE Introduction

1.1 Introduction

The challenge for any book tackling a subject as broad and deep as a completely new cellular radio standard is one of focus. The process of just creating the Long Term Evolution (LTE) specifications alone has taken several years and involved tens of thousands of temporary documents, thousands of hours of meetings, and hundreds of engineers. The result is several thousand pages of specifications. Now the hard work is underway, turning those specifications into real products that deliver real services to real people willing to pay real money. A single book of this length must therefore choose its subject wisely if it is to do more than just scratch the surface of such a complex problem.

The focus that Agilent has chosen for this book is a practical one: to explain design and measurement tools and techniques that engineering teams can use to accelerate turning the LTE specifications into a working system. The first half of the book provides an overview of the specifications starting in Chapter 2 with RF aspects and moving through the physical layer and upper layer signaling to the System Architecture Evolution (SAE) in Chapter 5. Due to limited space, the material in Chapters 2 through 5 should be viewed as an introduction to the technology rather than a deep exposition. For many, this level of detail will be sufficient but anyone tasked with designing or testing parts of the system will always need to refer directly to the specifications. The emphasis in the opening chapters is often on visual rather than mathematical explanations of the concepts. The latter can always be found in the specifications and should be considered sufficient information to build the system. However, the former approach of providing an alternative, more accessible explanation is often helpful prior to gaining a more detailed understanding directly from the specifications.

Having set the context for LTE in the opening chapters, the bulk of the remainder of the book provides a more detailed study of the extensive range of design and measurement techniques and tools that are available to help bring LTE from theory to deployment.

1.2 LTE System Overview

Before describing the LTE system it is useful to explain some of the terminology surrounding LTE since the history and naming of the technology is not intuitive. Some guidance can be found in the Vocabulary of 3GPP Specifications 21.905 [1], although this document is not comprehensive. The term LTE is actually a project name of the Third Generation Partnership Project (3GPP). The goal of the project, which started in November 2004, was to determine the long-term evolution of 3GPP's universal

LTE and the Evolution to 4G Wireless: Design and Measurement Challenges, Second Edition, Edited by Moray Rumney.

mobile telephone system (UMTS). UMTS was also a 3GPP project that studied several candidate technologies before choosing wideband code division multiple access (W-CDMA) for the radio access network (RAN). The terms UMTS and W-CDMA are now interchangeable, although that was not the case before the technology was selected.

In a similar way, the project name LTE is now inextricably linked with the underlying technology, which is described as an evolution of UMTS although LTE and UMTS actually have very little in common. The UMTS RAN has two major components: (1) the universal terrestrial radio access (UTRA), which is the air interface including the user equipment (UE) or mobile phone, and (2) the universal terrestrial radio access network (UTRAN), which includes the radio network controller (RNC) and the base station, which is also known as the node B (NB).

Because LTE is the evolution of UMTS, LTE's equivalent components are thus named evolved UTRA (E-UTRA) and evolved UTRAN (E-UTRAN). These are the formal terms used to describe the RAN. The system, however, is more than just the RAN since there is also the parallel 3GPP project called System Architecture Evolution that is defining a new all internet protocol (IP) packet-only core network known as the evolved packet core (EPC). The combination of the EPC and the evolved RAN (E-UTRA plus E-UTRAN) is the evolved packet system (EPS). Depending on the context, any of the terms LTE, E-UTRA, E-UTRAN, SAE, EPC, and EPS may get used to describe some or all of the system. Although EPS is the only correct term for the overall system, the name of the system will often be written as LTE/SAE or even simply LTE, as in the title of this book.

Figure 1.2-1 shows a high level view of how the evolved RAN and EPC interact with legacy radio access technologies.

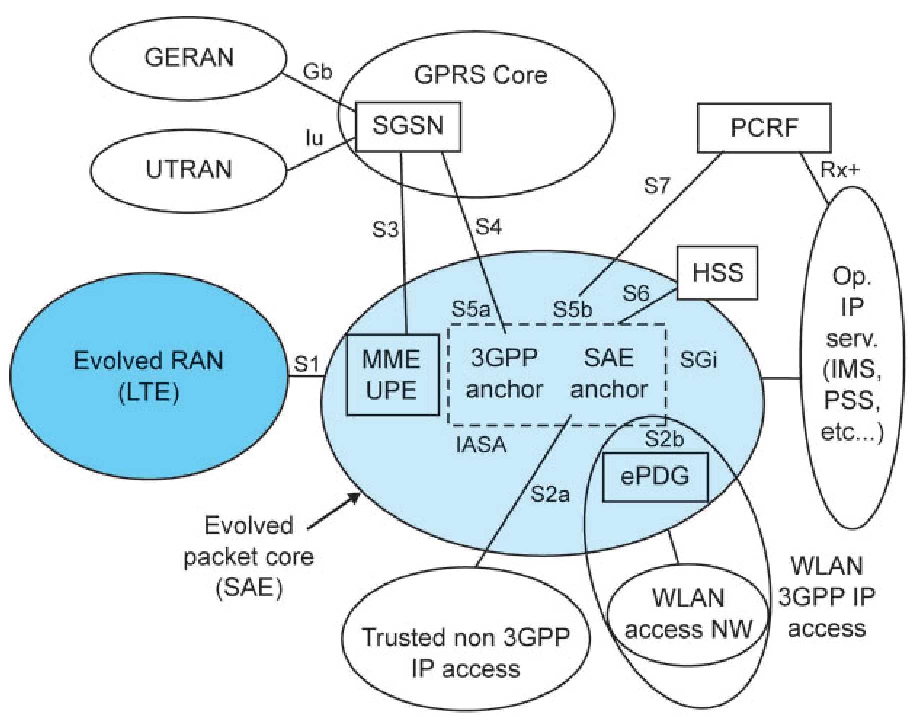

Figure 1.2-1. Logical high-level architecture for the evolved system (from 23.882 [2] Figure 4.2-1)

The 3GPP drive to simplify the existing hybrid circuit-switched/packet-switched core network is behind the SAE project to define an all-IP core network. This new architecture is a flatter, packet-only core network that is an essential part of delivering the higher throughput, lower cost, and lower latency that is the goal of the LTE evolved RAN. The EPC is also designed to provide seamless interworking with existing 3GPP and non-3GPP radio access technologies. The overall requirements for the System Architecture Evolution are summarized in 22.278 [3]. A more detailed description of the EPC is given in Chapter 5.

1.3 The Evolution from UMTS to LTE

The LTE specifications are written by 3GPP, which is a partnership of standards development organizations (SDOs). The work of 3GPP is public and, as will be described in Section 1.6, it is possible to gain access to all meeting reports, working documents, and published specifications from the 3GPP website: www.3gpp.org. The organizational partners that make up 3GPP are the Japanese Association of Radio Industries and Businesses (ARIB), the USA Alliance for Telecommunications Industry Solutions (ATIS), the China Communications Standards Association (CCSA), the European Telecommunications Standards Institute (ETSI), the Korean Telecommunications Technology Association (TTA), and the Japanese Telecommunications Technology Committee (TTC).

Table 1.3-1. Evolution of the UMTS specifications

Release	Functional freeze	Main UMTS feature of release
Rel-99	March 2000	Basic 3.84 Mcps W-CDMA (FDD & TDD)
Rel-4	March 2001	1.28 Mcps TDD (TD-SCDMA)
Rel-5	June 2002	HSDPA
Rel-6	March 2005	HSUPA (E-DCH)
Rel-7	Dec 2007	HSPA+ (64QAM downlink, MIMO, 16QAM uplink) LTE and SAE feasibility study
Rel-8	Dec 2008	LTE work item—OFDMA/SC-FDMA air interface, SAE work item—new IP core network, Dual-carrier HSDPA
Rel-9	December 2009	Home BS, MBMS, multi-standard radio, dual-carrier HSUPA, dual-carrier HSDPA with MIMO, dual-cell HSDPA
Rel-10	March 2011 (protocols 3 months later)	LTE-Advanced (carrier aggregation, 8x DL MIMO, 4x UL MIMO, relaying, enhanced inter-cell interference coordination (eICIC)), 4-carrier HSDPA
Rel-11	September 2012 (protocols 3 months later)	Further eICIC, coordinated multi-point transmission (CoMP), carrier aggregation scenarios, 8-carrier HSDPA
Rel-12	TBD—2014? (Stage 1 March 2013)	Further interference coordination, inter-site carrier aggregation, others TBD including dynamic TDD and LTE-D

Table 1.3-1 summarizes the evolution of the 3GPP UMTS specifications towards LTE. Each release of the 3GPP specifications represents a defined set of features. A summary of the contents of any release can be found at www.3gpp.org/releases.

The date given for the functional freeze relates to the date when no further new items can be added to the release. After this point any further changes to the specifications are restricted to essential corrections. The commercial launch date of a release depends on the period of time following the functional freeze before the specifications are considered stable and then implemented into commercial systems. For the first release of UMTS the delay between functional freeze and commercial launch was several years, although the delay for subsequent releases was progressively shorter. The delay between functional freeze and the first commercial launch for LTE/SAE was remarkably short, being less than a year, although it was two years before significant numbers of networks started operation. This period included the time taken to develop and implement the conformance test cases, which required significant work that could not begin until the feature set of the release was frozen and UEs had been implemented.

After Release 99, 3GPP stopped naming releases with the year and opted for a new scheme starting with Release 4. This choice was driven by the document version numbering scheme explained in Section 1.6. Release 4 introduced the 1.28 Mcps narrow band version of W-CDMA, also known as time division synchronous code division multiple access (TD-SCDMA). Following this was Release 5, in which high speed downlink packet access (HSDPA) introduced packet-based data services to UMTS in the same way that the general packet radio service (GPRS) did for GSM in Release 97 (1998). The completion of packet data for UMTS was achieved in Release 6 with the addition of high speed uplink packet access (HSUPA), although the official term for this technology is enhanced dedicated channel (E-DCH). HSDPA and HSUPA are now known collectively as high speed packet access (HSPA). Release 7 contained the first work on LTE/SAE with the completion of feasibility studies, and further improvements were made to HSPA such as downlink multiple input-multiple output (MIMO), 64QAM on the downlink, and 16QAM on the uplink. In Release 8, HSPA continued to evolve with the addition of numerous smaller features such as dual-carrier HSDPA and 64QAM with MIMO. Dual-carrier HSUPA was introduced in Release 9, four-carrier HSDPA in Release 10, and eight-carrier HSDPA in Release 11.

The main work in Release 8 was the specification of LTE and SAE, which is the main focus of this book. Work beyond Release 8 up to Release 12 is summarized in Chapter 8, although there are many references to features from these later releases throughout this second edition. Within 3GPP there are additional standardization activities not shown in Table 1.3-1 such as those for the GSM enhanced RAN (GERAN) and the IP multimedia subsystem (IMS).

1.4 LTE/SAE Requirements

The high level requirements for LTE/SAE include reduced cost per bit, better service provisioning, flexible use of new and existing frequency bands, simplified network architecture with open interfaces, and an allowance for reasonable power consumption by terminals. These are detailed in the LTE feasibility study 25.912 [4] and in the LTE requirements document 25.913 [5]. To meet the requirements for LTE outlined in 25.913 [5], LTE/SAE has been specified to achieve the following:

- Increased downlink and uplink peak data rates, as shown in Table 1.4-1. Note that the downlink is specified for single input single output (SISO) and MIMO antenna configurations at a fixed 64QAM modulation depth, whereas the uplink is specified only for SISO but at different modulation depths. These figures represent the physical limitation of the FDD air interface in ideal radio conditions with allowance for signaling overheads. Lower peak rates are specified for specific UE categories, and performance requirements under non-ideal radio conditions have also been developed. Comparable figures exist in [4] for TDD operation.

- Scalable channel bandwidths of 1.4 MHz, 3.0 MHz, 5 MHz, 10 MHz, 15 MHz, and 20 MHz in both the uplink and the downlink.
- Spectral efficiency improvements over Release 6 HSPA of 3 to 4 times in the downlink and 2 to 3 times in the uplink.
- Sub-5 ms latency for small IP packets.
- Performance optimized for low mobile speeds from 0 to 15 km/h supported with high performance from 15 to 120 km/h; functional support from 120 to 350 km/h. Support for 350 to 500 km/h is under consideration.
- Co-existence with legacy standards while evolving toward an all-IP network.

Table 1.4-1. LTE (FDD) downlink and uplink peak data rates (from 25.912 [4] Tables 13.1 & 13.1a)

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

The headline data rates in Table 1.4-1 represent the corner case of what can be achieved with the LTE RAN in perfect radio conditions; however, it is necessary for practical reasons to introduce lower levels of performance to enable a range of implementation choices for system deployment. This is achieved through the introduction of UE categories as specified in 36.306 [6] and shown in Table 1.4-2. These are similar in concept to the categories used to specify different levels of performance for HSPA.

Table 1.4-2. Peak data rates for UE categories (derived from 36.306 [6] Tables 4.1-1 and 4.1-2)

UE category	Peak downlink data rate (Mbps)	Number of downlink spatial layers	Peak uplink data rate (Mbps)	Number of uplink spatial layers	Support for 64QAM in uplink
Category 1	10.296	1	5.16	1	No
Category 2	51.024	2	25.456	1	No
Category 3	102.048	2	51.024	1	No
Category 4	150.752	2	51.024	1	No
Category 5	302.752	4	75.376	1	Yes
Category 6	301.504	2 or 4	51.024	1, 2, or 4	No
Category 7	301.504	2 or 4	10.2048	1, 2, or 4	No
Category 8	2998.56	8	149.776	8	Yes

Categories 6, 7, and 8 were added in Release 10 for the support of LTE-Advanced (see Section 8.3). There are other attributes associated with UE categories, but the peak data rates, downlink antenna configuration, and uplink 64QAM support are the categories most commonly referenced.

The emphasis so far has been on the peak data rates but what really matters for the performance of a new system is the improvement that can be achieved in average and cell-edge data rates. The reference configuration against which LTE/SAE performance targets have been set is defined in 25.913 [5] as being Release 6 UMTS. For the downlink the reference is HSDPA Type 1 (receive diversity but no equalizer or interference cancellation). For the uplink the reference configuration is single transmitter with diversity reception at the Node B. Table 1.4-3 shows the simulated downlink performance of UMTS versus the design targets for LTE. This is taken from the work of 3GPP during the LTE feasibility study [7]. Table 1.4-4 shows a similar set of results for the uplink taken from [8].

Table 1.4-3. Comparison of UMTS Release 6 and LTE downlink performance requirements

Case 1 500m inter-site distance	Spectrum efficiency		Mean user throughput		Cell-edge user throughput	
	(bps/Hz/cell)	x UTRA	[bps/Hz/user]	x UTRA	(bps/Hz/user)	x UTRA
UTRA baseline 1x2	0.53	x1.0	0.05	x1.0	0.02	x1.0
E-UTRA 2x2 SU-MIMO	1.69	x3.2	0.17	x3.2	0.05	x2.7
E-UTRA 4x2 SU-MIMO	1.87	x3.5	0.19	x3.5	0.06	x3.0
E-UTRA 4x4 SU-MIMO	2.67	x5.0	0.27	x5.0	0.08	x4.4

Table 1.4-4. Comparison of UMTS Release 6 and LTE uplink performance requirements

Case 1 500m inter-site distance	Spectrum efficiency		Mean user throughput		Cell-edge user throughput	
	(bps/Hz/cell)	x UTRA	(bps/Hz/user)	x UTRA	[bps/Hz/user]	x UTRA
UTRA baseline	0.332	x1.0	0.033	x1.0	0.009	x1.0
E-UTRA 1x2	0.735	x2.2	0.073	x2.2	0.024	x2.5
E-UTRA 1x2 MU-MIMO	0.675	x2.0	0.067	x2.0	0.023	x2.4
E-UTRA 1x4	1.103	x3.3	0.110	x3.3	0.052	x5.5
E-UTRA 2x2 SU-MIMO	0.776	x2.3	0.078	x2.3	0.010	x1.1

From these tables the LTE design targets of 2x to 4x improvement over UMTS Release 6 can be seen. Note, however, that UMTS did not stand still and there were Release 7 and Release 8 UMTS enhancements that significantly narrow the gap between UMTS and LTE. The evolution of UMTS continues through Release 12. Although the figures in Tables 1.4-3 and 1.4-4 are meaningful and user-centric, they were derived from system-level simulations and are not typical of the methods used to specify minimum performance. The simulations involved calculation of throughput by repeatedly dropping ten users randomly into the cell. From this data a distribution of performance was developed and the mean user throughput calculated. The cell-edge throughput was defined as the 5th percentile of the throughput cumulative distribution. For this reason the cell-edge figures are quoted per user assuming 10 users per cell, whereas the mean user throughput is independent of the number of users.

When it comes to defining minimum performance requirements for individual UE, the simulation methods used to derive the figures in Tables 1.4-3 and 1.4-4 cannot be used. Instead, the minimum requirements for UMTS and LTE involve spot measurement of throughput at specific high and low interference conditions, and for additional simplicity, this is done without the use of closed-loop adaptive modulation and coding. This approach to defining performance is pragmatic but it means that there is no direct correlation between the results from the conformance tests and the simulated system performance in Tables 1.4-3 and 1.4-4.

1.5 LTE/SAE Timeline

The timeline of LTE/SAE development is shown in Figure 1.5-1. This includes the work of 3GPP in drafting the specifications as well as the conformance test activities of the Global Certification Forum (GCF) and the trials carried out by the LTE/SAE Trial Initiative (LSTI). The work of GCF towards the certification of UE against the 3GPP conformance specifications is covered in some detail in Section 7.4. The LSTI, whose work was completed in 2011, was an industry forum and complimentary group that worked in parallel with 3GPP and GCF with the intent of accelerating the acceptance and deployment of LTE/SAE as the logical choice of the industry for next-generation networks. The work of LSTI was split into four phases. The first phase was proof of concept of the basic principles of LTE/SAE, using early prototypes not necessarily compliant with the specifications. The second phase was interoperability development testing (IODT), which was a more detailed phase of testing using standards-compliant equipment but not necessarily commercial platforms. The third stage was interoperability testing (IOT), similar in scope to IODT but using platforms intended for commercial deployment. The final phase was Friendly Customer Trials, which ran through 2010. GCF certified the first UE against the 3GPP conformance tests in April 2011. By November 2012 there were 102 FDD and 11 TDD commercial networks launched in 51 countries according to the Global Suppliers Association.

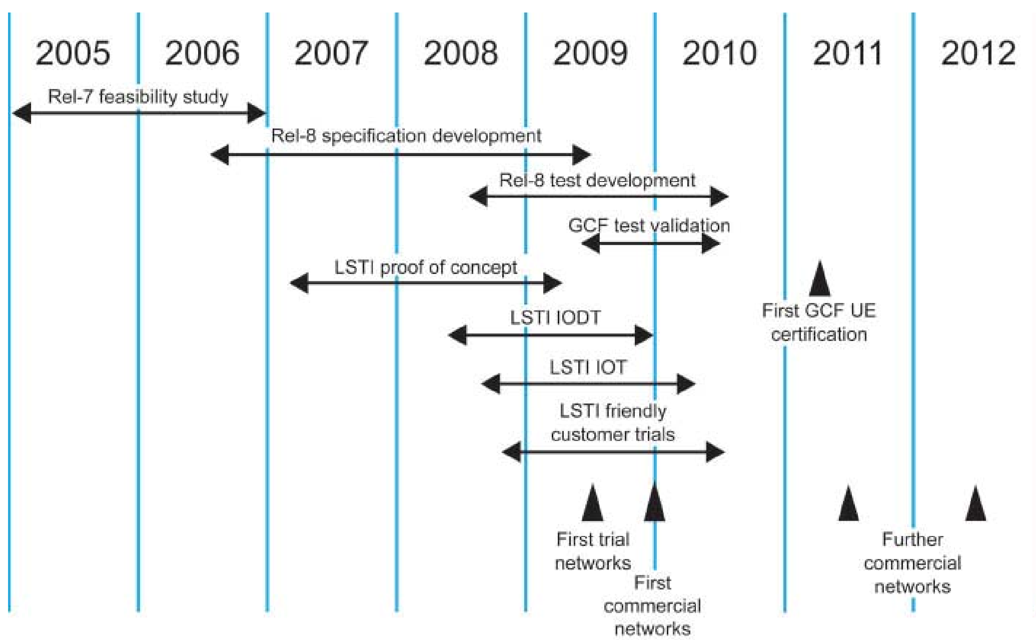

Figure 1.5-1. Projected LTE/SAE timeline

2

System Architecture Evolution

This chapter covers the high-level architecture of LTE. We begin by describing the hardware components in an LTE network and by reviewing the software protocols that those components use to communicate. We then look in more detail at the techniques used for data transport in LTE, before discussing the state diagrams and the use of radio spectrum. We will leave some more specialized architectural issues until later chapters, notably those related to quality of service, charging and inter-system operation.

Several specifications are relevant to this chapter. TS 23.401 [1] and TS 36.300 [2] are stage 2 specifications that include descriptions of the system architecture, while the relevant stage 3 specifications [3, 4] contain the architectural details. We will also note some other important specifications as we go along.

2.1 Architecture of LTE

2.1.1 High Level Architecture

Figure 2.1 reviews the high-level architecture of the evolved packet system (EPS). There are three main components, namely the user equipment (UE), the evolved UMTS terrestrial radio access network (E-UTRAN) and the evolved packet core (EPC). In turn, the evolved packet core communicates with packet data networks in the outside world such as the internet, private corporate networks or the IP multimedia subsystem. The interfaces between the different parts of the system are denoted Uu, S1 and SGi.

The UE, E-UTRAN and EPC each have their own internal architectures and we will now discuss these one by one.

2.1.2 User Equipment

Figure 2.2 shows the internal architecture of the user equipment [5]. The architecture is identical to the one used by UMTS and GSM.

The actual communication device is known as the *mobile equipment* (ME). In the case of a voice mobile or a smartphone, this is just a single device. However, the mobile

An Introduction to LTE: LTE, LTE-Advanced, SAE and 4G Mobile Communications, First Edition. Christopher Cox.

