Fundamentals of Information and Communication Technologies

Shun-Ping Chen

Fundamentals of Information and Communication Technologies

Fundamentals of Information and Communication Technologies

By

Shun-Ping Chen

Cambridge Scholars Publishing

Fundamentals of Information and Communication Technologies

By Shun-Ping Chen

This book first published 2020

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-5585-2
ISBN (13): 978-1-5275-5585-3

CONTENTS

Preface vii

Chapter 1 Introduction 1

Chapter 2 Information Theory 8
2.1 Shannon Capacity.. 8
2.2 Multimedia Data.. 12
2.3 Data Processing, Boolean Logics 18
2.4 Information Content, Entropy 21

Chapter 3 Source and Channel Coding 26
3.1 Source Coding... 26
3.2 Channel Coding... 29

Chapter 4 Modulation Schemes 39

Chapter 5 Internet 46

Chapter 6 Electromagnetic Waves 51

Chapter 7 Wave Propagation 61
7.1 Elementary Antennas 62
7.2 Wave Propagation Models 64
7.3 Antenna Array and Beam Focusing 72
7.4 MIMO and Beam Forming... 84

Chapter 8 Wireless Technologies 88
8.1 WLAN, WiFi 88
8.2 Bluetooth... 94
8.3 Other Wireless PAN and WAN Technologies 97
8.4 Satellite Communications 100
8.5 Broadcast Services 107

Chapter 9 Cellular Mobile Networks 111
9.1 GSM (2G), UMTS (3G)...113
9.2 LTE (4G)..119
9.3 5G Mobile Networks..121
9.4 Mobile Network Planning Aspects..128

Chapter 10 Wirelines and Waveguides 137
10.1 Copper Wire 137
10.2 Coaxial Cable 138
10.3 Optical Fibre 139
10.4 FTTC, FTTH and FTTBS 140

Chapter 11 Optical Fibre Communications 142
11.1 Optical Fibre 142
11.2 Attenuation 144
11.3 Modal Dispersion 146
11.4 Chromatic Dispersion 147
11.5 Polarisation Modal Dispersion 153
11.6 Nonlinearities 158
11.7 Laser Diodes as Transmitter 159
11.8 Photodiodes as Receiver 161
11.9 DWDM and CWDM 163

Chapter 12 Free Space Optical Communications 165
12.1. Background 165
12.2. Free Space Optical Link 166
12.3. Channel Model with Different Factors 166
12.4. Deep Space Optical Communications 170

Chapter 13 Network Security 174
13.1 Symmetrical Encryption 188
13.2 Asymmetrical Encryption 193
13.3 Authentication 196
13.4 Hash-value, integrity check 196

Chapter 14 Network Management 197
14.1 Telecommunications Management Network 197
14.2 SNMP 198
14.3. Functionalities of Network Management 201

Chapter 15 Trends and Future Developments 203

References 206
Figures 214
Tables 217
List of Acronyms 218

PREFACE

Information and Communication Technologies (ICT) have developed rapidly in the last decades, with performance improved significantly by using sophisticated software and hardware techniques, evolving from analogue technologies to efficient, reliable and inexpensive digital technologies.

The goals of the information and communication technologies are to efficiently process the multimedia and other data information and to reliably transmit them from the information source to the destination, or technically speaking, from the transmitter to the receiver. This overall end-to-end process involves many disciplines, like multimedia source coding, channel coding, modulation and wireline and wireless transmission. Each of these disciplines have been treated in numerous textbooks and taught in numerous courses at universities.

The author has been teaching various Bachelor and Masters courses at the Darmstadt University of Applied Sciences (Hochschule Darmstadt) in Germany since 2008. He is very lucky to be involved in many lectures directly related to the above-mentioned information and communication technology fields, leading to the idea to write a textbook, which does not primarily go into too much detail on all of these exciting fields, but rather provides an overview of the end-to-end information processing and communication.

This textbook is designed not only as lecture notes for students of Bachelor and Masters programmes at universities, but also for interested researchers and engineers, who would like to get an overview and cover the most fundamental aspects of information and communication technologies and networks, as well as for all other interested people in the scientific community and society. For this purpose the author also tries to describe the technical aspects more comprehensively on one hand, and on the other hand also focuses on his teaching and research areas like antenna arrays and fibre and free space optical communications as well as network security.