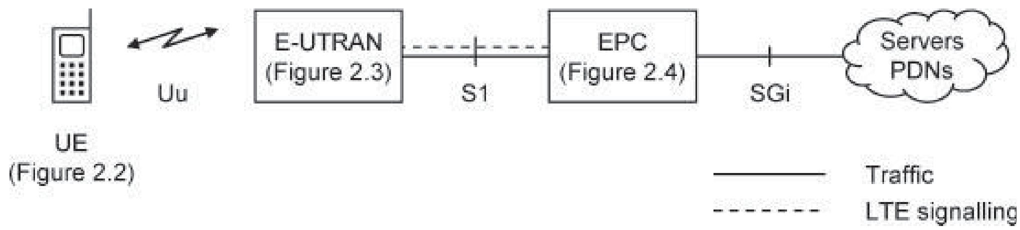

Figure 2.1 High level architecture of LTE.

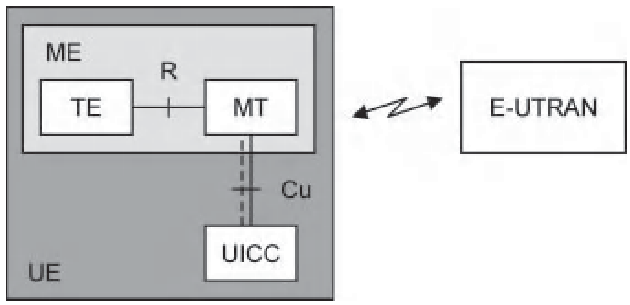

Figure 2.2 Internal architecture of the UE. Reproduced by permission of ETSI.

equipment can also be divided into two components, namely the *mobile termination* (MT), which handles all the communication functions, and the *terminal equipment* (TE), which terminates the data streams. The mobile termination might be a plug-in LTE card for a laptop, for example, in which case the terminal equipment would be the laptop itself.

The *universal integrated circuit card* (UICC) is a smart card, colloquially known as the SIM card. It runs an application known as the *universal subscriber identity module* (USIM) [6], which stores user-specific data such as the user's phone number and home network identity. The USIM also carries out various security-related calculations, using secure keys that the smart card stores. LTE supports mobiles that are using a USIM from Release 99 or later, but it does not support the *subscriber identity module* (SIM) that was used by earlier releases of GSM.

In addition, LTE supports mobiles that are using IP version 4 (IPv4), IP version 6 (IPv6), or dual stack IP version 4/version 6. A mobile receives one IP address for every packet data network that it is communicating with, for example one for the internet and one for any private corporate network. Alternatively, the mobile can receive an IPv4 address as well as an IPv6 address, if the mobile and network both support the two versions of the protocol.

Mobiles can have a wide variety of radio capabilities [7], which cover issues such as the maximum data rate that they can handle, the different types of radio access technology that they support and the carrier frequencies on which they can transmit and receive. Mobiles pass these capabilities to the radio access network by means of signalling messages, so that the E-UTRAN knows how to control them correctly. The most important capabilities are grouped together into the *UE category*. As shown in Table 2.1, the UE category mainly covers the maximum data rate with which the mobile can transmit and receive. It also covers some technical issues that are listed in the last three columns of the table, which we will cover in Chapters 3 and 5.

Table 2.1 UE categories. Reproduced by permission of ETSI

UE category	Release	Maximum # DL bits per ms	Maximum # UL bits per ms	Maximum # DL layers	Maximum # UL layers	Support of UL 64-QAM?
1	R8	10 296	5 160	1	1	No
2	R8	51 024	25 456	2	1	No
3	R8	102 048	51 024	2	1	No
4	R8	150 752	51 024	2	1	No
5	R8	299 552	75 376	4	1	Yes
6	R10	301 504	51 024	4	1	No
7	R10	301 504	102 048	4	2	No
8	R10	2 998 560	1 497 760	8	4	Yes

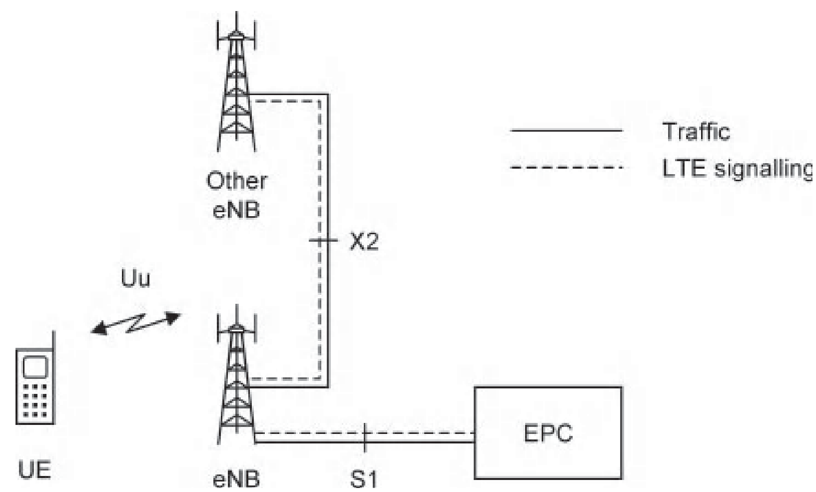

Figure 2.3 Architecture of the evolved UMTS terrestrial radio access network.

2.1.3 Evolved UMTS Terrestrial Radio Access Network

The evolved UMTS terrestrial radio access network (E-UTRAN) [8] is illustrated in Figure 2.3. The E-UTRAN handles the radio communications between the mobile and the evolved packet core and just has one component, the *evolved Node B* (eNB).

Each eNB is a base station that controls the mobiles in one or more cells. A mobile communicates with just one base station and one cell at a time, so there is no equivalent of the soft handover state from UMTS. The base station that is communicating with a mobile is known as its *serving eNB*.

The eNB has two main functions. Firstly, the eNB sends radio transmissions to all its mobiles on the downlink and receives transmissions from them on the uplink, using the analogue and digital signal processing functions of the LTE air interface. Secondly, the eNB controls the low-level operation of all its mobiles, by sending them signalling messages such as handover commands that relate to those radio transmissions. In carrying out these functions, the eNB combines the earlier functions of the Node B and the radio network controller, to reduce the latency that arises when the mobile exchanges information with the network.

Each base station is connected to the EPC by means of the S1 interface. It can also be connected to nearby base stations by the X2 interface, which is mainly used for signalling and packet forwarding during handover. The X2 interface is optional, in that the S1 interface can also handle all the functions of X2, albeit indirectly and more slowly. Usually, the S1 and X2 interfaces are not direct physical connections: instead, the information is routed across an underlying IP based transport network in the manner shown in Figure 1.4. The same issue will apply to the EPC's interfaces below.

A *home eNB* (HeNB) is a base station that has been purchased by a user to provide femtocell coverage within the home [9]. A home eNB belongs to a *closed subscriber group* (CSG) and can only be accessed by mobiles with a USIM that also belongs to the closed subscriber group. From an architectural point of view, a home eNB can be connected directly to the evolved packet core in the same way as any other base station, or can be connected by way of an intermediate device known as a *home eNB gateway* that collects the information from several home eNBs. Home eNBs only control one cell, and do not support the X2 interface until Release 10.

2.1.4 Evolved Packet Core

Figure 2.4 shows the main components of the evolved packet core [10, 11]. We have already seen one component, the home subscriber server (HSS), which is a central database that contains information about all the network operator's subscribers. This is one of the few components of LTE that has been carried forward from UMTS and GSM.

The *packet data network* (PDN) *gateway* (P-GW) is the EPC's point of contact with the outside world. Through the SGi interface, each PDN gateway exchanges data with one or more external devices or packet data networks, such as the network operator's servers, the internet or the IP multimedia subsystem. Each packet data network is identified by an *access point name* (APN) [12]. A network operator typically uses a handful of different APNs, for example one for its own servers and one for the internet.

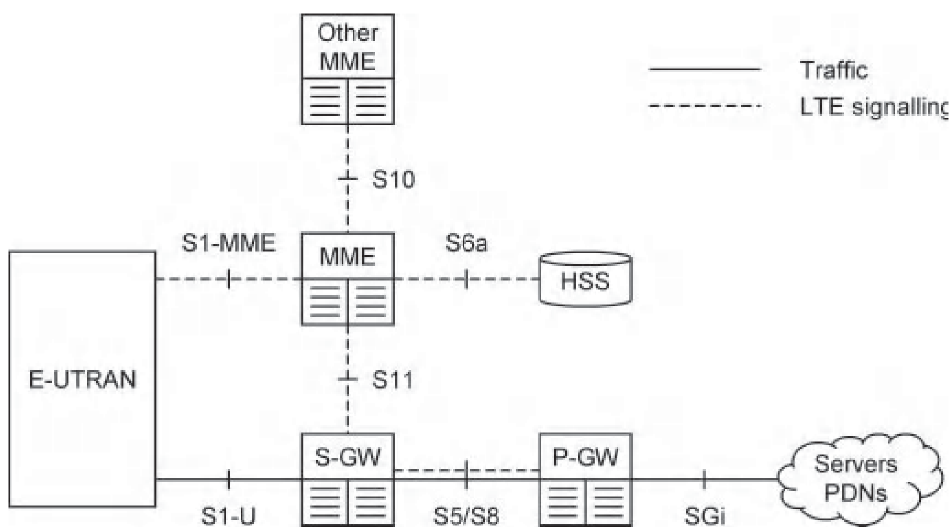

Figure 2.4 Main components of the evolved packet core.

Each mobile is assigned to a default PDN gateway when it first switches on, to give it always-on connectivity to a default packet data network such as the internet. Later on, a mobile may be assigned to one or more additional PDN gateways, if it wishes to connect to additional packet data networks such as private corporate networks. Each PDN gateway stays the same throughout the lifetime of the data connection.

The *serving gateway* (S-GW) acts as a router, and forwards data between the base station and the PDN gateway. A typical network might contain a handful of serving gateways, each of which looks after the mobiles in a certain geographical region. Each mobile is assigned to a single serving gateway, but the serving gateway can be changed if the mobile moves sufficiently far.

The *mobility management entity* (MME) controls the high-level operation of the mobile, by sending it signalling messages about issues such as security and the management of data streams that are unrelated to radio communications. As with the serving gateway, a typical network might contain a handful of MMEs, each of which looks after a certain geographical region. Each mobile is assigned to a single MME, which is known as its *serving MME*, but that can be changed if the mobile moves sufficiently far. The MME also controls the other elements of the network, by means of signalling messages that are internal to the EPC.

Comparison with UMTS and GSM shows that the PDN gateway has the same role as the gateway GPRS support node (GGSN), while the serving gateway and MME handle the data routing and signalling functions of the serving GPRS support node (SGSN). Splitting the SGSN in two makes it easier for an operator to scale the network in response to an increased load: the operator can add more serving gateways as the traffic increases, while adding more MMEs to handle an increase in the number of mobiles. To support this split, the S1 interface has two components: the S1-U interface carries traffic for the serving gateway, while the S1-MME interface carries signalling messages for the MME.

The EPC has some other components that were not shown in Figure 2.4. Firstly, the *cell broadcast centre* (CBC) was previously used by UMTS for the rarely implemented *cell broadcast service* (CBS). In LTE, the equipment is re-used for a service known as the *earthquake and tsunami warning system* (ETWS) [13]. Secondly, the *equipment identity register* (EIR) was also inherited from UMTS, and lists the details of lost or stolen mobiles. We will introduce further components later in the book, when we consider the management of quality of service, and the inter-operation between LTE and other mobile communication systems.

2.1.5 Roaming Architecture

Roaming allows users to move outside their network operators' coverage area by using the resources from two different networks. It relies on the existence of a *roaming agreement*, which defines how the operators will share the resulting revenue. There are two possible architectures [14], which are shown in Figure 2.5.

If a user is roaming, then the home subscriber server is always in the home network, while the mobile, E-UTRAN, MME and serving gateway are always in the visited network. The PDN gateway, however, can be in two places. In the usual situation of *home routed traffic*, the PDN gateway lies in the home network, through which all the user's traffic is all routed. This architecture allows the home network operator to see all the traffic

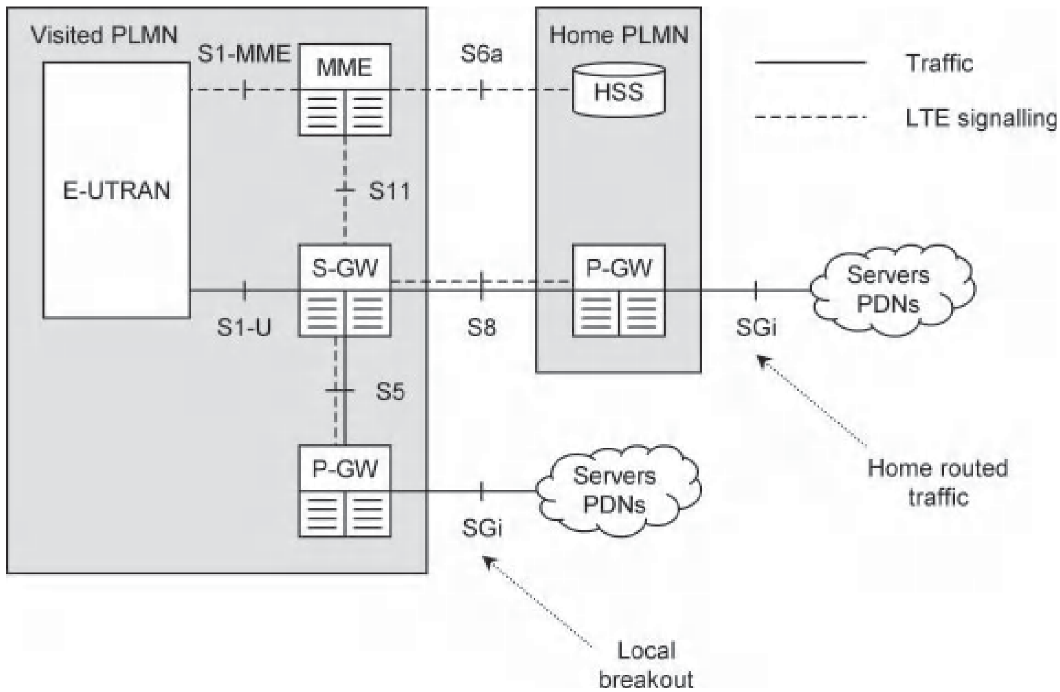

Figure 2.5 Architecture of LTE for a roaming mobile.

and to charge the user for it directly, but can be inefficient if the user is travelling overseas, particularly during a voice call with another user nearby. To deal with this situation, the specifications also support *local breakout*, in which the PDN gateway is located in the visited network. The HSS indicates whether or not the home network will permit local breakout, for each combination of user and APN [15].

The interface between the serving and PDN gateways is known as S5/S8. This has two slightly different implementations, namely S5 if the two devices are in the same network, and S8 if they are in different networks. For mobiles that are not roaming, the serving and PDN gateways can be integrated into a single device, so that the S5/S8 interface vanishes altogether. This can be useful because of the associated reduction in latency.

2.1.6 Network Areas

The EPC is divided into three different types of geographical area [16], which are illustrated in Figure 2.6.

An *MME pool area* is an area through which the mobile can move without a change of serving MME. Every pool area is controlled by one or more MMEs, while every base station is connected to all the MMEs in a pool area by means of the S1-MME interface. Pool areas can also overlap. Typically, a network operator might configure a pool area to cover a large region of the network such as a major city and might add MMEs to the pool as the signalling load in that city increases.

Similarly, an *S-GW service area* is an area served by one or more serving gateways, through which the mobile can move without a change of serving gateway. Every base

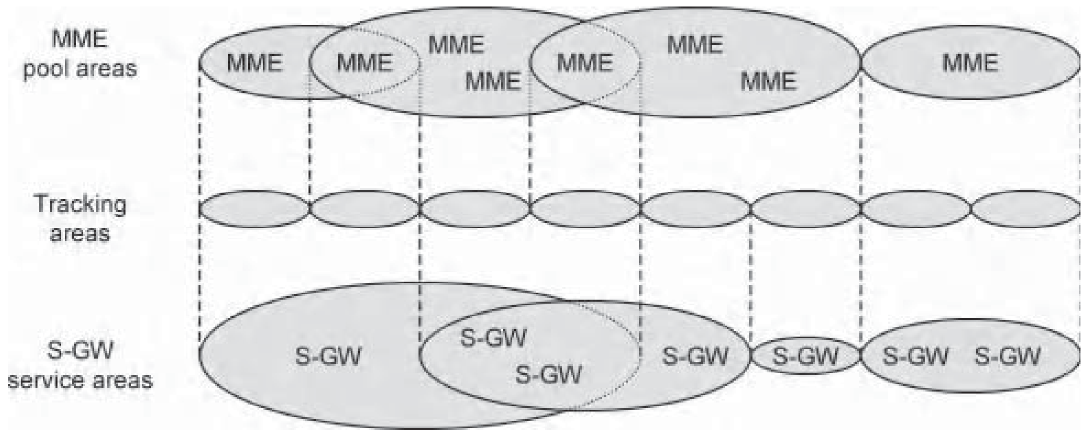

Figure 2.6 Relationship between tracking areas, MME pool areas and S-GW service areas.

station is connected to all the serving gateways in a service area by means of the S1-U interface. S-GW service areas do not necessarily correspond to MME pool areas.

MME pool areas and S-GW service areas are both made from smaller, non-overlapping units known as *tracking areas* (TAs). These are used to track the locations of mobiles that are on standby and are similar to the location and routing areas from UMTS and GSM.

2.1.7 Numbering, Addressing and Identification

The components of the network are associated with several different identities [17]. As in previous systems, each network is associated with a *public land mobile network identity* (PLMN-ID). This comprises a three digit *mobile country code* (MCC) and a two or three digit *mobile network code* (MNC). For example, the mobile country code for the UK is 234, while Vodafone's UK network uses a mobile network code of 15.

Each MME has three main identities, which are shown as the shaded parts of Figure 2.7. The 8 bit *MME code* (MMEC) uniquely identifies the MME within all the pool areas that it belongs to. By combining this with a 16 bit *MME group identity* (MMEGI), we arrive at a 24 bit *MME identifier* (MMEI), which uniquely identifies the MME within a particular

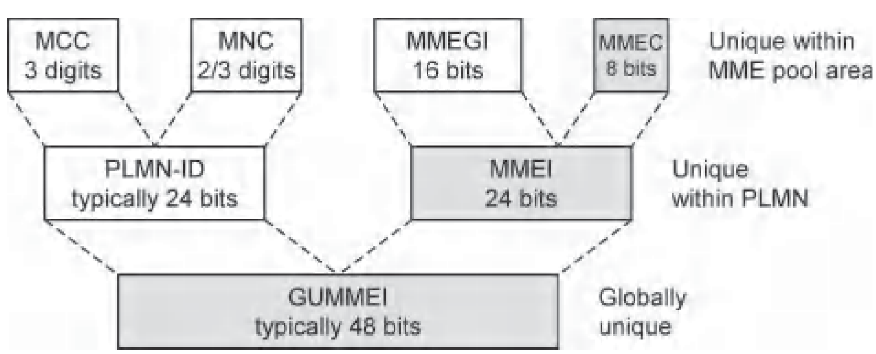

Figure 2.7 Identities used by the MME.

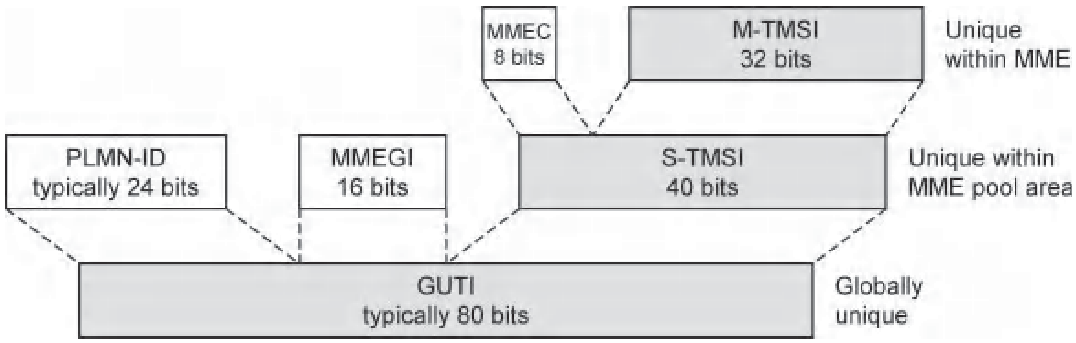

Figure 2.8 Temporary identities used by the mobile.

network. By bringing in the network identity, we arrive at the *globally unique MME identifier* (GUMMEI), which identifies an MME anywhere in the world.

Similarly, each tracking area has two main identities. The 16 bit *tracking area code* (TAC) identifies a tracking area within a particular network. Combining this with the network identity gives the globally unique *tracking area identity* (TAI).

Cells have three types of identity. The 28 bit *E-UTRAN cell identity* (ECI) identifies a cell within a particular network, while the *E-UTRAN cell global identifier* (ECGI) identifies a cell anywhere in the world. Also important for the air interface is the *physical cell identity*, which is a number from 0 to 503 that distinguishes a cell from its immediate neighbours.

A mobile is also associated with several different identities. The most important are the *international mobile equipment identity* (IMEI), which is a unique identity for the mobile equipment, and the *international mobile subscriber identity* (IMSI), which is a unique identity for the UICC and the USIM.

The IMSI is one of the quantities that an intruder needs to clone a mobile, so we avoid transmitting it across the air interface wherever possible. Instead, a serving MME identifies each mobile using temporary identities, which it updates at regular intervals. Three types of temporary identity are important, and are shown as the shaded parts of Figure 2.8. The 32 bit *M temporary mobile subscriber identity* (M-TMSI) identifies a mobile to its serving MME. Adding the MME code results in the 40 bit *S temporary mobile subscriber identity* (S-TMSI), which identifies the mobile within an MME pool area. Finally, adding the MME group identity and the PLMN identity results in the most important quantity, the *globally unique temporary identity* (GUTI).