The author appreciates many fruitful discussions with colleagues, students and research partners. Especially, he would like to thank his colleague, friend and advisor Professor Heinz Schmiedel for proofreading the manuscript, for many helpful advices and for his continuous support and encouraging discussions.

The author would like to thank Laurence Fenton for careful proofreading and Rebecca Gladders for many helpful advices and support.

CHAPTER 1

INTRODUCTION

Following on from the development of wireline and wireless technologies in the nineteenth century, such as the telegraph and the telephone, the modern information and communication technologies witnessed tremendous advances in the twentieth century, from the inventions of transistors, computers, laser diodes and satellite communications to digital cellular mobile networks, optical fibres and, last but not least, the internet. All are being permanently further developed and continuously improved.

Starting with the basic architecture of the information and communication technologies and networks, a goal is a reliable, high-quality communication link for transmitting the multimedia information data between the source of the information and the destination. To ensure high availability and reliability we want low bit error rates, low blocking probability, low delay and low jitter in the ideal case. The applications could be conversational phone calls, video conferences, internet browsing, video streaming, IP television, multimedia data transfer, etc. In the upcoming 5G mobile communications, new promising applications like eMBB (enhanced Multimedia Broadband), mMTC (massive Machine Type Communications, like sensor applications) and uRLLC (ultra-Reliable Low-Latency Communications) will be enabled.

In daily life, different types of information, like classical text, still images, voice, data, music or motion pictures and video streaming, will be processed and transmitted. This kind of communication is defined as H2H (human-to-human) communication.

Along with the rapid improvement of digitalisation and new applications in industrial branches like autonomous driving, power grids, smart factories, smart homes, intelligent traffic management and the internet of things (IoT), where an extremely large number of sensors will be applied, the measurement and control data must be transmitted reliably and with extremely low delay, especially for so-called mission critical applications. Therefore more and more machine type communications (MTC) will be indispensable. Different types of sensors can detect parameters like temperature, forces, velocity, acceleration, humidity, etc.

Also bio-chemical sensors for detecting the content of certain chemical components, blood glucose, taste, odour, flavour, etc., are now being developed for the healthcare, medicine and chemical industries. These machine type communication technologies provide the possibility for automatic data processing in relevant fields like autonomous driving, health care, traffic management, smart homes, smart grids, smart factories and so on.

The challenging tasks for the information and communication engineers will be to design intelligent systems to convert these primary physical and chemical parameters to electrical information data sources, usually in digital, binary format, which can then be transmitted from the information source to the central information processing unit, and then further to the receivers by utilising sophisticated digital communication networks and technologies.

The transmission can take place via wireline systems (copper wires, optical fibres, coaxial CATV infrastructure and hybrid fibre coaxial HFC as well as the 50/60 Hz power lines that are partially used to transmit digital information in the higher frequency bands) or wireless communication systems (WLAN/WiFi, Bluetooth, ZigBee, RFID, cellular mobile networks like GSM, UMTS, LTE, satellite and future 5G mobile networks).

Driven by the invention of transistors, laser diodes, optical fibres, cellular mobile network concepts and the internet in the twentieth century, traditional analogue communication technologies have been almost completely replaced by digital communication technologies. Therefore in this book we only focus on digital information and communication technologies.

Chapter 2 discusses the fundamentals of information theory.

In Figure 1 the end-to-end digital communication channel model from the information source to the destination, or (respectively) from the transmitter to the receiver, is illustrated schematically. Prior to the source coding, the multimedia information sources for H2H and MTC will be first converted from the analogue information sources (voice, data, images and moving pictures, etc.) to a digital information data stream in terms of binary bits by using an ADC (Analogue Digital Converter).

These digital, binary data streams will be optimised by compressing techniques, where the redundancy and irrelevance will be removed, in order to efficiently utilise the scarce network bandwidth in the transmission networks. This process is called source coding, and will be discussed in Chapter 3.