2.2 Communication Protocols

2.2.1 Protocol Model

Each of the interfaces from the previous section is associated with a protocol stack, which the network elements use to exchange data and signalling messages. Figure 2.9 shows the high-level structure of those protocol stacks.

The protocol stack has two planes. Protocols in the *user plane* handle data that are of interest to the user, while protocols in the *control plane* handle signalling messages that

2

LTE Overview

Cinzia Sartori, Anssi Juppi, Henning Sanneck, Seppo Hämäläinen
and Miikka Poikselkä

LTE encompasses a set of aggressive requirements that aim at improving the end-user throughput, the cell capacity and reducing the user plane latency. These requirements, together with full mobility, will bring substantial benefits to user experience.

LTE is designed to support all kind of IP data traffic and voice is supported as Voice over IP (VoIP) for better integration with multimedia services. LTE aggressive requirements lead to the definition of a new Network Architecture, the Evolved Packet System (EPS), which comprises the Enhanced RAN (E-UTRAN or LTE) and the Evolved Packet Core (EPC). Both data and voice services are supported over the same packet switched network. E-UTRAN and EPC have been defined in 3GPP Release 8 and enhanced in further 3GPP Releases.

LTE paved the way to a new standardisation approach. In Release 8 LTE network and OAM have been standardised at the same time, yielding tremendous opportunities to design an overall optimised system with built in SON features.

The scope of this chapter is to give first a short introduction, without digging into technical details of the EPS network architecture, both E-UTRAN and EPC (deep inside is addressed in LTE technology specific books, for example, (Holma and Toskala, 2011) for LTE Radio Access Network). The chapter continues with a brief description of LTE-Advanced (LTE-A is specified in 3GPP Release 10), while the last part of the chapter is dedicated to LTE Radio Access Network Scenarios, their evolution and potential SON component.

2.1 Introduction to LTE and SAE

Long-Term Evolution (LTE) is a 3GPP project that provides extensions and modifications of the UMTS system allowing high data rate, low latency and packet optimised radio access networks. System Architecture Evolution (SAE) is an associated 3GPP project working on

LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency, First Edition.
Edited by Seppo Hämäläinen, Henning Sanneck and Cinzia Sartori.

3GPP core network evolution. The new air interface and network architecture aim at providing decreased cost per transmitted bit, achieved by:

- Advanced modulation techniques that allow optimised use of radio frequency.
- Flat architecture that minimises the number of network elements and optimises the usage of the transmission network.
- Capability to serve high quality, low latency real-time traffic, allowing both voice and data services to be provided over a single all-IP network.

The present chapter gives an overview of 3GPP as well as LTE Requirements and Specifications.

2.1.1 3GPP Structure, Timeline and LTE Specifications

The LTE standard, as well as WCDMA and the latest phase of GSM Evolution, have been developed by the 3rd Generation Partnership Project (3GPP). 3GPP is a collaborative standardisation model, uniting telecommunications standards bodies. It was formed by the European Telecommunications Standards Institute (ETSI), which defined the successful GSM standard, together with its counterparts on the other continents to be a global partnership which today includes more than 300 individual member companies worldwide.

3GPP standardisation covers specification work for GSM based 2G and WCDMA based 3G radio technologies in addition to 4G LTE technology. Further, 3GPP does standardisation for services and system aspects and core networks. To cope with such a huge dimension 3GPP has defined a very structured working procedure; each Technical Specification Group (TSG) is further sub-divided in Working Groups (WGs), each of those covering a specific aspect. For example RAN TSG is split into RAN Radio layer, Radio Layers 2 and 3, Radio performance and protocol aspects, and Mobile terminal conformance testing related working groups, as shown in Figure 2.1.

LTE requirements were defined in the first half of 2005 and they have been the basis for the LTE Study Item (SI). 3GPP Study Items are feasibility studies that are carried out for bigger topics before actual standardisation starts in Work Item (WI) phase. The focus of LTE SI was defining of the new LTE radio access technology in terms of both multiple access and system architecture which can satisfy such requirements. The LTE Study Item was formally closed in September 2006 after which LTE work item was started. The first LTE specifications were contained as a part of the 3GPP Release 8.

LTE was further enhanced in Release 9 and Release 10. The recent Release 10 provides a big step forward with LTE-Advanced features which significantly improves the user data rate, the coverage extension as well as reduces latency at the same time. LTE-Advanced will be briefly introduced in Section 2.1.8. The 3GPP roadmap is shown in Figure 2.2.

The outcome of the work item is a technical specification describing standardised technology. Technical specifications are grouped in categories, each category focusing their specific technology area. The specifications for the *E-UTRAN* are contained in the 36 series of Release 8, Release 9 and Release 10 and divided into the following subcategories:

- 36.100 series covering radio specifications and evolved Node B (eNB) conformance testing.
- 36.200 series covering layer 1 (physical layer) specifications.

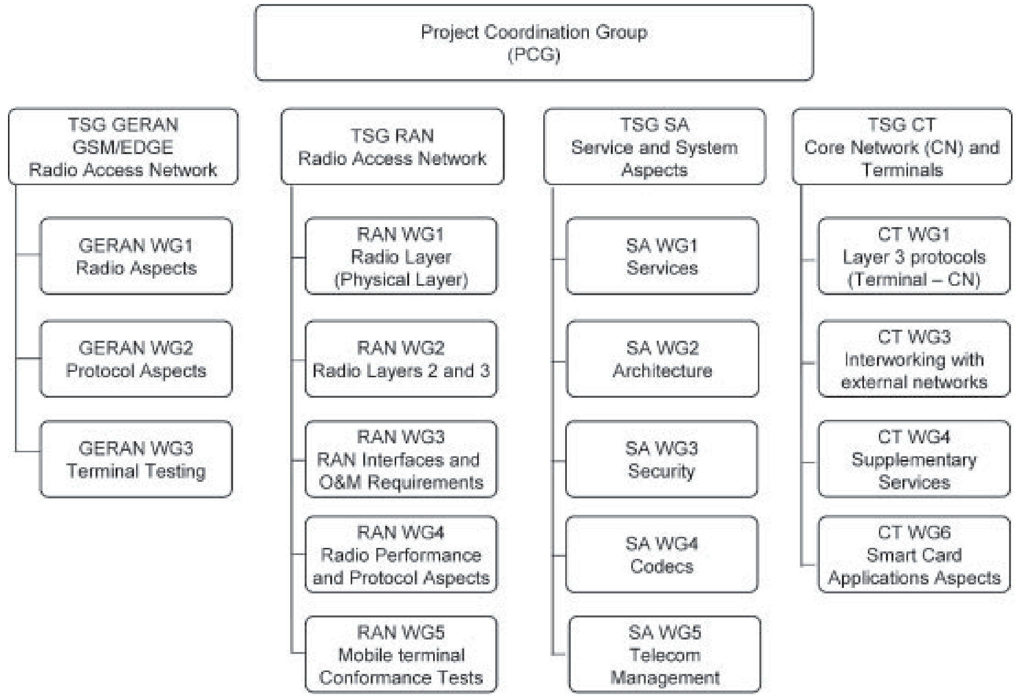

Figure 2.1 3GPP working structure. Adapted with permission from 3GPP.

- 36.300 series covering layer 2 and 3 air interface signalling specifications.
- 36.400 series covering network signalling specifications.
- 36.500 series covering user equipment conformance testing.
- 36.800 and 36.900 series, which are technical reports containing background information.

The specifications for *SAE* are scattered to many different specifications. The following documents cover the high level architecture of the SAE:

- 23.401 GPRS Enhancements for E-UTRAN access.
- 23.402 Architecture enhancements for non-3GPP accesses.

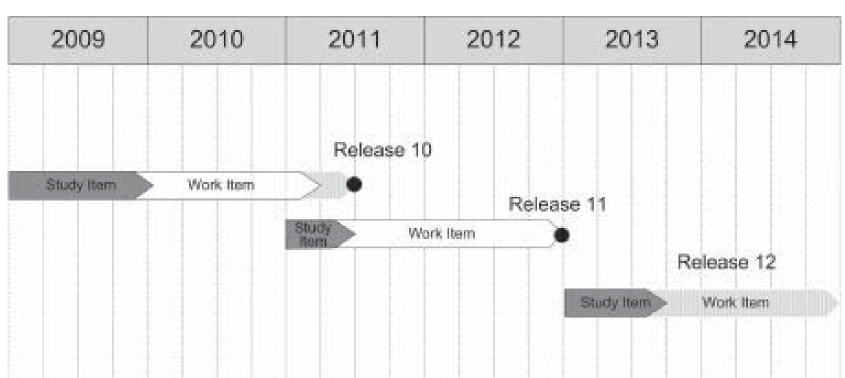

Figure 2.2 3GPP Roadmap.

Telecom Management LTE relevant specifications are:

- 32.100 series covering management principles, architecture and requirements, Fault Management Integration Reference Points (IRPs).
- 32.200 series covering charging management.
- 32.300 series: Common Management IRPs.
- 32.400 series: Performance and Trace Management IRPs.
- 32.500 series: Self-Organising Networks IRPs.
- 32.600 series: Configuration Management IRPs.

The latest versions of the LTE and SAE specifications can be found at the 3GPP site (http://www.3gpp.org/ftp/specs).

2.1.2 LTE Requirements

The LTE, the long-term evolution of the 3GPP radio-access technology, started with a Study Item in 3GPP with the scope of ensuring competitiveness for the next 10 years and beyond. LTE Requirements are described in (3GPP TR25.913, 2009). A key requirement set for LTE was that its performance should be superior if compared with 3G HSPA.

The LTE key performance requirements have been defined as comparison with HSPA R6:

- Spectral efficiency (bits/sec/Hz/site) in a loaded network two to four times more than HSPA R6 downlink and two to three times more than HSPA R6 enhanced uplink.
- Peak user throughput should be minimum 100 Mbps in downlink and 50 Mbps uplink within a 20 MHz spectrum allocation. The peak data rate should scale linearly with the size of the spectrum allocation.
- Frequency bandwidth flexibility from below 1.5 MHz up to 20 MHz allocations.
- Enables round trip time <10 ms.
- Packet switched optimised.
- Seamless mobility.

Table 2.1 summarises the main LTE requirements.

2.1.3 System Architecture Overview

As mentioned in the introduction 3GPP has defined a new system architecture for LTE. The EPS consists of the Evolved UTRAN (E-UTRAN), the Evolved Packet Core (EPC) and the connectivity to 3GPP and non-3GPP access systems. EPS solutions for 3GPP access are typically selected by operators who want to introduce EPS as smooth evolution to their existing 2G/3G infrastructure. EPS solutions for non-3GPP access are typically selected by operators who want to maximise the deployment of generic, non-3GPP protocols and to minimise the deployment of 3GPP specific protocols. The EPS system architecture is described in Figure 2.3. In this figure involved logical elements and interfaces between them are shown for the basic E-UTRAN configuration. In addition to this basic configuration, there are various architecture reference models specified in (3GPP TS23.402, 2011).

Table 2.1 LTE main requirements (3GPP TR25.913, 2009). Adapted with permission from 3GPP

Metric	Requirement	Conditions
System		
Round trip time	<10 ms	
Connection set-up latency	<100 ms	for idle to active
Operating bandwidth	1.4–20 MHz	
Coverage (cell sizes)	5–100 km	with slight degradation after 30 km
Mobility support	Up to 500 kmph but optimised for low speeds from 0 to 15 kmph	
Downlink		
Peak spectral efficiency	>5 bps/Hz	LTE in 20 MHz FDD
Peak transmission rate	>100 Mbps	2×2 spatial multiplexing
Cell edge spectral efficiency	>0.04–0.06 bps/Hz/user	2×2 spatial multiplexing Interference Rejection Combining receiver (IRC) 10 users per cell (high load)
Uplink		
Peak spectral efficiency	>2.5 bps/Hz	LTE in 20 MHz FDD
Peak transmission rate	>50 Mbps	Single antenna transmission
Cell edge spectral efficiency	>0.02–0.03 bps/Hz/user	Single antenna transmission Interference Rejection Combining receiver (IRC) 10 users per cell (high load)

In the EPS architecture only the Radio Access and the Core Networks are new, while the Service Connectivity Layer, that is the UE and Services, remains unchanged and is functionally the same as the other 3GPP systems.

The scope of EPS is to provide the IP Connectivity Layer and protocol layers in such a way that the elements are highly optimised for IP connectivity. EPS uses the concept of EPS bearer, with which an associated Quality of Service (QoS) parameters define the pipe where the IP traffic is routed from UE to the packet data network and vice-versa. One of the targets for EPS was simplified QoS scheme if compared to 3G.

The EPS architecture has three key aspects which address the performance requirements for LTE/EPC:

- Reduction of the number of network elements on the data path, compared to GPRS/UMTS;
- Streamlining of RAN functionality, by providing it in a single node.
- Separation of the control and user plane network elements (MME and S-GW).

The EPS system applies flat network architecture leading with two big architectural changes compared to EDGE and WCDMA. In the LTE access network all radio functionalities are collapsed into one network element only, the eNB and Controllers such as the RNC in

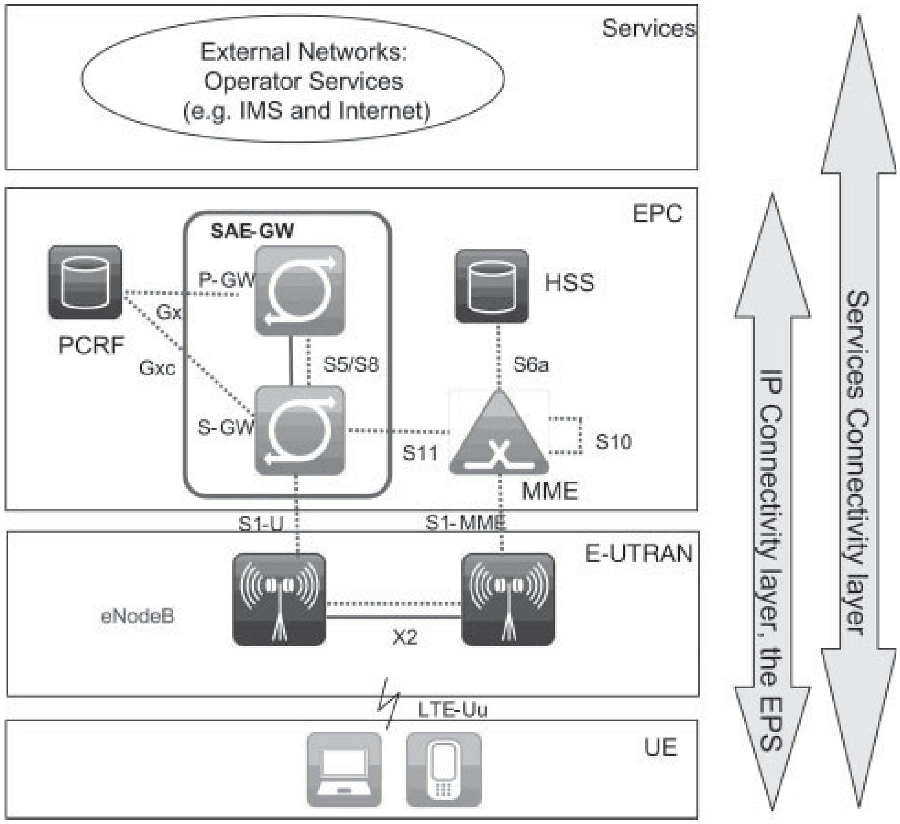

Figure 2.3 LTE System Architecture (Holma and Toskala, 2011). Reproduced with permission from John Wiley & Sons, Ltd.

WCDMA or the BSC in GERAN are not used. Also, the Core of the EPC does not contain the circuit switched domain.

2.1.4 *Evolved UTRAN*

LTE is based on Orthogonal Frequency Division Multiplexing (OFDM) for downlink radio transmission and data are carried simultaneously by narrow-band subcarriers. The signal is organised into subframes of 1 ms each as shown in Figure 2.4. This small duration, together with the flat network architecture, enables very short latency for both data and signalling.

For the uplink the requirement to limit as much as possible the power consumption of the UE transmitter leads to the use of Single Carrier FDMA.

LTE is very scalable with system bandwidth ranging from 1.4 MHz up to 20 MHz and it allows both paired (downlink and uplink use different frequency bands) and unpaired spectrum (downlink and uplink use same frequency band), using Frequency Division Multiplexing (FDD) and Time Division Multiplexing (TDD) both sharing the same downlink subframe structure.

Table 2.2 LTE performances for FDD at 20 MHz (3GPP TR25.913, 2009). Adapted with permission from 3GPP

Requirement	LTE	HSPA+
Peak transmission rate	DL: 150–300 Mbps	DL: 42–168 Mbps
At 20 MHz BW	UL: 75 Mbps	UL: 11–54 Mbps
Spectral efficiency (average) 4-rx mobile	1.7–2.7 bps/Hz/cell	1.21–1.9 bps/Hz/cell
Coverage (link budget)	162 dB	162 dB

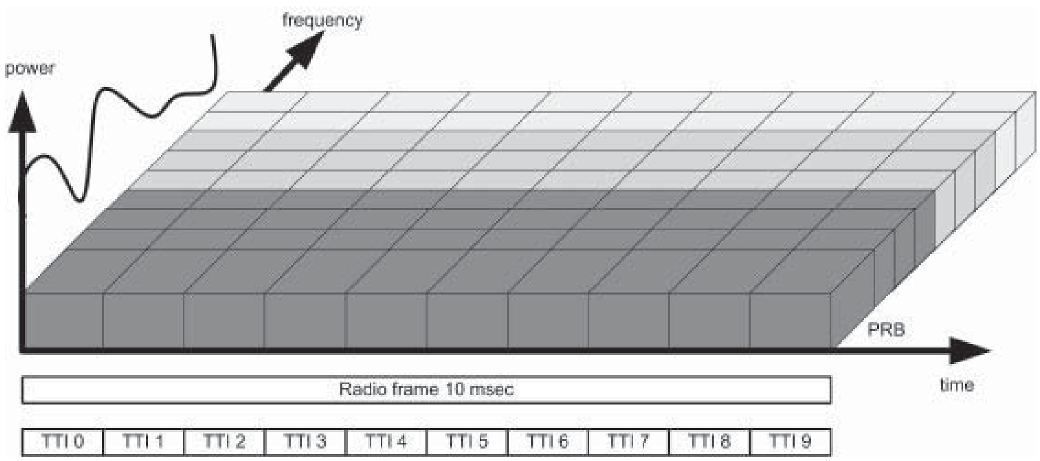

Figure 2.4 Time and frequency domain of the LTE system resources.

Multiple antennas are embedded in LTE. All UEs support at least two receive antennas, allowing downlink receive diversity. More advanced techniques, such as Transmit diversity, spatial multiplexing (Single-User and Multi-Users MIMO) and beam-forming are also supported.

LTE peak data rate increases by a factor of about 100 compared to HSPA+. Spectral efficiency also increases significantly while there are no substantial improvements in coverage. LTE and HSPA+ performances for a FDD system at 20 MHz are shown in Table 2.2.

2.1.5 E-UTRAN Functional Elements

The E-UTRAN consists of one node only, the eNB (eNodeB) that interfaces to the UE. The Radio Access Network is simply the meshed eNBs connected to neighbours eNBs through the X2 interface (Figure 2.5).

All radio functionalities are collapsed into the eNB, which means that the eNB is the termination point of all radio related protocols, the PHYsical (PHY) layer, the Medium Access Control (MAC) layer and the Packet Data Control Protocol (PDCP) layer that includes the user plane header compression and encryption.

The Control plane includes Radio Resource Control (RRC) functionality, with radio resource management, admission control and scheduling according to QoS policies.

Control and User planes in basic system architecture are shown in Figure 2.6.

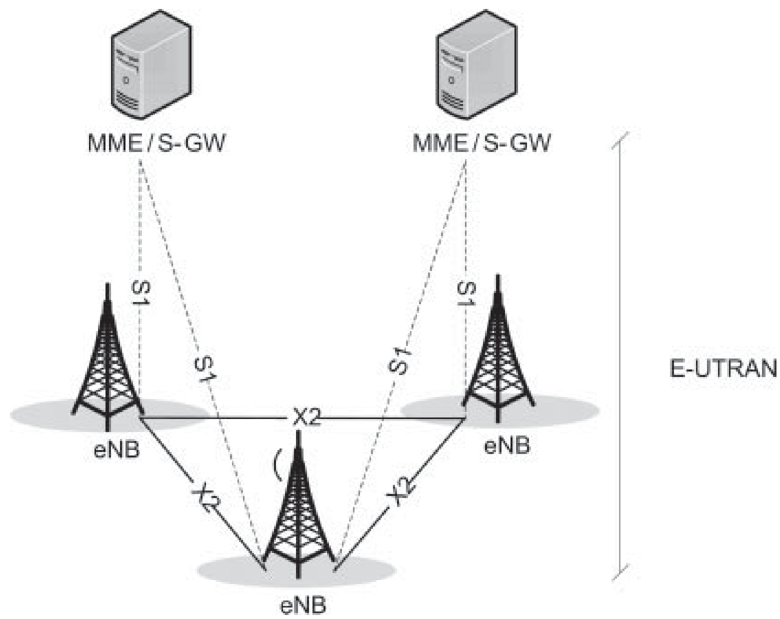

Figure 2.5 E-UTRAN Overall Architecture (3GPP TS36.300, 2011). Reproduced with permission from 3GPP.

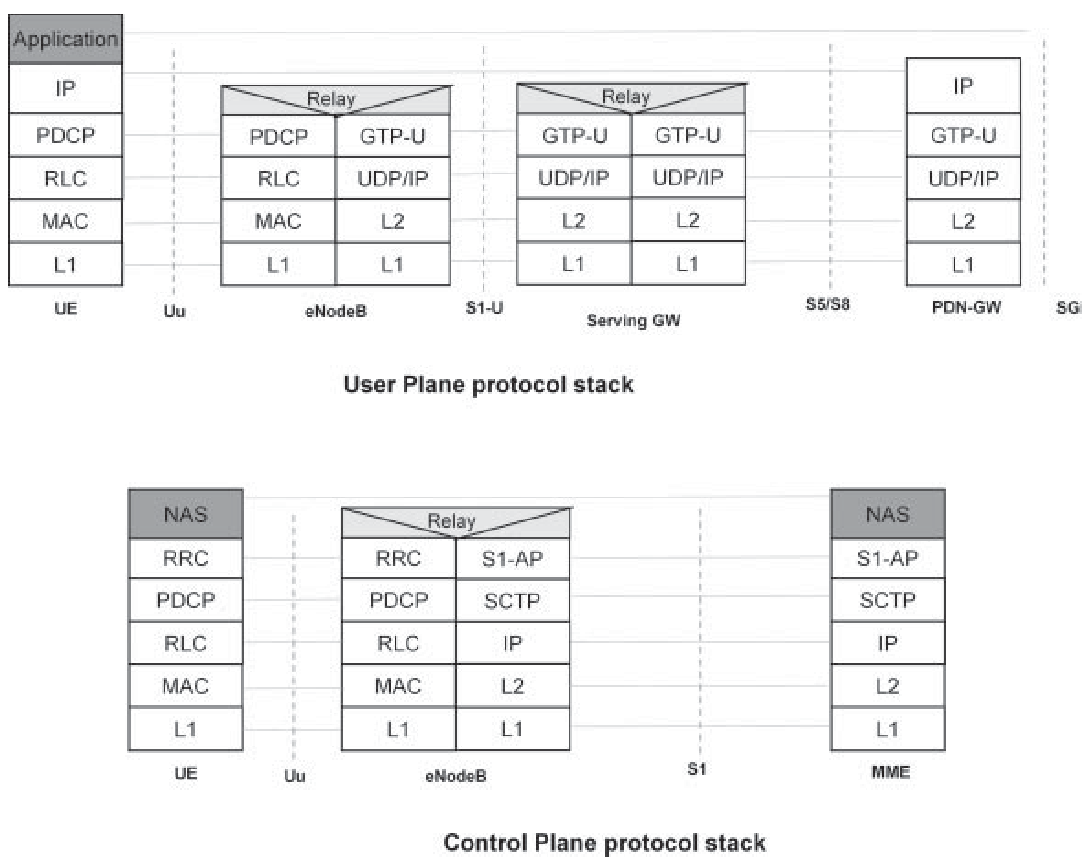

Figure 2.6 User and Control Planes in basic system architecture (3GPP TS36.300, 2011). Reproduced with permission from 3GPP.

2.1.6 Evolved Packet Core

The core network, EPC, is optimised for IP connectivity and it does not involve support for circuit switched domain. EPC contains the following logical network elements for controlling EPS bearer and UE: Mobility Management Entity (MME), Serving Gateway (S-GW) and Packet Data Network Gateway (P-GW).

MME takes the role of SGSN in current GPRS networks. MME is the control plane element allowing the user plane traffic to bypass over a direct tunnel.

The role of GGSN to provide connectivity to operator service networks and the Internet is taken by two gateway elements in the EPC; S-GW is the user plane (U-plane) gateway to the E-UTRAN and P-GW is the U-plane gateway to the packet data network (for example, the Internet or the operator's IP Multimedia Subsystem, IMS). S-GW and P-GW can be colocated or separated network elements. When S-GW and P-GW are colocated the gateway is called SAE-GW.

One of the targets for EPC is to arrange optimised interworking with other 3GPP access networks and other wireless access networks.

2.1.6.1 Evolved Packet Core Functional Elements

The Evolved Packet Core is composed of several functional blocks (Figure 2.7).

For the establishment of EPS bearer and overall UE control the EPC includes the MME, S-GW, and P-GW.

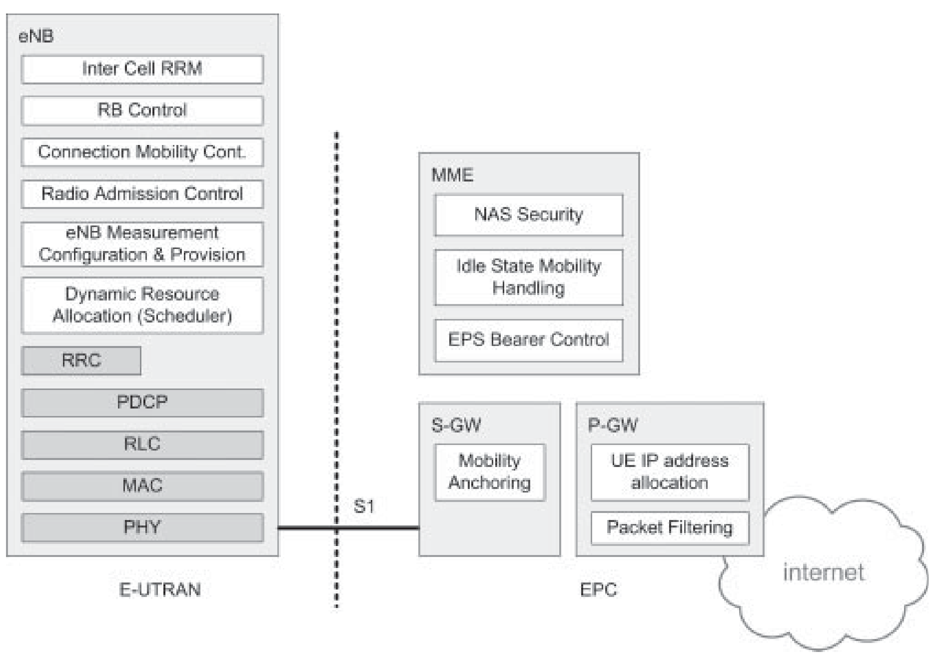

Figure 2.7 E-UTRAN and EPC functional split (3GPP TS36.300, 2011). Reproduced with permission from 3GPP.

For the user subscription data, both permanent and location-based, and user's policy and charging control the EPC includes the:

- Home Subscriber Service (HSS).
- Policy Control and Charging Rules Functions (PCRF).

The functional split between E-UTRAN and EPC is represented in Figure 2.7.

2.1.6.2 Mobility Management Entity

The MME is the central element in the EPC and it provides a logically direct Control plane connection (*Non-Access Stratum* or *NAS* signalling) with the UE. The MME manages and stores UE context, generates temporary identities and allocates them for UEs, authenticates the user, manages mobility and bearers, and is the termination point for NAS signalling. MME is responsible for:

- Mobility Management procedures:
  - Idle and active mode UE tracking in its service area and paging procedures.
  - MME controls the setting up and release of resources based on the UEs activity changes and participates to handover signalling.
- Authentication and security:
  - Interacting with the HSS MME verifies the UE authentication credentials and, to protect the UE privacy, it allocates each UE a temporary identity called the Globally Unique Temporary Identity (GUTI) so that the need to send the permanent UE identity, the IMSI, over the air interface is minimised.
  - Lawful interception of signalling for mobility between 2G-3G and LTE access networks through the S3 interface with the SGSN.
  - Providing control plane functions.
- Management of subscription profile and service connectivity. The MME retrieves the UE subscription profile from the home network, which determines the packet data network to which the UE should be connected at attachment.

2.1.6.3 Serving Gateway

S-GW is the U-plane gateway to the E-UTRAN. The primary function of the S-GW is a user plane tunnel meaning that all user plane packets go through it. The S-GW serves as an anchor point both for user moving across LTE cells and making inter-eNB handovers and anchor for mobility between LTE P-GW and other 3GPP technologies, meaning handovers to and from 2G/3G systems. The anchor point is the point which is common for all 3GPP accesses technologies, 2G, 3G and LTE. User data is routed to UE independently of underlying radio technology or changing radio technology due to handover.

S-GW is also responsible for packet forwarding, routing, and buffering of downlink data for UEs that are in LTE-IDLE state. In addition it terminates downlink data path for users in idle state and triggers paging when downlink data arrives for the UE. S-GW also replicates user traffic for lawful interception.

2.1.6.4 Packet Data Network Gateway

P-GW is the U-plane gateway to the packet data network (for example, the Internet or the operator's IP Multimedia Subsystem: IMS). The P-GW is responsible for policy enforcement, charging support, and user's IP address allocation. It also serves as a global mobility anchor for mobility between 3GPP and non-3GPP access, and LTE and pre-Release 8 3GPP access.

The P-GW is the edge router between EPS and external packet data networks. Typically the P-GW allocates the IP address to the UE and includes the Policy Control Enforcement Function (PCEF) which means it performs throttling, gating and filtering functions for user data including charging information reporting. A UE may have multiple and simultaneous connections to many external networks.

2.1.6.5 LTE/EPC Related Legacy Network Elements

The legacy network elements that interoperate with EPS (Figure 2.8) are the following:

- Serving GPRS Support Node (SGSN).
- Home Subscription Server (HSS).
- Policy and Charging Resource Function (PCRF).
- Authentication, Authorisation and Accounting function (AAA).

SGSN is responsible for the transfer of packet data between the Core Network and the legacy 2G/3G RAN. For EPS this node is only of interest from the perspective of inter-system mobility management.

HSS is the Core Network entity responsible for managing user profiles, performing the authentication and authorisation of users. The user profiles managed by HSS consist of

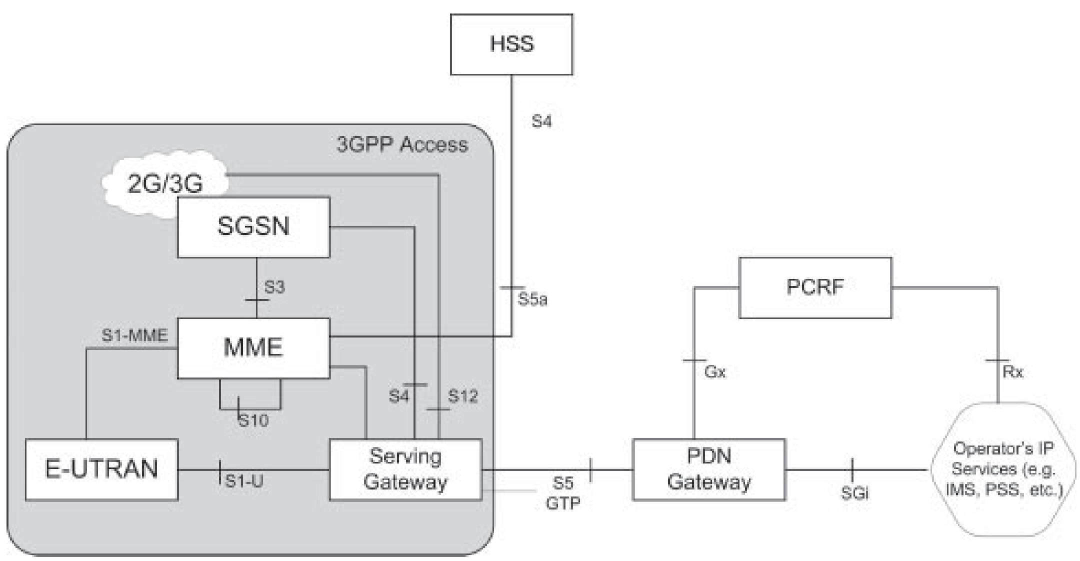

Figure 2.8 Non-roaming reference architecture for 3GPP (2G/3G/LTE) access within EPS using GTP based S5). Reproduced with permission from 3GPP.

subscription and security information as well as details on the physical allocation of the user. The HSS stores the user subscription data, indicating the services the user may use and the PDNs to which the user can connect to. In addition it records the location of the user of the last visited network MME.

The PCRF is responsible for Policy and Charging control and it makes decisions on how to handle services in terms of QoS and provides devices to the PCEF located in the P-GW.

AAA is responsible for relaying authentication and authorisation information to and from non-3GPP access network connected to EPC.

2.1.7 Voice over LTE (VoLTE)

Earlier in this chapter it has been explained that EPS is an all-IP technology. We can get to the conclusion that the voice service will have to be delivered in a new way. Voice in this IP world, would be implemented as Voice over IP (VoIP). 3GPP specified way to support VoIP is IP Multimedia Subsystem (IMS). IMS is an access-independent and standard-based IP connectivity and service control architecture that enables various types of multimedia services to end-users using common Internet-based protocols (3GPP TS23.228, 2011). The IMS based VoLTE solution puts the IMS in the centre of the voice core network, managing the connectivity between subscribers, taking care of charging and the implementation of policy control. The voice service (supplementary service, service continuity between LTE and CS, etc.) is then further managed by voice application servers on top of the IMS.

In addition to the 3GPP specifications defining the detailed architecture as well as protocol level requirements for these services, GSM Association (GSMA) has published more specific profile for *IMS profile for voice and SMS* as IR.92 document (GSMA PRD IR.92, 2010). The IR.92 contains agreed mandatory and optimum set of functionalities for the UE, the LTE Access Network, the Evolved Packet Core Network and the IP Multimedia Subsystem functionalities to manage voice and SMS in LTE based on 3GPP specifications. In a sense, this technical profile gives all industry stakeholders a level playing field on which to enhance their VoLTE service as they see fit, but most importantly a level playing field that enables the basic working, and interworking, of VoLTE across the entire industry landscape because over the time.

Although the road towards LTE, IMS and All-IP is clear, 3GPP has defined an alternative approach, CS-FallBack (CSFB) (3GPP TS23.272, 2011). In CSFB, whenever an LTE mobile generates or receives a voice call it is automatically transferred to the 2G or 3G networks. Once the call is finished the mobile reverts to LTE. This is a valid solution, but it relies on interrupting the LTE connection when the terminal is forced to move to the 2G or 3G network. This might be a big problem, depending on the application that is being used prior to the voice call. The CSFB solution is assumed to be interim VoLTE roaming solution. Furthermore, it fits to operators who plan to provide hotspot LTE coverage and/or wish to re-use their legacy networks for a while.

2.1.8 LTE-Advanced

Looking ahead, the exponential growth of data traffic will continue in upcoming years; enabling factors are the adoption of mobile broadband services, increase in usage intensity, great availability and choices of devices and machine-to-machine communications.

Table 2.3 LTE-A Requirements (3GPP TR36.913, 2011). Adapted with permission from 3GPP

LTE-A Requirement		Conditions
Peak transmission rate (20 MHz × 5) BW	DL: 1 Gbps UL: 500 Mbps	
Peak spectrum efficiency	DL: 30 bps/Hz UL: 15 bps/Hz	4 antennas BS and 2 antennas terminal
Average spectrum efficiency	DL: 2.6 bps/Hz UL: 2.0 bps/Hz	
Cell edge spectrum efficiency	DL: 0.09 bps/Hz UL: 0.07 bps/Hz	
Latency	User plane: 10 ms Control Plane: 50 ms	

In addition to these key drivers, analysis of data shows that data traffic is/will be distributed in an uneven way. These facts call for higher bandwidth and higher network efficiency which can be obtained by combining several tools, each of those tailored for specific network scenarios.

While Release 8 and Release 9 LTE have been optimised for conventional wide area deployment based on macro base stations and dual receiver and single transmit antenna single band terminals, LTE-Advanced targets more complex scenarios. In fact focus for LTE-Advanced is not to introduce a new air interface technology, but rather to extend features and capabilities of LTE to support new network deployments and ensuring optimal distribution of services.

LTE-Advanced is defined in Release 10 targeting IMT-Advanced requirements, as defined by ITU-R, while maintaining backwards compatibility with previous LTE versions; backwards compatibility means that an LTE Release-8 and Release 9 terminal will be able to operate in the LTE-Advanced network and vice versa. Key areas of enhancements are improved data-rate, coverage extension and latency reduction as well as inter-working with other technologies and global roaming (3GPP TR36.913, 2011), (Ghosh *et al.*, 2010), (Mogensen *et al.*, 2009). A summary is reported in Table 2.3.

LTE-Advanced includes several technology components as shown in Figure 2.9: Carrier Aggregation (CA) allows achieving peak data rate of 1 Gbps in downlink and 500 Mbps in uplink. Enhancements for uplink and downlink MIMO target at improving the spectral efficiency; Relay Nodes (in-band backhauling) and Coordinated Multipoint transmission and reception (CoMP; CoMP is not included in Release 10) cope with high-cell interference at cell-edge. Optimised interworking between cell layers is also introduced for Heterogeneous Networks. (see Section 2.1.8.5 and Chapter 10).

Each of the LTE-Advanced features is shortly presented in next paragraphs. As far as their SON component is concerned, a detailed description of Relay Self-Configuration and Automatic Neighbour Relations (ANR) is included in Chapter 4, and Chapter 10 is dedicated to Heterogeneous Networks.

At the time of writing SON applicability to the other LTE-Advanced feature is under analysis.

2.1.8.1 Carrier Aggregation

The very high peak data rate targets are fulfilled by means of bandwidth extensions.

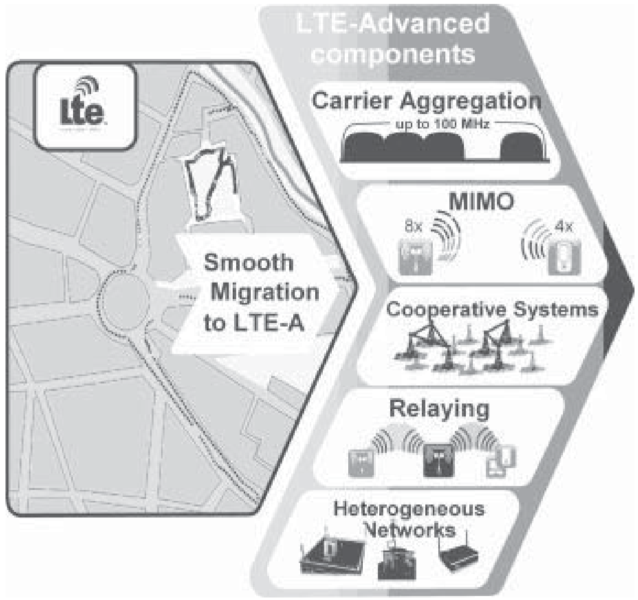

Figure 2.9 LTE-Advanced components.

Carrier Aggregation (CA) allows for combining up to five LTE Release 8 Component Carriers (CC) to achieve a transmission bandwidth of 100 MHz (5×20 MHz) and enhancing end-users peak data rates to 1 Gbps in downlink and 500 Mbps in uplink while maintaining backwards compatibility. Aggregation of Component Carriers is per user based, so that different users in a cell may be differently configured (Figure 2.10).

In order to enable operators to provide high throughput without wide continuous frequency band allocations, 3GPP defines, in addition to contiguous spectrum in a single band other two levels of aggregation, the 'non contiguous spectrum in single band' and the 'non contiguous spectrum in multiple band'. In addition, Carrier Aggregation allows asymmetrical bands into use with FDD since there can be uplink or downlink only frequency bands.

2.1.8.2 Improved MIMO Schemes

Multi-antenna technologies include spatial diversity and beam-forming. Antenna diversity is very effective in multi-path propagation handling, where the signal is reflected along multiple paths which may destructively interfere with one another, before being finally received.

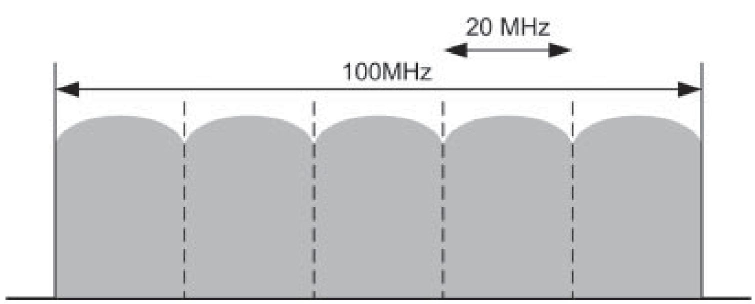

Figure 2.10 Carrier aggregation.

If the multiple transmissions is for a single user then the technology is called Single-User MIMO (SU-MIMO), for multiple users Multi-User MIMO (MU-MIMO).

MIMO performance is subject to a large number of parameters: the number of transmitting and receiving antennas, reference signals and algorithms for channel estimation, feedback of channel estimation data from the receiver to the transmitter and spatial encoding methods.

Release 8 and Release 9 support MIMO with up to four transmitting and receiving antennas in downlink, but with only one antenna in uplink. Release 10 extends downlink MIMO to support up to eight transmitting and receiving antennas, and uplink to support up to four transmit and eight receiver antennas.

2.1.8.3 Relay

The very high data targets of LTE-Advanced require a tight infrastructure (short distance between base stations) in order to be able to reach cell-edge users, which usually suffer from relatively low Signal-to-Interference plus Noise ratio (SINR). An attractive solution to this problem is provided by multi-hop technologies by means of Relay Nodes. The idea behind Relay Nodes is the improvement of the link budget by reducing the transmitter-to-receive distance for those users close to the cell edge.

In fact, deploying Relay Nodes near to the cell edge will help to increase the capacity or, alternatively, to extend the cell coverage area (Figure 2.11). The Donor eNB will 'donate' part of its air-interface capacity to providing backhaul connection for one or more Relay Nodes while still serving own users. Two types of Relay Nodes are envisaged. The simplest form is the conventional Amplify and Forward (AF) Relay, also called the repeater. AF simply amplify-and-forwards the signal received from Donor eNB. The drawback of such a solution is quite obvious, that is the AF Relays amplify both interference and noise in addition to the desired signal and are therefore not effective in presence of noise or interference.