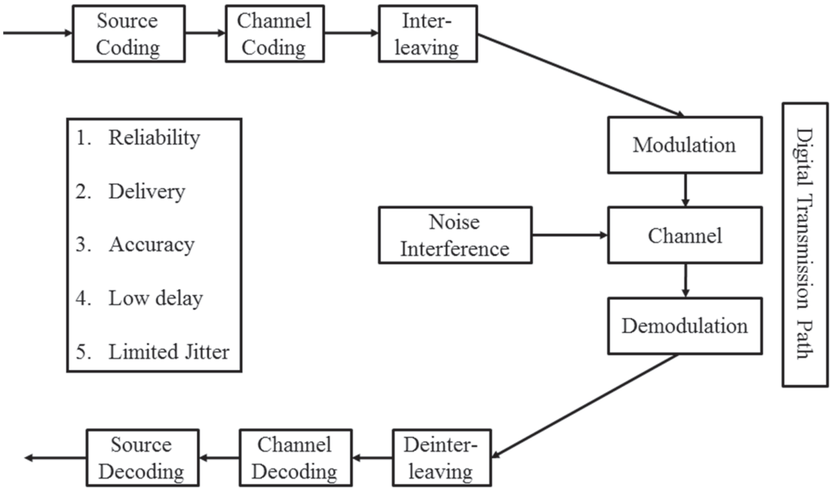


Figure 1 Digital communication channel model

The source coding techniques of the multimedia information sources will be discussed, in order to understand how they achieve the best performance with the best reliability for the limited physical transmission network resources. If necessary, the source-coded information will be further encrypted in order to guarantee confidentiality. The integrity of the data will be checked by using the hash value, which is like the fingerprint of the information content.

In the second step some useful redundant bits and bytes are added, in order to enable FEC (forward error correction). This can be done by sophisticated channel coding schemes like Viterbi coding and convolutional coding. After that, interleaving is used to separate the neighbouring bits and reorder them in a proper manner, in order to separate the burst errors to single bit errors, and thus to make the FEC more efficient.

Source coding reduces the data volume to a minimum, whereas channel coding is used to enable automatic bit error correction, therefore to guarantee the performance and reliability of the transmission networks.

This digital information format will be modulated from the base band to the corresponding carrier frequency for certain physical media like wireline, or free space by using properly defined modulation schemes to achieve high bit rates, with low bit errors and low delay. The frequency could be f = 900 MHz, 2 GHz for the classical cellular mobile communications in the form of electromagnetic wave propagating in the free space, and approximately f = 193.5 THz (λ = 1.55 μm) for the guided wave propagation in optical fibres. During the transmission either in free space or in the wirelines like optical fibres, different mechanisms lead to the gradual reduction of the signal power level from the transmitter to the receiver (attenuation), the spreading of the binary signal pulses in the time domain (dispersion) and the disturbance of the information signal by interferences by other users or other disturbing noises in the same channel or the same frequency band, so that the distance between the transmitter and receiver, and achievable bit rates, will be always limited by the signal power level relative to the noise or interference. At a certain threshold, also called receiver sensitivity, the signal as binary 0s and 1s can hardly be recognised and received anymore without bit errors. Therefore so-called repeaters and regenerators to refresh, regenerate and re-amplify the signal are used to increase the difference between the wanted signal power level and the noise or interference power level. The ratio between them is also defined as SINR (signal to interference and noise ratio). The better the SINR, the better the channel quality will be, and correspondingly a higher throughput will be achieved.

The quality of such a transmission channel is generally characterised by the fundamental technical parameters like reliability, availability, bit error rates, accuracy, delay or jitter (variation of the delay), etc. Besides other network management measures, it is important to reduce the time needed for repairs in case of failures, outages and performance degradation of single network elements, network sections or application servers.

At the receiver site, the information signal will be transformed back to the format of the source step by step. Firstly the binary information bits and bytes are demodulated back to the base band, de-interleaved back to the original order, channel-decoded by performing the FEC and then source-decoded. The last step is to convert the digital data stream to the analogue information. The digital binary bits and bytes will then be converted to the analogue information – i.e. text, language, music, video, measurement data, etc. – using a DAC (Digital Analogue Converter).

Concerning the transmission path, the medium could be free space or air, by utilising the free space propagation of electromagnetic waves (like GSM, UMTS, LTE, Wireless LAN/WiFi, satellite communications, radio, TV and free space optical communications (FSOC)), or by guiding the electromagnetic waves in metallic waveguides or optical fibres consisting of silica quartz glass. For lower frequencies, the classical twisted copper wires, either schielded or unschielded, can be used, for example, in the xDSL (ADSL, HDSL, VDSL) digital subscriber lines.