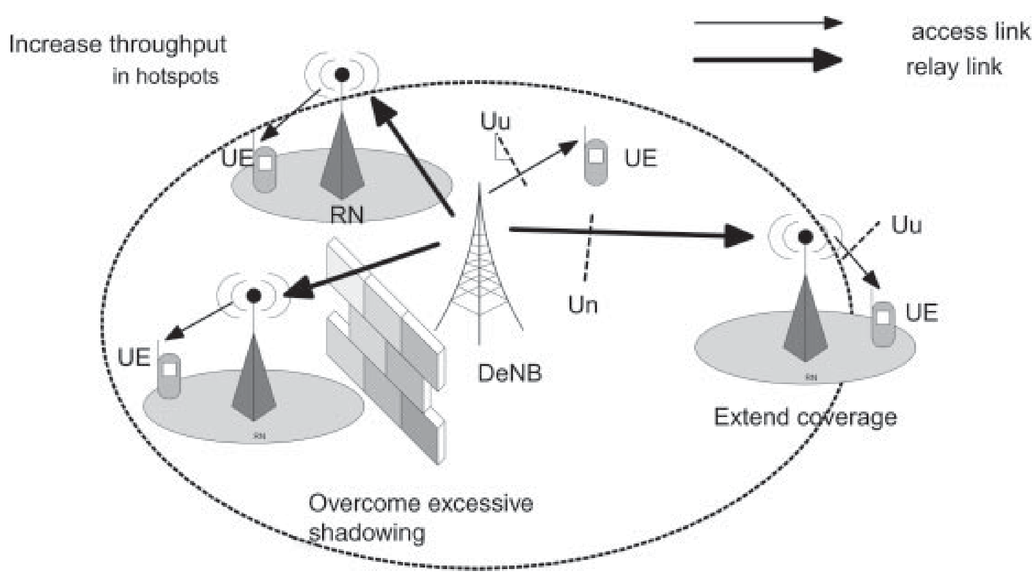

Figure 2.11 Donor eNB and relay nodes.

The most advanced Relay Nodes are the Decode and Forward (DF) ones, which detect and decode the desired signal and then re-encode and forward it. Therefore DF relays are applicable both in interference and noise limited environments and thus, they can be used to improve system capacity.

Relay Nodes can be seen as a special case in a Heterogeneous Network. For Relay Nodes the backhaul is being provided by the Donor eNB, the Relay node has some special configuration procedures and Automatic Neighbours relations (ANR) procedure is a bit different compared to other nodes. Relay self-configuration and ANR are described in dedicated paragraphs in Chapter 4.

2.1.8.4 Coordinated Multipoint Transmission and Reception (CoMP)

Coordinated Multipoint transmission and reception (CoMP) shows great potential to improve cell edge data rates and system capacity. It deals with low signal quality at cell edges and interference levels: due to frequency reuse one (i.e. systems where frequency channels are reused for achieving high level of area coverage), the cell-edge performance of LTE is limited by co-channel (inter-cell) interference.

Despite studies have shown high potential from CoMP for a single user, this was less evident in the case of large scale networks operation.

The Coordinated Multipoint processing requires close cooperation between a number of geographically separated eNBs which coordinate amongst them to provide joint scheduling and transmissions as well as performing joint processing of the received signals (see Figure 2.12). This leads to very high signalling load between the cells: as a consequence, the Intra-site CoMP deployment, where the communication is between the sectors of a single eNB, is likely to be the most feasible system solution.

Due to these reasons the technology was not seen mature enough to be included it in Release 10, and studies will continue in Release 11 with focus on practical concepts and real performance benefits, including transport technologies.

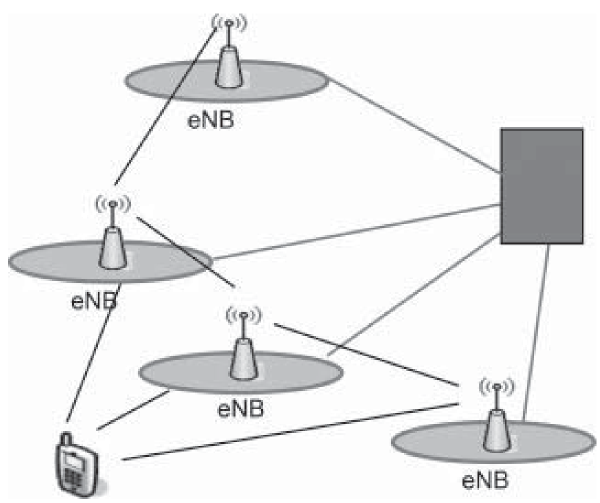

Figure 2.12 Coordinated multi-point.

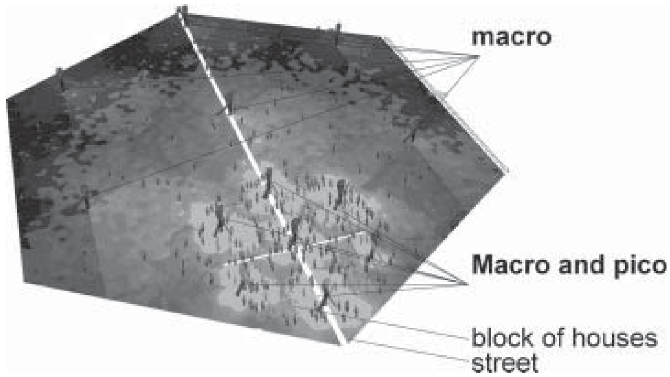

Figure 2.13 Heterogeneous Networks.

2.1.8.5 Heterogeneous Networks

Further spectral efficiency can be achieved with multi-layers network topologies; in this scenario macro cells provide the continuous and wide area coverage while small cells (served by low-power nodes) improve hot spot performance (Figure 2.13). Small cells can be served by a micro eNB, pico eNBs or Home eNodeBs (HeNB, also known as femto eNB).

Low-power nodes such as pico eNBs are usually deployed in a coordinated way that is under the operator's control by means of conventional network planning. While Enterprise HeNBs are installed by the operator, the Residential HeNBs are typically installed by the end-user without operator intervention. Therefore the latter is also called 'uncoordinated deployment'.

HeNB operating modes are *Open Subscriber Group* (*OSG*), *Closed Subscriber Group* (*CSG*) or *Hybrid* (meaning they operate with both OSG and CSG). CSG HeNB usage is constrained to its owner only or an otherwise limited set of users. Hybrid HeNB allows all UEs of the given operator to access the HeNB, but CSG members may receive preferential handling. Open and Closed Subscriber Group HeNBs are further explained in Section 10.2.

No interference problems exist if the operator is able to allocate different frequencies to different cell types (e.g. macro cell versus femto cell). If this cannot be done and the same frequency has to be used for the different cell types, the cardinal problem is the interference between users served by a macro eNB and the users served by a HeNB (see Figure 2.14).

Another problem with the HeNBs is that they may not have a fixed position (they can be moved inside the house) and/or be switched off at any time by the owner. In 3GPP Release 10, the Enhanced Inter-Cell Interference Coordination (eICIC) has been specified to mitigate interference. Heterogeneous Networks and the related interference mitigation mechanisms are described in Chapter 10.

Heterogeneous Networks have lately attracted very much attention from the operators. Pico-eNBs may be used both for coverage and capacity reasons (e.g. in shopping malls). In this constellation the cells are deployed in a coordinated way, that is, by the operators, as they operate in Open Subscriber Group mode, and can be placed both, indoor or outdoor.

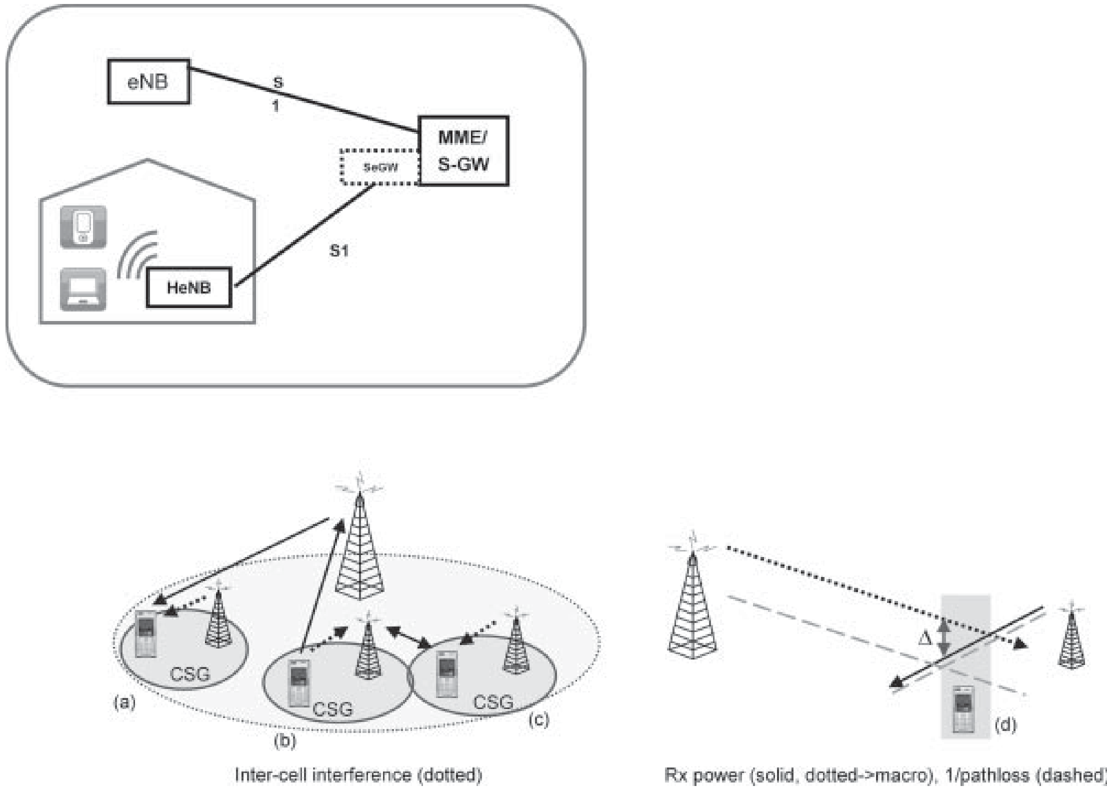

Figure 2.14 Residential HeNB and interference with macro (3GPP TR36.814). Reproduced with permission from 3GPP.

Actually multi-layer networks are already part of 3GPP Release 8 and 9 specifications. New in Release 10 are the mechanisms that very much improve the performances of such networks.

Chapter 10 analyses Heterogeneous Networks and their management in detail. Due to the complexity of this deployment and the lack of operator planning, a high degree of automation is needed to manage and properly use the network resources.

2.1.9 Network Management

LTE follows the same 3GPP management reference model as 3G as specified in 3GPP TS32.101 (2010) (Figure 2.15). It introduces several interfaces from operations systems to NEs:

- Itf-S (interface 1 in Figure 2.15): Interface between the Network Element (NE) and the Domain Manager (DM). This interface is vendor specific.
- Itf-N (interface 2 in Figure 2.15): Interface between the DM and the Network Manager (NM). This is a standardised open interface and thus facilitates multi-vendor management.
- The Itf-P2P (interface 4a in Figure 2.15), interface between the DM as well as the Element Manager (EM) embedded into the Network Element. Itf-P2P interfaces have been not used in real deployments.

- Collaborative algorithms for dynamic spectrum management exploiting cognition techniques, including spectrum sensing and information provisioning mechanisms.
- Collaborative algorithms for Joint Radio Resource Management (JRRM) enhanced with cognition techniques.
- Collaborative algorithms for the management of Flexible Base Stations (FBSs).
- Cognitive Pilot Channel (CPC) definition as enabler for autonomous and collaborative algorithms, for example, regarding message structures, bit rates required, mesh optimisation.

As with the SOCRATES project, the major impact of E^3 came through the participation of the project partners in and the corresponding dissemination of project results to 3GPP standardisation and the NGMN forum.

3.3.3.4 COST 2100 SWG 3.1

COST is a European, inter-governmental cooperation framework in the field of Scientific and Technical Research (http://www.cost2100.org). It is based on Actions, which are networks of coordinated national research projects in fields of interest to participants coming from different COST and non-COST countries. COST2100 is the Action on Pervasive Mobile & Ambient Wireless Communications and belongs to the ICT Domain. This Action basically addresses the various topics that are emerging in the area of mobile and wireless communications. The Sub-Working Group (SWG) 3.1 focuses on mobile wireless network optimisation aspects with special attention to the data used as an input to the optimisation process. The goal is to substitute artificially generated data from existing simulation- and prediction-based optimisation tools by measured data taken from the running system (e.g. OAM system, network elements and probe/drive testing tools). The SWG3.1 activities include:

- Network performance criteria development, measurement data definition, acquisition, filtering, classification, correlation, grouping and separation.
- Optimisation parameters definition and selection, parameter correlation with counters and quality indicators.
- Network modelling and model tuning with use of measured data, optimisation method development, algorithm implementation.
- Comparison between measurement and simulation based optimisation approaches including reference scenario definition, field tests for verification with practical trials.

The COST 2100 SWG3.1 has cooperated with other research projects in the area of SON and mobile network optimisation, for example, the GANDALF and SOCRATES projects. As COST2100 is actually a framework and not a funded research project, results are available to the participating institutions only.

3.4 Architecture

This section discusses the architecture of a SON system; in particular, general considerations are presented on where SON functions should be located in the OAM hierarchy. The considerations comprise both use case-specific and system-specific critera.

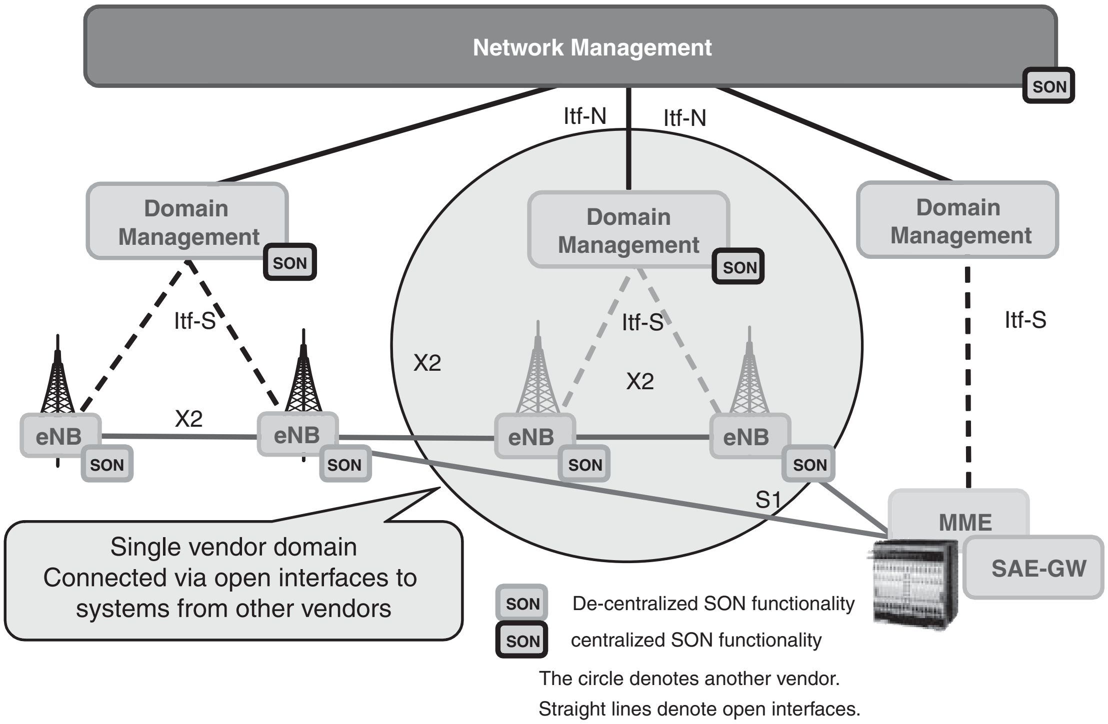

Figure 3.11 Location of SON functions in the 3GPP OAM architecture (Sanneck, Bouwen and Troch, 2010). Reproduced with permission of © 2010 IEEE.

Figure 3.11 shows the 3GPP OAM architecture as it has been introduced in Section 2.1.9. A network element (such as an eNB) is managed by its (vendor-specific) Domain Manager (DM). Different such vendor domains can then be managed in a uniform way via the northbound interface (Itf-N) at the Network Management (NM) level. In the figure, it is shown that SON functions can be located at the:

- **NM level:** *centralised SON* (using *standard* or proprietary interfaces).
- **DM level:** *centralised SON* (using a *proprietary* interface).
- **NE level:**
  - *distributed SON* (using *standard* or proprietary interfaces);
  - *local SON* (having no dependencies to other NE, hence requiring no interfaces).

A combination of a set of SON functions located at different levels can be called a *hybrid SON* system. A single SON function can also be hybrid in that different function components execute at different levels (an example being the ANR function, cf. Section 4.2.3).

It should be noted that SON functions themselves are entities which are 'managed' using the OAM architecture depicted in Figure 3.11. However, 'management' refers here to setting targets, configuring the SON function behaviour at a high level, and monitoring SON function results rather than directly changing the low-level configuration and monitor low-level performance indicators for the SON function (cf. the 3GPP SA5-related part in Section 3.2.3).

Data (KPIs, alarms) which drive the SON functions are obviously generated at the individual NEs. That means that for the distributed SON approach, information (data, requested

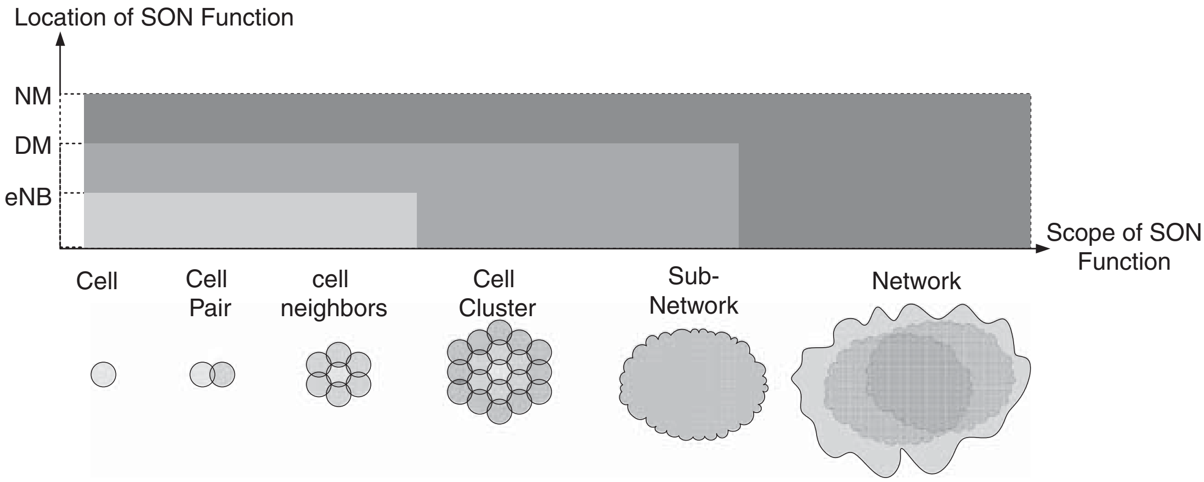

Figure 3.12 SON function spatial scope versus SON function execution location.

reconfigurations) needs to be exchanged between the NEs. In a centralised SON approach, first data needs to be transferred from the NE to the DM/NM-level. After a SON function has been executed the requested reconfiguration needs to be conveyed to the relevant NE in turn.

In the following sections, criteria for the choice to locate a SON function at any of these levels are discussed.

3.4.1 Use-Case Related Criteria

- **Spatial scope:** Figure 3.12 shows the relationships between the scope of a SON function (ranging from an individual cell to an entire network) and its execution location (NE/DM/NM-level). On one hand, it can be seen as natural to locate a function with a network-wide scope (covering multiple vendor domains) at the NM level, because the function is driven by data conceptually residing at that level (e.g. self-configuring a parameter which needs to be identical for all NEs in a network). On the other hand, it is also technically possible to execute such functions in a fully distributed way (for the example given above this would mean that an NE retrieves a network-wide parameter from the NM-level which is then propagated from NE to NE).

  For functions having a rather limited scope like a cell pair, also, on one hand it is natural to execute them on the NE, because the function is driven by data already available on two NEs. On the other hand, there is also no fundamental technical blocking point to execute it at the DM/NM level instead.

  Note that the key SON functions discussed in the following chapters of this book have typically a scope of up to a cell cluster level.
- **Timing requirements:** SON functions have very specific time intervals in which they acquire data, execute and request reconfigurations of NEs (this is elaborated in Section 9.1 on SON function interactions). These time intervals (ranging from minutes to days) are typically larger than those of RRM functions (cf. Section 5.6) and smaller than those of conventional network optimisation functions. The novelty with SON compared to choosing

conventional OAM function locations is that in fact these time intervals are somewhere 'in the middle' of these two extremes, hence the choice where SON functions should reside in the architecture is not straightforward.

It can be argued that SON functions requiring relatively frequent data acquisition and execution advocate more a distributed location of functions, whereas functions which rarely execute should be realised in a centralised way. Similarly to the spatial scope consideration above, this is just a rough guideline. If the loop of data acquisition, transfer, execution/decision making and transfer of the reconfiguration requests is fast enough and sufficiently dimensioned in the concrete system instantiation, any function can be implemented in a centralised way and be scalable (this is the link to the system-level criteria presented below).

- **Standardised versus proprietary parameters:** In a typical cellular network there are tens of thousands of network elements, leading to millions of parameters to be maintained. Even though most of the parameters can be set to their default or vendor-defined values, or can be specified through network element and environment specific templates, the number of residual parameters is very high. So even if only few percent of all parameters need to be considered for daily configuration and optimisation work, still several hundred thousand parameter values are subject of daily management in a typical network. Parameters can be classified being either vendor-specific or standardised also contributing to the execution location choice for SON functions manipulating those parameters:
  - **Standardised parameters (3GPP, IETF, ATM Forum, etc.):** These parameters have been standardised in order to configure properties of standardised external interfaces (air interface, backhaul transmission line), cf. Section 4.1 (e.g. VLAN, IP addresses), Section 4.2 (e.g. frequency bands, cell IDs) and Chapter 5 (e.g. handover parameters in Section 5.1.3). They have common semantics across all vendors and thus can be configured and optimised independently from the vendor of the specific NE. This class of parameters accounts for about 10–20% of all parameters (note: a radio cell typically has about 500 parameters). Many of those parameters are touched during daily work in network operation.
  - **Vendor-specific parameters for call processing features:** For example, 3GPP intentionally did *not* standardise the logic within the elements to decide on handover (for handover the message sequence between the network elements is standardised, but in order to offer possibility for vendors to differentiate and to introduce enhanced features, the logic to decide on handover is not standardised; cf. Section 5.1). So most parameters that configure the decision logic for handover are vendor-specific. In most cases those parameters cannot be mapped across vendors, even if semantics might be similar. Thus they *cannot* be optimised in a vendor-independent way, but nevertheless must be tuned very carefully.
  - **Vendor specific parameters for low level hardware and internal properties of the NE:** Those parameters usually are set during commissioning and hardly ever are modified during the lifetime of a network element. Thus they are not further considered in the context of SON.

  In total, vendor specific parameters account for about 80–90% of all parameters.
- **Amount of data/processing required:** NE resources for processing are usually limited due to the relatively high cost of installing memory and processing power in NE when comparing to the DM/NM level. This advocates centralised realisation. On the other hand, the transfer of data mentioned above from the NE to the DM/NM level consumes OAM bandwidth and

causes delay. Again these use-case specific constraints need to be considered in the framework given by the concrete system instantiation.

- **Decision making:** Centralised decision making based on inputs from several NEs is usually straightforward, whereas distributed decision making needs to be carefully planned and controlled to avoid classical concurrency issues in distributed systems like oscillations, race conditions and deadlocks. Also, distributed decision making may work efficiently for spatial scopes like a cell cluster, but may not scale to, for example, domain-wide tasks.

3.4.2 System-Level Criteria

- **Scalability:** As mentioned above, SON functions usually have a scope up to a cell cluster-level, based on which a per-function scalability analysis for a distributed versus a centralised approach needs to be performed. Additionally (for both conventional OAM and SON functions), there is an upper bound on the number of NEs which can be treated within a centralised administrative OAM/SON domain. That means, for scenarios with a very high number of NEs (cf. Chapter 2.1.8.5) it may be required to introduce an additional (DM) level of hierarchy to improve scalability, whereas in a distributed SON approach, no additional such means are required.
- **Reliability and availability:** It can be argued that the centralised SON approach contains with the DM-/NM-level entities a single point of failure, whereas the distributed approach has some inherent redundancy being a distributed system. While the general argumentation is valid, of course, centralised SON functions are added to an OAM system which usually already contains redundancy mechanisms like server clustering to improve reliability and availability.
- **Multi-vendor capability:** Both the NE- and the NM-level provide for 3GPP-standardised multi-vendor integration whereas at the DM-level only a proprietary integration is possible. SON-related standardisation at the NE-level (3GPP RAN WGs) is orientated along the tight call processing standardisation, that is, the interoperability regarding the exchange of data is assured. However, this comes at the price of a rather long standardisation process. On the other hand, NM-level standardisation for SON could be (similarly to OAM standardisation in general) considered to be faster but to leave more room for vendor-specific interpretation. In general, the interoperability regarding (vendor-specific) algorithms inside SON functions is not directly covered in the standards, cf. 'standard versus proprietary parameters' above.
- **Management/controllability:** Particularly in the initial SON deployment phase (cf. Section 3.6 on operational challenges), operators want to exercise still a lot of control on the system and tightly monitor what SON functions are doing. New operational workflows (cf. Sections 3.2 and 3.6) have to be realised by the combination of SON functions with existing (human-level) workflows.

  Both requirements are facilitated by a centralised approach, because all the SON-related data and the automated SON decision making are co-located with the function via which the operator exercise control and monitoring. In a distributed approach, however, specific additional instrumentation of SON functions may be required (i.e. additional data may need to be transferred from the NE to the OAM system just to satisfy the operator's control and monitoring requirements, thereby assuring consistency of the distributed configuration changes with operator policy).

- **Extensibility:** Similar to the previous point, a single entry point (as in a centralised architecture) for doing upgrades for the existing SON functionality and adding new SON functions facilitates the operability of the SON-enabled system itself.
- **System legacy and lifecycle:** There are evolution paths to SON functions from both the RRM as well as the network optimisation tool domains. If an existing optimisation tool is evolved into a set of SON functions, the natural architecture choice will be centralised SON. If RRM functions (cf. Section 5.6) serve as the baseline for a SON function, the approach will typically be a distributed one.

There may also exist a distributed versus centralised distinction regarding the system lifecycle: in initial LTE deployments it may be sufficient to rollout individual distributed SON functions for basic network configuration and optimisation, requiring few or no support from the centralised OAM system. Later in the life cycle, when the full set of optimisation and troubleshooting functions are required, the OAM systems have been upgraded to provide centralised SON support thereby covering domain- and network-wide operational workflows.

In summary, concluding from the above points, it is crucial to select the SON function execution location and thus the SON architecture mainly on a *per use case basis*. The following chapters comment on the function execution location choice in the architecture (e.g. Section 4.2.4 for Dynamic Radio Configuration functions, Section 5.1.5.2 for Mobility Robustness Optimisation, Section 5.4 on Coverage and Capacity Optimisation) for the respective discussed SON function.

In addition, as outlined above, also the system-level, operational constraints need to be taken into account for a set of SON functions. This is elaborated further in Chapter 9 (SON operation).

3.5 Business Value

In this section an overview about the business impact of SON shall be given. The analysis will concentrate on the case of macro eNB deployments. Since the achievable benefit of SON depends very much on the specific situation of an operator and the deployment scenario, a universal quantification of the benefit of the different use case is impossible.

Thus this Section will start in Section 3.5.1 with outlining the general economical model of an eNB. In Section 3.5.2 the different types of benefit of SON will be described. Based on this information Sections 3.5.3 to 3.5.5 will provide a more detailed analysis of the expected benefit of selected SON features.

3.5.1 The Economics of eNB Sites

The costs of a mobile network as seen by a mobile operator go far beyond the pure purchase price of the individual network elements. Thus, it is commonly accepted to use the concept of the Total Cost of Ownership (TCO) in the context of telecom infrastructure equipment. The TCO aggregates all costs, which occur over the entire lifetime of technical solution (in this case an eNB) in one single figure. The TCO can be calculated as the sum of three components, which are Capital Expenditures (Capex), Implementational Expenditures (Impex) and Operational Expenditures (Opex).

10

SON for Heterogeneous Networks (HetNet)

Cinzia Sartori, Henning Sanneck, Klaus Pedersen, Johanna Pekonen and Ingo Viering

10.1 Introduction

As introduced in Chapter 2, mobile operators acknowledge that the Heterogeneous Network (‘HetNet’) is an attractive solution, able to cope with the enormous data traffic increase.

In fact migration from HSPA to LTE brings additional improvements in the terms of higher spectral efficiency, but those alone are expected to be insufficient when compared to the traffic growth predictions from the industry. This essentially means that other performance boosters such as macro-site densification, improved receivers, higher order sectorisation and addition of small cells are likely to be needed. As the spectral efficiency per link for LTE is approaching the theoretical Shannon limit, it is postulated that the addition of small cells is amongst the most promising solutions for building improved spectral efficiency per area. Thus, the migration from macro-only networks to multi-layer topology networks, often referred to as Heterogeneous Networks or simply HetNet (see also Figure 10.1), is expected to further accelerate in the future. This chapter mainly focuses on LTE multi-layer networks and outlines some of the supportive mechanisms that help enable easy rollout and operation of such multi-layer deployments. In particular, we consider cases where the different LTE cell types as summarised in Table 10.1 are deployed. Notice that the main characteristics in Table 10.1 shall only be considered as an example, since vendors of course have the freedom to also develop cell types with other power settings and antenna gains as listed in this table. From Table 10.1 it is worth noticing that there are huge differences in the Equivalent Isotropic Radiated Power (EIRP),

LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency, First Edition.
Edited by Seppo Hämäläinen, Henning Sanneck and Cinzia Sartori.

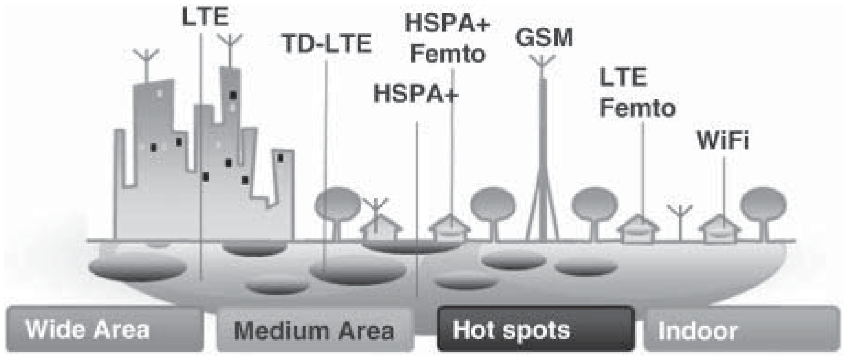

Figure 10.1 Multi-RAT and Multi-Layer network.

which essentially means that the coverage area of each cell type differs significantly, as well as the generated interference to surrounding cells.

The macro, micro and pico eNBs are characterised by the same architecture, while the Home eNBs (HeNB) have different architecture attributes as explained in more detail in Section 10.2. In addition macro eNBs are typically installed by the operator after careful radio network planning considerations, pico eNBs may be installed on an ad hoc basis without prior detailed radio network planning. On the other extreme, HeNBs are typically installed by an end-user who is not a technical expert and they can be moved to a different location and/or switched on or off at any time by the end-user. HeNBs are therefore deployed in an uncoordinated manner without any direct control through the operator. Furthermore, HeNBs may likely be deployed with a restricted access configuration that, amongst others, can results in rather challenging interference scenarios calling for autonomous interference management solutions.

A major precondition for a successful LTE multi-layer deployment is that especially the small eNBs (pico, HeNB) are self-configuring and self-optimising, as conducting configuration and performance management operations manually is simply impossible for the deployment of a large number of small base stations. In Section 10.2 first an overview about the main HeNB characteristics is provided, in order to further set the scene and define this particular small cell type. In Section 10.3 the main self-configuration mechanisms relevant for LTE multi-layer deployments are outlined, while Section 10.4 and 10.5 deal with key self optimisation Sections 10.4 and 10.5, the key self-optimisation techniques for LTE multi-layer deployments such as interference management, mobility optimisation and load balancing. All of the aforementioned techniques are considered to be key enablers for efficient and successful LTE multi-layer deployment and operation.

SON for LTE Relay nodes, SON for Multi-RAT and Energy Saving are not directly addressed in this chapter, since they have already been described in Chapters 4 and 5, respectively.

Table 10.1 Example of approximate main characteristics of different LTE cell types @ 10 MHz bandwidth

Cell type	Transmit power	Antenna gain	EIRP	Range
Macro	46–49 dBm	14 dBi	60–63 dBm	>100 m
Micro	37–40 dBm	5 dBi	42–45 dBm	100 m
Pico	24–37 dBm	4 dBi	28–41 dBm	20–50 m
HeNB	10–20 dBm	0 dBi	10–20 dBm	10–20 m

10.2 Standardisation and Network Architecture

The 3GPP standardisation for LTE has addressed the HetNet deployment scenarios already from the first LTE Release onwards (i.e. 3GPP Release 8). As an example the E-UTRAN base station class specifications in (3GPP TS36.104, 2011) defines pico and HeNB in addition to the macro BS class. Additionally the Closed Subscriber Group (CSG) concept was introduced from the first LTE Release onwards in relation to the HeNB subsystem (see also Section 2.1.8.5).

While pico eNBs can simply be described as small base stations (the typical downlink TX power for a micro eNB is 36 dBm, and for a pico eNB it is 24 dBm, both for 10 MHz bandwidth), having the same architecture as a macro eNB (e.g. the X2 interface is present between macro and pico eNBs and amongst pico eNBs), this is definitely not the case for HeNB. Specifications of the HeNB are created by 3GPP with clear directions given by the Femto Forum (http://www.femtoforum.org).

The Next Generation Mobile Networks (NGMN) Alliance also puts requirements to the multi-layer scenario with 'Interaction between Home and Macro BTS' in the SON session of the (NGMN, 2010). In particular, NGMN addressed the interference management between macro eNB and HeNB.

The 3GPP SON features introduced originally for the macro environment in 3GPP are all applicable for micro and pico eNBs, too. The HeNBs, however, are required to support self-management procedures, which do not necessarily require signalling interactions with the macro, micro or pico eNB neighbours located in the same HetNet environment. Figure 10.2 gives an overview of the HeNB related 3GPP standardisation topics from Release 8 to Release 10 which are explained in more detail in the following.

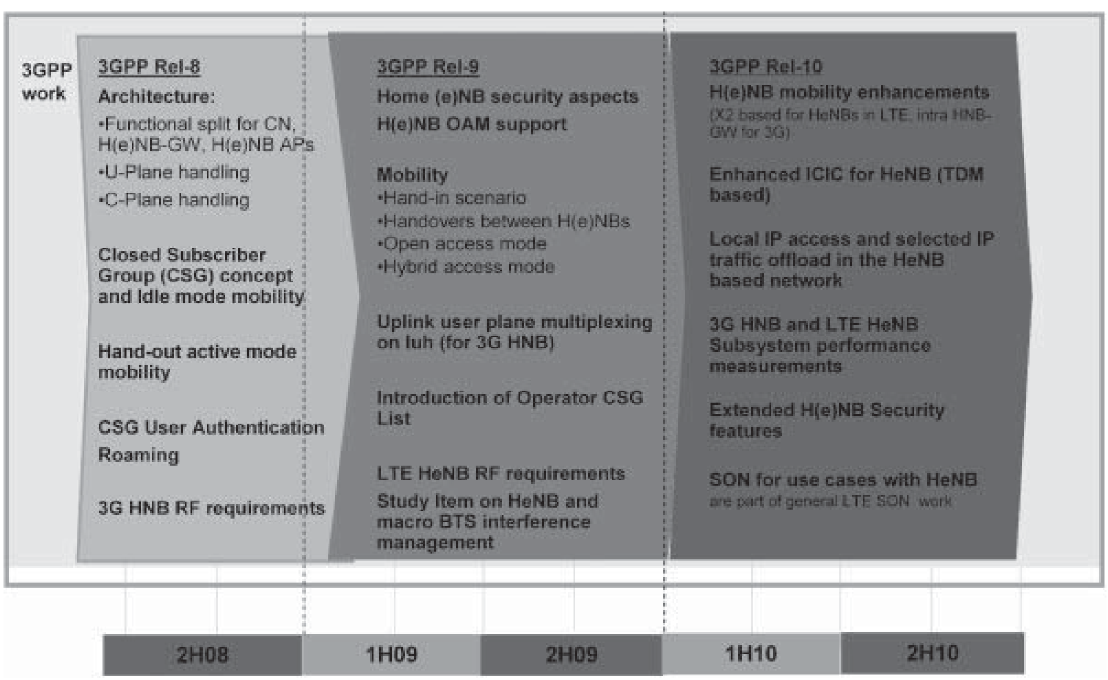

Figure 10.2 HeNB 3GPP Standardisation.

3GPP Release 8

3GPP Release 8 defines the HeNB basic architecture and *closed access mode* concept (which only allows CSG users to be served, see Section 2.1.8.5) focusing especially on the functionalities needed in residential HetNet deployments. HeNB specific functionalities like Network Listening Mode (NLM, which is responsible for acquiring information about the surrounding network environment, see also Section 10.3.2) and Autonomous HeNB power calibration were introduced to support self-configuration capabilities for HeNBs deployed in an uncoordinated manner. Defining the idle mode mobility rules for the HetNet scenarios with CSG cells from the first LTE Release onwards solved the legacy terminal issues, because based on the 3GPP Release 8 definitions any LTE terminal, which has a CSG subscription (i.e. it is allowed to camp on a CSG cell) is aware of CSG specific rules and functionalities.

The active mode mobility (handover) from a HeNB to a macro cell is based on the same principles as the macro layer internal mobility. The support of this *outbound mobility* (from an uncoordinated HeNB to a macro/micro/pico eNB) is a necessary feature for residential scenarios in order to maintain the ongoing call or data connection when the subscriber is moving out from the indoor HeNB coverage to outdoor macro cell coverage. Additionally the CSG subscriptions are supported for the roaming scenarios already from 3GPP Release 8 onwards.

3GPP Release 9

Self-optimisation use cases and supporting procedures were introduced in 3GPP Release 9 for macro eNB scenarios. Those procedures are also applicable to HetNet deployments with macro; small cell or inter-small cell scenarios if an X2 interface is present. The macro eNB–HeNB interference management was addressed with dedicated study items, for FDD mode in (3GPP TR36.922, 2011) and for TDD mode in (3GPP TR36.922, 2011), which finally led to the corresponding specification work in 3GPP Release 10.

3GPP Release 9 furthermore enhanced HeNB subsystem operation and CSG related definitions, including the OAM system (based on recommendations by the Broadband Forum, for example, BBF TR-069, 2010 and BBF TR-196, 2011) and H(e)NB security architecture and requirements.

One of the key features for HetNet environments is the handover support. The active mode *inbound mobility* (from a macro/micro/pico eNB to an uncoordinated HeNB) has also been introduced in 3GPP Release 9. In a typical scenario the high HeNB density may furthermore require Physical Cell Identity (PCI) reuse amongst the HeNBs within macro cell coverage, which may lead to PCI confusion in the serving macro cell when the UE is reporting neighbour cell measurement results; therefore the serving macro cell needs additionally the cell identity of the measured HeNB cell in order to select the correct target for the handover request. Two new access modes for HeNBs were defined: *open access mode* allowing all UEs to use the HeNBs (i.e. the cell appears as any other normal cell to the UE) and the *hybrid access mode* allowing all UEs still to use the cell, but the CSG member UEs with higher priority.

3GPP Release 10

3GPP Release 10 adds further enhancements, especially addressing the enterprise and public environments. The defined features cover further optimisation for the inter-HeNB mobility (over X2 connection), Local IP Access (LIPA) from HeNB, and methods for interference management, of which enhanced Inter-Cell Interference Coordination (eICIC) is certainly the most relevant feature (see Section 10.4).

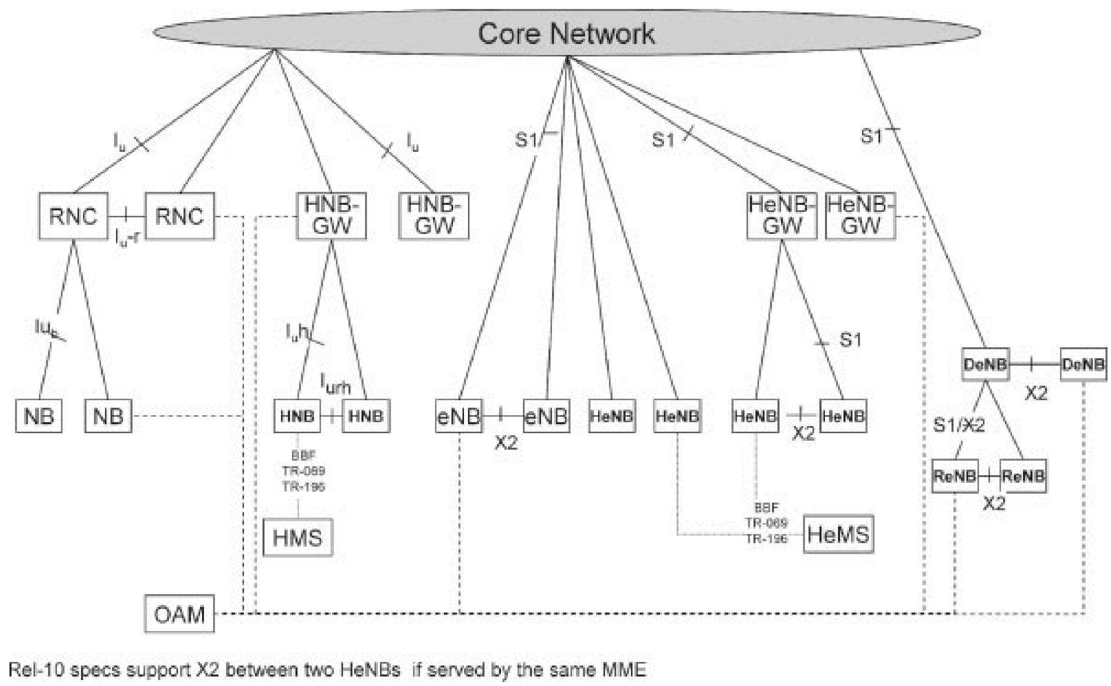

Figure 10.3 Network architecture for Multi-RAT HetNet.

As for macro cells, specifications for HeNBs solutions assure that they can work in real multi-vendor environments with open interfaces between the network elements, and can be fully integrated with other 3GPP based cellular systems. Simultaneously to the LTE HeNB specification work similar standardised concepts have been defined for 3G Home NodeBs (HNB).

The functional split of the LTE HeNB subsystem and the CSG mobility principles are defined in (3GPP TS36.300, 2011). The 3G HNB subsystem Stage 2 descriptions are described in (3GPP TS25.467, 2011).

10.2.1 Network Architecture for HetNet

Figure 10.3 shows a complete multi-RAT Heterogeneous Network architecture which, for completeness, includes 3G and LTE macro (e)NBs, H(e)NBs and LTE Relays.

10.2.1.1 HeNB Deployment Scenarios

The LTE HeNB subsystem architecture, as defined in 3GPP Release 8, allows three different architecture deployment scenarios:

- HeNB with direct S1 interface connection to the Evolved Packet Core (EPC): the HeNB connects to the EPC like a macro eNB and may have S1 connections to multiple Mobility Management Entities (MME)/Serving Gateways (SGW).
- HeNB with EPC connection via a HeNB Gateway (HeNB GW): the HeNB GW serves as a concentrator for the C-Plane in case of high number of HeNBs and may terminate the User

Plane towards the HeNB and towards the Serving GW. The HeNB GW appears towards the MME during S1 setup like an eNB with multiple cells and towards the HeNB like an MME.
- HeNB with EPC connection via a HeNB GW for the control plane (C-Plane) only.