The governing Maxwell's equations can be used to explain almost all the wave propagation mechanisms, from free space microwave propagation to antenna problems and optical fibre waveguides. This will be discussed in Chapter 4, Chapter 5 and Chapter 6.

Besides the technical impairments like attenuations, thermal noise, interferences and non-linear distortion, security issues also become more and more crucial and important. Since the internet is an open infrastructure enabling communications between people all over the world, malicious persons will misuse this opportunity to eavesdrop, manipulate confidential data, defraud, steal, blackmail and benefit financially from their criminal activities. Countermeasures both for the network operations of enterprises and private persons are most important in order to avoid any loss of confidential data and to defend themselves from any criminal activities. Therefore the network security and network management issues will be important parts of the information and communication technologies and networks, as discussed in Chapter 7.

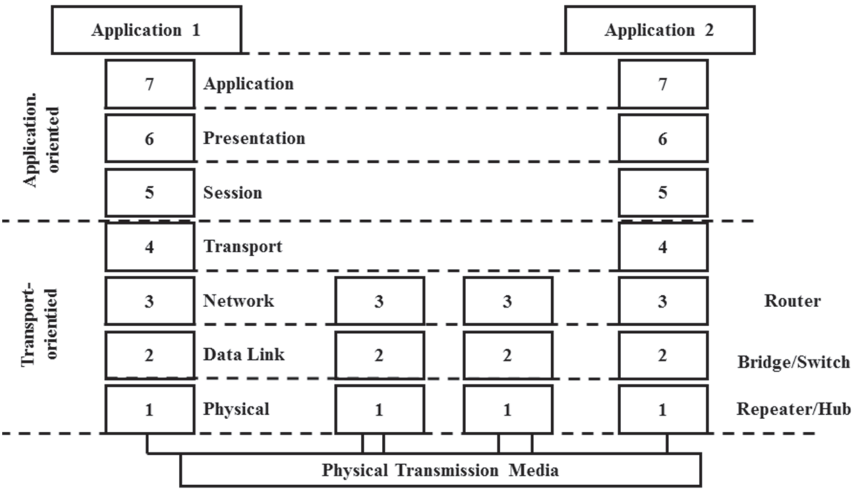


Figure 2 Open system interconnection model OSI

Open System Interconnection OSI model (Fig. 2) defines the communication standards. It defines seven layers to be treated separately and efficiently on the one hand, while on the other hand it allows that systems designed by different system suppliers or vendors, in accordance with the common OSI standards and other relevant technical standards like ITU, ETSI, IEEE and IETF, can be simply interconnected with each other flexibly and reliably. This OSI will be also the foundation for many network aspects.

Roughly speaking, the different layers fulfil different tasks and interact as a kind of master and slave to each other, top-down directly with the underlying layers:

- Application Layer: To allow access to network resources.
- Presentation Layer: To translate, encrypt and compress data.
- Session Layer: To establish, manage and terminate sessions.
- Transport Layer: To provide reliable process-to-process message delivery and error recovery.
- Network Layer: To move packets from source to destination and to provide internetworking.
- Data Link Layer: To organise bits into frames and to provide hop-to-hop delivery.
- Physical Layer: To transmit bits over a medium according to electrical and mechanical interface specifications.

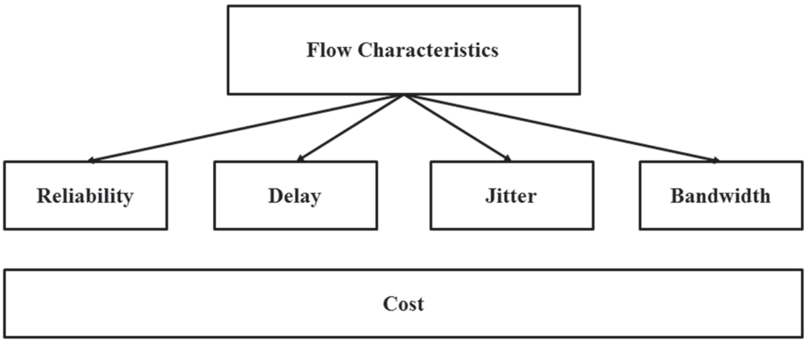


Figure 3 QoS parameters for the flow characteristics

ITU-T-defined QoS (quality of service) parameter sets, including service support, service operability, service accessibility, service retainability, service integrity, maintainability, availability, maintenance support, etc., were created to guarantee the reliable, timely and secure transmission of information with lowest bit error rates and low delay. In this book we will focus on the major technical parameters for the flow characteristics, i.e. availability or reliability, delay, jitter and available bandwidth (Fig. 3).