There are some HeNB specific functionalities (e.g. paging optimisation), which are part of HeNB GW. In HeNB GW-less scenario these functionalities are supported by the MME.

The CSG subscription concept is supported in LTE specifications from 3GPP Release 8 onwards. Therefore it has been possible to define the MME to be responsible for the CSG access control in all scenarios. The LIPA support in the HeNB subsystem requires the integration of Local-GW functionality into the HeNB, which is described in (3GPP TS36.300, 2011). The HeNB is also required to support the S5-interface towards the EPC.

10.2.1.2 Residential and Enterprise Scenarios

The HeNB features required for HetNets with enterprise scenarios are different from the HeNB features required for residential scenarios:

- The residential HeNB is typically operated in closed or hybrid access mode, in some seldom cases in open access mode. The handovers with the macro layer are performed via the S1 interface. Inter-HeNB mobility is not foreseen.
- The enterprise HeNB should support handovers with the macro layer (inbound and outbound), and with neighbouring HeNBs. For the inter-HeNB mobility the handover procedures over X2-connection are recommended in order to reduce the signalling load in the MME caused by the enterprise internal handovers. The enterprise HeNBs are typically operated in hybrid or open access mode to ensure coverage also for visitors; however, the closed access mode is also possible.

10.2.1.3 3G HNB Subsystem

In the 3G HNB subsystem concept the presence of legacy systems and terminals requires particular solutions for the CSG subscription control. Hence, the 3G HNB Gateway (HNB GW) has some dedicated functionalities, which are not required for the HeNB GW. The main differences of the 3G HNB subsystem compared to the LTE HeNB subsystem are:

- HNB GW is a mandatory network element of HNB subsystem.
- Support for legacy terminal base, CSG access control related functions in HNB GW.
- Support of Circuit Switched User Plane.
- The Iu based interface between the HNB and HNB GW contains some HNB subsystem specific extensions for HNB and legacy UE registration at HNB GW.

10.3 Self-Configuration

Self-Configuration is a key requirement for small cells. The newly added base station (BS) has to ‘integrate’ itself into the network with minimum human intervention: the BS must be able to boot, connect to the network with proper security credentials, and get the appropriate software

1

Drivers for 5G: The ‘Pervasive Connected World’

Firooz B. Saghezchi,[1] Jonathan Rodriguez,[1] Shahid Mumtaz,[1]
Ayman Radwan,[1] William C. Y. Lee,[2] Bo Ai,[3] Mohammad Tauhidul Islam,[4]
Selim Akl[4] and Abd-Elhamid M. Taha[5]

[1] *Instituto de Telecomunicações, Aveiro, Portugal*
[2] *School of Advanced Communications, Peking University, China*
[3] *State Key Laboratory of Rail Traffic Control and Safety, Beijing, China*
[4] *School of Computing, Queen’s University, Kingston, Ontario, Canada*
[5] *College of Engineering, Alfaisal University, Riyadh, KSA*

1.1 Introduction

We have been witnessing an exponential growth in the amount of traffic carried through mobile networks. According to the Cisco visual networking index [1], mobile data traffic has doubled during 2010–2011; extrapolating this trend for the rest of the decade shows that global mobile traffic will increase 1000x from 2010 to 2020.

The surge in mobile traffic is primarily driven by the proliferation of mobile devices and the accelerated adoption of data-hungry mobile devices – especially smart phones. Table 1.1 provides a list of these devices along with their relative data consumptions. In addition to the increasing adoption rate of these high-end mobile devices, the other important factor associated with the tremendous mobile traffic growth is the increasing demand for advanced multi-media applications such as Ultra-High Definition (UHD) and 3D video as well as augmented reality and immersive experience. Today, mobile video accounts for more than 50% of global mobile data traffic, which is anticipated to rise to two-thirds by 2018 [1]. Finally, social networking has become important for mobile users, introducing new consumption behaviour and a considerable amount of mobile data traffic.

The growth rate of mobile data traffic is much higher than the voice counterpart. Global mobile voice traffic was overtaken by mobile data traffic in 2009, and it is forecast that Voice over IP (VoIP) traffic will represent only 0.4% of all mobile data traffic by 2015. In 2013, the number of mobile subscriptions reached 6.8 billion, corresponding to a global

Fundamentals of 5G Mobile Networks, First Edition. Edited by Jonathan Rodriguez.

Table 1.1 Data consumption of different mobile terminals.

Device	Relative data usage
Feature phone	1x
Smart phone	24x
Handheld gaming console	60x
Tablet	122x
Laptop	515x

penetration of 96%. The ever-growing global subscriber rate spurred on by the world population growth will place stringent new demands on potential 5G networks to cater for one billion new customers.

Apart from 1000x traffic growth, the increasing number of connected devices imposes another challenge on the future mobile network. It is envisaged that in the future connected society, everyone and everything will be inter-connected – under the umbrella of Internet of Everything (IoE) – where tens to hundreds of devices will serve every person. This upcoming 5G cellular infrastructure and its support for Big Data will enable cities to be smart. Data will be generated everywhere by both people and machines, and will be analysed in a real-time fashion to infer useful information, from people's habits and preferences to the traffic condition on the streets, and health monitoring for patients and elderly people. Mobile communications will play a pivotal role in enabling efficient and safe transportation by allowing vehicles to communicate with each other or with a roadside infrastructure to warn or even help the drivers in case of unseen hazards, paving the way towards autonomous self-driving cars. This type of machine-to-machine (M2M) communications requires very stringent latency (less than 1 ms), which imposes further challenges on the future network.

The 1000x mobile traffic growth along with trillions of connected devices is pushing the cellular system to a broadband ubiquitous network with extreme capacity and Energy Efficiency (EE) and diverse Quality of Service (QoS) support. Indeed, it is envisaged that the next-generation cellular system will be the first instance of a truly converged wired and wireless network, providing fibre-like experience for mobile users. This ubiquitous, ultra-broadband, and ultra-low latency wireless infrastructure will connect the society and drive the future economy.

1.2 Historical Trend of Wireless Communications

A new generation of cellular system appears every 10 years or so, with the latest generation (4G) being introduced in 2011. Following this trend, the 5G cellular system is expected to be standardised and deployed by the early 2020s. The standardisation of the new air interfaces for 5G is expected to gain momentum after the International Telecommunication Union-Radiocommunication Sector's (ITU-R) meeting at the next World Radiocommunication Conference (WRC), to be held in 2015. Table 1.2 summarises the rollout year as well as the International Mobile Telecommunications (IMT) requirements for the peak and the average data rates for different generations of the cellular system. Although IMT requirements for 5G are yet to be defined, the common consensus from academic researchers and industry is that in principle it should deliver a fibre-like mobile Internet experience with peak rates of up to 10 Gbps in static/low mobility conditions, and 1 Gbps blanket coverage for highly mobile/cell edge users (with speeds of > 300 km/h). The round-trip time latency of the state-of-the-art 4G

Table 1.2 Specifications of different generations of cellular systems.

Generation	Rollout year	IMT requirement for data rate	
		Mobile users	Stationary users
1G	1981	–	–
2G	1992	–	–
3G	2001	384 Kbps	>2 Mbps
4G	2011	100 Mbps	1 Gbps
5G	2021	1 Gbps	10 Gbps

system (Long-Term Evolution – Advanced; LTE-A) is around 20 ms, which is expected to diminish to less than 1 ms for 5G.

Global standards are a fundamental cornerstone in reaching ubiquitous connectivity, ensuring worldwide interoperability, enabling multi-vendor harmonisation and economies of scale. ITU-R is responsible for defining IMT specifications for next-generation cellular systems. Having defined two previous specifications (IMT-2000 for 3G and IMT-Advanced for 4G), it has already commenced activities towards defining specifications for 5G, which is aimed for completion around 2015. ITU-R arranges WRCs every three to four years to review and revise radio regulations. Allocation of new spectrum for mobile communications is already on the agenda of the next WRC, to be held in November 2015.

To understand where we want to be in terms of 5G, it is worthwhile to appreciate where it all started and to mark where we are now. The following provides a roadmap of the evolution towards 5G communications:

- **Before 1G (<1983):** All the wireless communications were voice-centric and used analogue systems with single-side-band (SSB) modulation.
- **1G (1983–):** All the wireless communications were voice-centric. In 1966, Bell Labs had made a decision to adopt analogue systems for a high-capacity mobile system, because at that time the digital radio systems were very expensive to manufacture. An analogue system with FM radios was chosen. In 1983, the US cellular system was named AMPS (Advanced Mobile Phone Service). AMPS was called 1G at the time.
- **2G (1990–):** During this period, all the wireless communications were voice-centric. European GSM and North America IS-54 were digital systems using TDMA multiplexing. Since AT&T was divested in 1980, no research institute like Bell Labs could develop an outstanding 2G system as it did for the 1G system in North America. IS-54 was not a desirable system and was abandoned. Then, GSM was named 2G at the time when 3G was defined by ITU in 1997. Thus, we could say that moving from 1G to 2G means migrating from the analogue system to the digital system.
- **2.5G (1995–):** All the wireless communications are mainly for high-capacity voice with limited data service. The CDMA (code division multiple access) system using 1.25 MHz bandwidth was adopted in the United States. At the same time, European countries enhanced GSM to GPRS and EDGE systems.
- **3G (1999–):** In this generation, the wireless communications platform has voice and data capability. 3G is the first international standard system released from ITU, in contrast to previous generation systems. 3G exploits WCDMA (Wideband Code Division Multiple

Access) technology using 5 MHz bandwidth. It operates in both frequency division duplex (FDD) and time division duplex (TDD) modes. Thus, we could say that by migrating from 2G to 3G systems we have evolved from voice-centric systems to data-centric systems.

- **4G (2013–):** 4G is a high-speed data rate plus voice system. There are two 4G systems. The United States has developed the WiMAX (Worldwide Interoperability for Microwave Access) system using orthogonal frequency-division multiplexing (OFDM), evolving from WiFi. The other is the LTE system that was developed after WiMAX. The technology of LTE and that of WiMAX are very similar. The bandwidth of both systems is 20 MHz. The major cellular operators are favourable to LTE, and most countries around the world have already started issuing licences for 4G using current developed LTE systems. The cost of licensing through auction is very high. Thus, we could say that migrating from 3G to 4G means a shift from low data rates for Internet to high-speed data rates for mobile video.
- **5G (2021–):** 5G is still to be defined officially by standardisation bodies. It will be a system of super high-capacity and ultra-high-speed data with new design requirements tailored towards energy elicited systems and reduced operational expenditure for operators. In this context, 5G envisages not only one invented technology, but a technology ecosystem of wireless networks working in synergy to provide a seamless communication medium to the end user. Thus, we can say that moving from 4G to 5G means a shift in design paradigm from a single-discipline system to a multi-discipline system.

1.3 Evolution of LTE Technology to Beyond 4G

A summary of IMT-Advanced requirements for 4G is as follows:

- Peak data rate of 100 Mbps for high mobility (up to 360 km/h) and 1 Gbps for stationary or pedestrian users.
- User-plane latency of less than 10 ms (single-way UL/DL (uplink/downlink) delay).
- Scalable bandwidth up to 40 MHz, extendable to 100 MHz.
- Downlink peak spectral efficiency (SE) of 15 bit/s/Hz.
- Uplink peak SE of 6.75 bit/s/Hz.

Paving the way to 5G entails both evolutionary and revolutionary system design. While disruptive radio access technologies (RATs) are needed to provide a step up to the next level of performance capability, we also need to improve the existing RATs. In this regard, we need to further improve the LTE system to beyond 4G (B4G). First targeting the IMT-Advanced requirements, LTE standard Release (R)-8 was unable to fulfil the requirements in the downlink direction (although it could meet all the requirements in the uplink direction) with a single antenna element at the User Equipment (UE) and four receive antennas at the Evolved Node B (eNB) [2]. In contrast, LTE-A is a true 4G technology (meeting all the IMT-Advanced requirements), requiring at least two antenna elements at the UE. As such, it was accepted as IMT-Advanced 4G technology in November 2010 [3]. Figure 1.1 illustrates the evolution of the LTE standard by the 3rd Generation Partnership Project (3GPP) to B4G. The innovations on this roadmap mainly include improving the SE and the area capacity while reducing the network operational cost to ensure fixed marginal cost for the operators. Finally, Table 1.3 summarises the main features of different Releases of LTE from R-8 to R-13, the latest one revealed in December 2013.

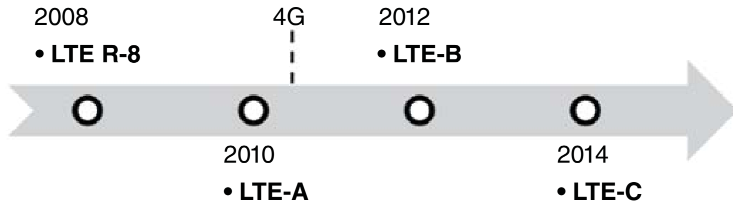

Figure 1.1 Evolution of LTE standard to beyond 4G.

Table 1.3 Main features of different LTE Releases.

Release		Features
LTE	R-8	• Supporting both frequency division duplex (FDD) and time division duplex (TDD) • Scalable frequency spectrum in six different bandwidths: 1.4, 3, 5, 10, 15 and 20 MHz • OFDM • Supporting up to four-layer spatial multiplexing with Single-User Multiple-Input Multiple-Output (SU-MIMO) • Achieving 300 Mbps in DL and 75 Mbps in UL • User-plane latency of less than 20 ms
	R-9	• Multicast and broadcast functionality
LTE-A	R-10	• Carrier aggregation to utilise up to 100 MHz bandwidth • Supporting up to eight-layer spatial multiplexing with SU-MIMO • Enhanced Multi-User (MU-)MIMO • Extended and more flexible reference signal • Relaying functionality • Peak data rate beyond 1 Gbps in DL and 500 Mbps in UL • User-plane latency of less than 10 ms
	R-11	• Coordinated multipoint (CoMP) transmission and reception • Enhanced support for Heterogeneous Network (HetNet)
LTE-B	R-12	• Local area enhancement (soft cell) • Lean carrier • Beamforming enhancement • Enhanced machine-type communication (MTC) • 3D-MIMO • Enhanced CoMP • Enhanced self-organising networks (eSON)
	R-13	• Radio Access Network (RAN) sharing enhancement

1.4 5G Roadmap

Figure 1.2 illustrates the roadmap for 5G [4]. We are in the early research stage for prototyping now. New spectrum is expected to be agreed upon in the WRC 2015, enabling IMT to define the requirements. This will be followed by the standardisation activities and the product development phase until 2020. It is expected that the first wave of 5G networks will be operational around 2021.

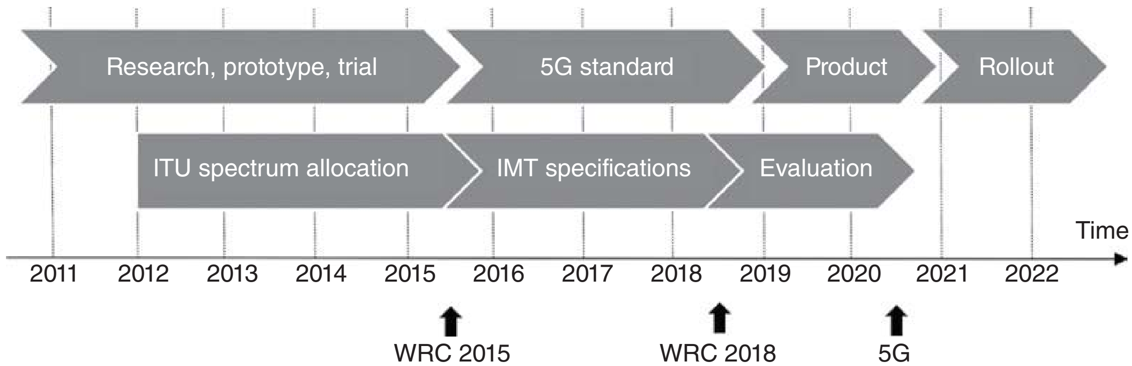

Figure 1.2 Roadmap for 5G.

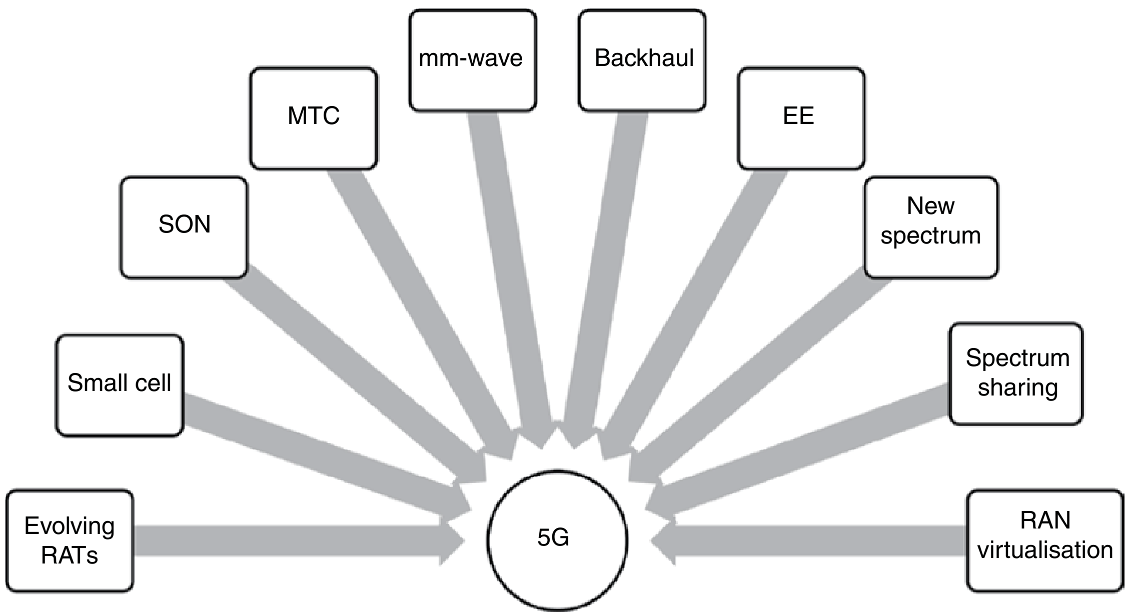

Figure 1.3 10 pillars of 5G.

1.5 10 Pillars of 5G

We identify 10 key building blocks for 5G, illustrated by Figure 1.3. In the following, we elaborate each of these blocks and highlight their role and importance for achieving 5G.

1.5.1 Evolution of Existing RATs

As mentioned before, 5G will hardly be a specific RAT, rather it is likely that it will be a collection of RATs including the evolution of the existing ones complemented with novel revolutionary designs. As such, the first and the most economical solution to address the 1000x capacity crunch is the improvement of the existing RATs in terms of SE, EE and latency, as well as supporting flexible RAN sharing among multiple vendors. Specifically, LTE needs to evolve to support massive/3D MIMO to further exploit the spatial degree of

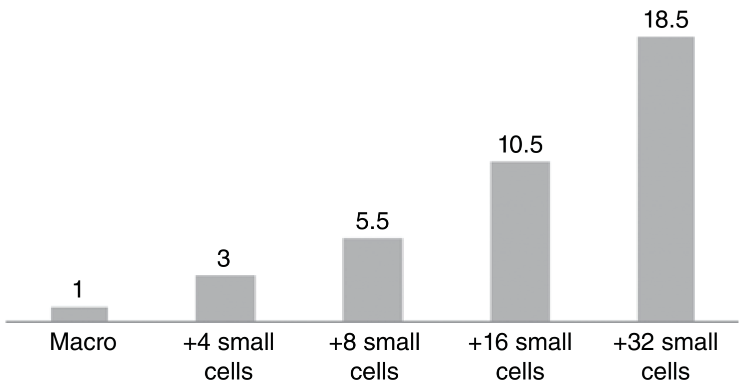

Figure 1.4 Capacity scales linearly with the number of added small cells.

freedom (DOF) through advanced multi-user beamforming, to further enhance interference cancellation and interference coordination capabilities in a hyperdense small-cell deployment scenario. WiFi also needs to evolve to better exploit the available unlicensed spectrum. IEEE 802.11ac, the latest evolution of the WiFi technology, can provide broadband wireless pipes with multi-Gbps data rates. It uses wider bandwidth of up to 160 MHz at the less polluted 5 GHz ISM band, employing up to 256 Quadrature Amplitude Modulation (QAM). It can also support simultaneous transmissions up to four streams using multi-user MIMO technique. The incorporated beamforming technique has boosted the coverage by several orders of magnitude, compared to its predecessor (IEEE 802.11n). Finally, major telecom companies such as Qualcomm have recently been working on developing LTE in the unlicensed spectrum as well as integrating 3G/4G/WiFi transceivers into a single multi-mode base station (BS) unit. In this regard, it is envisioned that the future UE will be intelligent enough to select the best interface to connect to the RAN based on the QoS requirements of the running application.

1.5.2 Hyperdense Small-Cell Deployment

Hyperdense small-cell deployment is another promising solution to meet the 1000x capacity crunch, while bringing additional EE to the system as well. This innovative solution, also referred to as HetNet, can help to significantly enhance the area spectral efficiency (b/s/Hz/m^2). In general, there are two different ways to realise HetNet: (i) overlaying a cellular system with small cells of the same technology, that is, with micro-, pico-, or femtocells; (ii) overlaying with small cells of different technologies in contrast to just the cellular one (e.g. High Speed Packet Access (HSPA), LTE, WiFi, and so on). The former is called multi-tier HetNet, while the latter is referred to as multi-RAT HetNet.

Qualcomm, a leading company in addressing 1000x capacity challenge through hyperdense small-cell deployments, has demonstrated that adding small cells can scale the capacity of the network almost in a linear fashion, as illustrated by Figure 1.4 [5]. That is, the capacity doubles every time we double the number of small cells. However, reducing the cell size increases the inter-cell interference and the required control signalling. To overcome this drawback, advanced inter-cell interference management techniques are needed at the system level along with complementary interference cancellation techniques at the UEs. Small-cell enhancement was the focal point of LTE R-12, where the New Carrier Type (NCT) (also known as the Lean Carrier) was introduced to assist small cells by the host macro-cell. This allows more efficient control plane functioning (e.g. for mobility management, synchronisation, resource allocation, etc.) through the macro-layer while providing a high-capacity and spectrally efficient data plane through the small cells [6]. Finally, reducing the cell size can also improve the EE of the network by bringing the network closer to the UEs and hence shrinking the power budget of the wireless links.

1.5.3 Self-Organising Network

Self-Organising Network (SON) capability is another key component of 5G. As the population of the small cells increases, SON gains more momentum. Almost 80% of the wireless traffic is generated indoors. To carry this huge traffic, we need hyperdense small-cell deployments in homes – installed and maintained mainly by the users – out of the control of the operators. These indoor small cells need to be self-configurable and installed in a plug and play manner. Furthermore, they need to have SON capability to intelligently adapt themselves to the neighbouring small cells to minimise inter-cell interference. For example, a small cell can do this by autonomously synchronising with the network and cleverly adjusting its radio coverage.

1.5.4 Machine Type Communication

Apart from people, connecting mobile machines is another fundamental aspect of 5G. Machine type communication (MTC) is an emerging application where either one or both of the end users of the communication session involve machines. MTC imposes two main challenges on the network. First, the number of devices that need to be connected is tremendously large. Ericsson (one of the leading companies in exploring 5G) foresees that 50 billion devices need to be connected in the future networked society; the company envisages 'anything that can benefit from being connected will be connected' [7]. The other challenge imposed by MTC is the accelerating demand for real-time and remote control of mobile devices (such as vehicles) through the network. This requires an extremely low latency of less than a millisecond, so-called "tactile Internet" [8], dictating 20x latency improvement from 4G to 5G.

1.5.5 Developing Millimetre-Wave RATs

The traditional sub-3 GHz spectrum is becoming increasingly congested and the present RATs are approaching Shannon's capacity limit. As such, research on exploring cm- and mmWave bands for mobile communications has already been started. Although the research on this field is still in its infancy, the results look promising.

There are three main impediments for mmWave mobile communications. First, the path loss is relatively higher at these bands, compared to the conventional sub-3GHz bands. Second, electromagnetic waves tend to propagate in the Line-Of-Sight (LOS) direction, rendering the radio links vulnerable to being blocked by moving objects or people. Last but not least, the penetration loss through the buildings is substantially higher at these bands, blocking the outdoor RATs for the indoor users.

Despite these limitations, there are myriad advantages for mmWave communications. An enormous amount of spectrum is available in mmWave band; for example, at 60 GHz, there is

9GHz of unlicensed spectrum available. This amount of spectrum is huge, especially when we think that the global allocated spectrum for all cellular technologies hardly exceeds 780 MHz [9]. This amount of spectrum can completely revolutionise mobile communications by providing ultra-broadband wireless pipes that can seamlessly glue the wired and the wireless networks. Other advantages of mmWave communications include the small antenna sizes ($\lambda/2$) and their small separations (also around $\lambda/2$), enabling tens of antenna elements to be packed in just one square centimetre. This in turn allows us to achieve very high beamforming gains in relatively small areas, which can be implemented at both the BS and the UE. Incorporating smart phased array antennas, we can fully exploit the spatial degree of freedom of the wireless channel (using Space-Division Multiple Access (SDMA)), which can further improve the system capacity. Finally, as the mobile station moves around, beamforming weights can be adjusted adaptively so that the antenna beam is always pointing to the BS.

Recently, Samsung Electronics, an industry leader in exploring mmWave bands for mobile communications, has tested a technology that can achieve 2 Gbps data rate with 1 km range in an urban environment [10]. Furthermore, Professor Theodore Rappaport and his research team at the Polytechnic Institute of New York University have demonstrated that mobile communications at 28 GHz in a dense urban environment such as Manhattan, NY, is feasible with a cell size of 200 m using two 25 dBi antennas, one at the BS and the other at the UE, which is readily achievable using array antennas and the beamforming technique [9].

Last but not least, foliage loss for mmWaves is significant and may limit the propagation. Furthermore, mmWave transmissions may also experience significant attenuations in the presence of a heavy rain since the raindrops are roughly the same size as the radio wavelengths (millimetres) and therefore can cause scattering. Therefore, a backup cellular system operating in legacy sub-3 GHz bands might be needed as part of the mmWave solution [9].

1.5.6 Redesigning Backhaul Links

Redesigning the backhaul links is the next critical issue of 5G. In parallel to improving the RAN, backhaul links also need to be reengineered to carry the tremendous amount of user traffic generated in the cells. Otherwise, the backhaul links will soon become bottlenecks, threatening the proper operation of the whole system. The problem gains more momentum as the population of small cells increases. Different communication mediums can be considered, including optical fibre, microwave and mmWave. In particular, mmWave point-to-point links exploiting array antennas with very sharp beams can be considered for reliable self-backhauling without interfering with other cells or with the access links.

1.5.7 Energy Efficiency

EE will remain an important design issue while developing 5G. Today, Information and Communication Technology (ICT) consumes as much as 5% of the electricity produced around the globe and is responsible for approximately 2% of global greenhouse gas emissions – roughly equivalent to the emissions created by the aviation industry. What concerns more is the fact that if we do not take any measure to reduce the carbon emissions, the contribution is projected to double by 2020 [11]. Hence, it is necessary to pursue energy-efficient design approaches from RAN and backhaul links to the UEs.

The benefit of energy-efficient system design is manifold. First, it can play an important role in sustainable development by reducing the carbon footprint of the mobile industry itself. Second, ICT as the core enabling technology of the future smart cities can also play a fundamental role in reducing the carbon footprint of other sectors (e.g. transportation). Third, it can increase the revenue of mobile operators by reducing their operational expenditure (Opex) through saving on their electricity bills. Fourth, reducing the 'Joule per bit' cost can keep mobile services affordable for the users, allowing flat rate pricing in spite of the 10 to 100x data rate improvement expected by 2020. Last but not least, it can extend the battery life of the UEs, which has been identified by the market research company TNS [12] as the number one criterion of the majority of the consumers purchasing a mobile phone.

1.5.8 Allocation of New Spectrum for 5G

Another critical issue of 5G is the allocation of new spectrum to fuel wireless communications in the next decade. The 1000x traffic surge can hardly be managed by only improving the spectral efficiency or by hyper-densification. In fact, the leading telecom companies such as Qualcomm and NSN believe that apart from technology innovations, 10 times more spectrum is needed to meet the demand. The allocation of around 100 MHz bandwidth at the 700 MHz band and another 400 MHz bandwidth at around 3.6 GHz, as well as the potential allocation of several GHz bandwidths in cm- or mmWave bands to 5G will be the focal point of the next WRC conference, organised by ITU-R in 2015.

1.5.9 Spectrum Sharing

Regulatory process for new spectrum allocation is often very time consuming, so the efficient use of available spectrum is always of critical importance. Innovative spectrum allocation models (different from the traditional licensed or unlicensed allocation) can be adopted to overcome the existing regulatory limitations. Plenty of radio spectrum has traditionally been allocated for military radars where the spectrum is not fully utilised all the time (24/7) or in the entire geographic region. On the other hand, spectrum cleaning is very difficult as some spectrum can never be cleaned or can only be cleaned over a very long time; beyond that, the spectrum can be cleaned in some places but not in the entire nation. As such, the Authorised/Licensed Shared Access (ASA/LSA) model has been proposed by Qualcomm to exploit the spectrum in small cells (with limited coverage) without interfering with the incumbent user (e.g. military radars) [13]. This kind of spectrum allocation model can compensate the very slow process of spectrum cleaning. It is also worth mentioning that as mobile traffic growth accelerates, spectrum refarming becomes important, to clean a previously allocated spectrum and make it available for 5G. Cognitive Radio concepts can also be revisited to jointly utilise licensed and unlicensed spectrums. Finally, new spectrum sharing models might be needed as multi-tenant network operation becomes widespread.

1.5.10 RAN Virtualisation

The last but not least critical enabler of 5G is the virtualisation of the RAN, allowing sharing of wireless infrastructure among multiple operators. Network virtualisation needs to be pushed from the wired core network (e.g. switches and routers) towards the RAN. For network virtualisation, the intelligence needs to be taken out of the RAN hardware and controlled in a centralised manner using a software brain, which can be done in different network layers. Network virtualisation can bring myriad advantages to the wireless domain, including both Capex (Capital Expenditure) and Opex savings through multi-tenant network and equipment sharing, improved EE, on-demand up- or down-scaling of the required resources, and increased network agility through the reduction of the time-to-the-market for innovative services (from 90 hours to 90 minutes), as well as easy maintenance and fast troubleshooting through increased transparency of the network [14]. Virtualisation can also serve to converge the wired and the wireless networks by jointly managing the whole network from a central orchestration unit, further enhancing the efficiency of the network. Finally, multi-mode RANs supporting 3G, 4G or WiFi can be adopted where different radio interfaces can be turned on or off through the central software control unit to improve the EE or the Quality of Experience (QoE) for the end users.

1.6 5G in Europe

Past research efforts in Europe have delivered many advances in mobile communications we take for granted today. These include the 2G GSM standard (used today by 80% of the world’s mobile networks) and the technologies used in the 3G Universal Mobile Telecommunications System (UMTS) and the 4G LTE standards. Timely development of the 5G technology is now of paramount importance for Europe to drive the economy, strengthen the industry’s competitiveness, and create new job opportunities.

Leading the development of 5G technology is critically important for the European Union (EU), primarily because of its vital role in economic growth. As a whole, the ICT sector represents approximately 5% of EU GDP, with an annual value of €660 billion. It generates 25% of total business expenditure in Research and Development (R&D), and investments in ICT account for 50% of all European productivity growth.

Second, pioneering 5G is vitally important because this technology will play a key role in securing Europe’s leadership in the global mobile industry. Historically, the European telecom industry was at the forefront of global competition from the early days of GSM technology to the UMTS and LTE technologies. It still represented approximately 40% of the worldwide telecom market of nearly €200 billion in 2012 in terms of network infrastructure supply. However, Europe is now falling behind its competitors and wants to catch up by leading the 5G technology.

Last but not least, leading 5G technology is of great importance for the EU as it can bring new job opportunities to Europe. European Commission Vice President Neelie Kroes announced during the Mobile World Congress 2013 in Barcelona: ‘I want 5G to be pioneered by European industry, based on European research and creating jobs in Europe’.

However, the emergence of new eastern competitors such as China and South Korea may challenge these key ambitions.

1.6.1 Horizon 2020 Framework Programme

Europeans use ‘Framework Programmes’ as financial instruments to coordinate and fund their future research and innovation. They have successfully exercised this model by developing 3G (UMTS) and 4G (LTE) standards; now they intend to use the same model for 5G.

need', said Takehiro. But he insisted that 5G should be a technology that industry should carefully take into consideration.

1.9 5G Architecture

As illustrated by Figure 1.6, 5G will be a truly converged system supporting a wide range of applications from mobile voice and multi-Giga-bit-per-second mobile Internet to D2D and V2X (Vehicle-to-X; X stands for either Vehicle (V2V) or Infrastructure (V2I)) communications, as well as native support for MTC and public safety applications. 3D-MIMO will be incorporated at BSs to further enhance the data rate and the capacity at the macro-cell level. System performance in terms of coverage, capacity and EE will be further enhanced in dead and hot spots using relay stations, hyperdense small-cell deployments or WiFi offloading; directional mmWave links will be exploited for backhauling the relay and/or small-cell BSs. D2D communications will be assisted by the macro-BS, providing the control plane. Smart grid is another interesting application envisaged for 5G, enabling the electricity grid to operate in a more reliable and efficient way. Cloud computing can potentially be applied to the RAN,

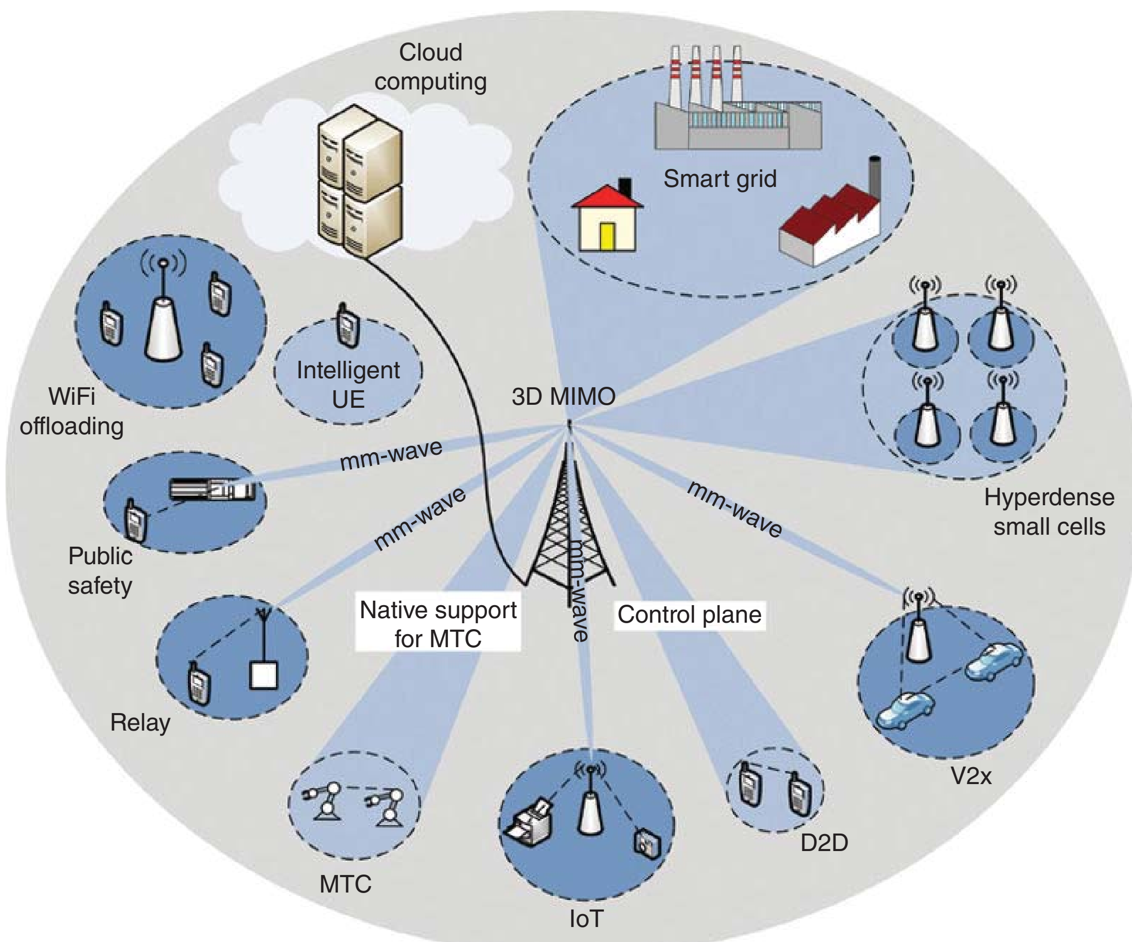

Figure 1.6 5G system architecture.

and beyond that, to mobile users that can form a virtual pool of resources to be managed by the network. Bringing the applications through the cloud closer to the end user reduces the communication latency to support delay-sensitive real-time control applications.

It is envisaged that 5G will seamlessly integrate the existing RATs (e.g. GSM, HSPA, LTE and WiFi) with the complementary new ones invented in mmWave bands. MmWave technology will revolutionise the mobile industry not only because of plenty of available spectrum at this band (readily allowing Gbps wireless pipes), but also because of diminishing antenna sizes, enabling the fabrication of array antennas with hundreds or thousands of antenna elements, even at the UE. Smart antennas with beamforming and phased array capabilities will be employed to point out the antenna beam to a desired location with high precision, rotated electronically through phase shifting. The narrow pencil beams will enable the exploiting of the spatial DOF, without interfering with other users. The small antenna sizes will enable Massive/3D MIMO at BSs and eventually at UEs. The mmWave technology will also provide ultra-broadband backhaul links to carry the traffic from/to either the small BSs or the relay stations, allowing further deployment flexibility for the operators, compared to the wired (copper or fibre) backhaul link. Hyperdense small-cell deployment is another promising solution for 5G to meet the 1000x capacity challenge. Small cells have the potential to provide massive capacity and to minimise the physical distance between the BS and the UEs to achieve the required EE enhancement for 5G. The traditional sub-3 GHz bands will be employed for macro-cell blanket coverage, while the higher frequency bands (e.g. cm- and mmWave bands) will be employed for small cells to provide a spectral- and energy-efficient data plane, assisted by a control plane served by the macro-BS [38].

Along with the development of new RATs and the deployment of hyperdense small cells, the existing RATs will continue to evolve to provide higher SE and EE. The data plane latency (round-trip time) of the LTE-A system is around 20 ms, which is expected to be reduced to less than 1 ms in its future evolutions [30]. Moreover, the SE of the existing HSPA system is 1 b/s/Hz/cell, which is expected to increase 10x by 2020 [30]. The EE of the cellular system is expected to improve 1000x by 2015, compared to the 2010 level [39]. The PHY (physical) and MAC (medium access control) layer techniques will be revisited for carrying short and delay-sensitive packets for MTCs [18]. Virtualisation will also play a key role in 5G for efficient resource utilisation in cellular systems, through a multi-tenant network where a mobile operator will not need to own a complete set of dedicated network equipment; rather, network equipment (e.g. BS) will be shared among different operators. The existing cloud network concept mainly involves the data centres. Mobile network virtualisation will push this concept towards the backhaul and the RAN to allow sharing of backhaul links and BSs among different operators. Last but not least, it is envisaged that 5G UEs will be multi-mode intelligent devices. These UEs will be smart enough to autonomously choose the right interface to connect to the network based on the channel quality, its remaining battery power, the EE of different RANs, and the QoS requirement of the running application. These smart and efficient 5G UEs will be able to support 3D media with speeds up to 10 Gbps.

1.10 Conclusion

5G is expected to be deployed around 2020, providing pervasive connectivity with 'fibre-like' experience for mobile users. Apart from the expected 10 Gbps peak data rate, the major challenge for 5G is the massive number of connected machines and the 1000x growth in mobile traffic. The ultra-broadband and green cellular system will be the driving engine for the future connected society where anyone and anything will be connected at anytime and anywhere. In this chapter, we gave an overview of the potential enablers of 5G along with research and development activities around the globe, including Europe, North America and the Asia-Pacific region. Being in the prototype stage, standardisation is the next milestone to achieving 5G, which will be followed by the development phase for two to three years. The last phase is network deployment and marketing, which may take another couple of years, foreseeing a potential commercial deployment by around 2020. In the final section of this chapter, we illustrated the foreseen architecture for 5G, harnessing all the common views on the current technology trends and the emerging applications. In a nutshell, mmWave technology, hyper-dense HetNet, RAN virtualisation and massive MTC are all major breakthroughs being considered for upgrading the cellular system to achieve 5G capability. However, these technology developments need to be fuelled by the allocation of new spectrum for mobile communications, expected to happen in the upcoming WRC meeting.

Acknowledgements

The authors would like to acknowledge the ECOOP project (sponsored by the Instituto de Telecomunicações/FCT – PEst-E/EEI/LA0008/2013), which has provided valuable inputs to the compilation of this chapter. Firooz B. Saghezchi would also like to acknowledge his PhD grant funded by the Fundação para a Ciência e a Tecnologia (FCT-Portugal) with reference number SFRH/BD/79909/2011.

References

[1] Cisco (2015) Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html (last accessed 4 February 2015).
[2] Ghosh, A., Ratasuk, R., Mondal, B. *et al.* (2010) LTE-Advanced: Next-Generation Wireless Broadband Technology [invited paper]. *IEEE Wireless Communications*, 17(3), 10–22.
[3] Liu, L., Chen, R., Geirhofer, S. *et al.* (2012) Downlink MIMO in LTE-Advanced: SU-MIMO vs. MU-MIMO. *IEEE Communications Magazine*, 50(2), 140–147.
[4] Huawei (2013) 5G: A Technology Vision. http://www.huawei.com/5gwhitepaper/ (last accessed 19 December 2014).
[5] Qualcomm (2014) 1000x: More Small Cells – Hyper-Dense Small Cell Deployments. https://www.qualcomm.com/documents/1000x-more-small-cells (last accessed 4 February 2015).
[6] Hoymann, C., Larsson, D., Koorapaty, H. and Cheng, J. F. (2013) A Lean Carrier for LTE. *IEEE Communications Magazine*, 51(2), 74–80.
[7] Ericsson (2013) 5G Radio Access – Research and Vision. White paper. http://www.ericsson.com/news/130625-5g-radio-access-research-and-vision_244129228_c (last accessed 19 December 2014).
[8] Fettweis, G. and Alamouti, S. (2014) 5G: Personal Mobile Internet beyond What Cellular Did to Telephony. *IEEE Communications Magazine*, 52(2), 140–145
[9] Rappaport, T. S., Sun, S., Mayzus, R. *et al.* (2013) Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access*, 1, 335–349.
[10] Pi, Z. and Khan, F. (2011) An Introduction to Millimeter-Wave Mobile Broadband Systems. *IEEE Communications Magazine*, 49(6), 101–107.
[11] Saghezchi, F. B., Radwan, A., Rodriguez, J. and Dagiuklas T. (2013) Coalition Formation Game toward Green Mobile Terminals in Heterogeneous Wireless Networks. *IEEE Wireless Communications*, 20(5), 85–91.