In fact, the QoS requirements for different services could be different. For example, voice or conversational video services are very sensitive to delay, but less sensitive to bit errors, whereas a data file transfer is extremely sensitive to bit errors, but less or even not at all sensitive to delay.

In order to provide reliable and secure communication between transmitter and receiver, the communication networks must be managed in a perfectly organised way. This is standardised by the network management systems TMN (Telecommunications Management Networks) and SNMP (Simple Network Management Protocol).

Throughout this book these terms of the communication channel model (Fig. 1) and the OSI model (Fig. 2) will be used often. At the end of this book, as outlook, future developments and trends will be discussed.

CHAPTER 2

INFORMATION THEORY

2.1 Shannon Capacity

A digital signal can be transmitted in a channel defined by its bandwidth (in Hertz) and by the noise and unwanted interference from other users. Since the total available frequency spectrum is limited, we are interested in the maximum data rate that can be transmitted over a channel with limited bandwidth and an existing noise and interference level.

In principle the data rate (in bps or bit/s) can be increased by increasing the number of bits per time. For simple modulation schemes, such as binary phase shift keying (see detailed explanation in Chapter 4), more bits/s consequently will increase the frequency spectrum of the signal and thus will require a larger bandwidth. Other modulation schemes such as higher order QAM (see Chapter 4) put the information bits, not only by using the phase shift, but also into amplitude steps. Obviously these systems require lower noise and interference within the channel, so that the receiver can reconstruct the original transmitted signal with its fine amplitude steps. Thus a higher data rate can be obtained with lower noise and interference, or a higher signal-to-noise ratio. The possible maximum data rate is called the channel capacity.

Claude Shannon published the first channel capacity estimation for digital data transmission stating that the total maximum channel capacity, or maximum data rate, depends basically on the available bandwidth B and the signal-to-noise ratio SNR.

$$C = B \cdot \log_2\left(1 + \frac{S}{N}\right). \tag{2.1}$$

The signal noise ratio S/N can be also written as SNR [1]. This formula represents the maximum limit of the channel capacity C depending on the available spectrum or bandwidth B and the signal-to-noise ratio SNR. To be more precise, the SNR normally considers the ratio of the signal power level to all disturbing factors like noise and interferences. In the general case of wireline or wireless transmissions, not only the distance-dependant reduction of signal power *S* during the wave propagation in the air, the thermal noises *N* of transmitter and receiver circuits, as well as the transmission medium, but also the interferences *I* from the neighbouring transmitters in the same frequency band, will reduce the signal-to-noise ratio. In some papers this term is also written as $SINR = S/(I+N)$, so that the Shannon capacity formula can be written also as

$$C = B \cdot \log_2 \left(1 + \frac{S}{I+N} \right). \tag{2.2}$$

Even though Shannon did not explicitly explain how to achieve the best channel capacity, this formula is the best approach to estimate the maximum channel capacity limit. In fact, all researchers working towards the goal to achieve the maximum capacity, to approach this maximum channel capacity limit, use sophisticated coding and modulation schemes as well as forward error correction techniques to exploit the given channel conditions.

Example:
Given the radio frequency bandwidth of 5 MHz and the signal-to-noise ratio of the wireless wave propagation environment SINR of 40 dB, what is the maximum channel capacity one can expect?

Hint: $$SINR_{dB} = 10 \cdot \log_{10}(SINR)\,. \tag{2.3}$$

Hint: $$SINR = 10^{(SINR_{dB}/10)}\,. \tag{2.4}$$

Example:
Given the available bandwidth of 10 THz for an optical fibre and the signal-to-noise ratio SINR (allowing for the attenuation loss of the optical wave propagation) to be 25 dB, what is the maximum channel capacity one can expect?

For the modern information and communication technologies, the information is transmitted as either an electrical or optical signal in digital modulation formats, as will be discussed in later chapters. Generally all information sources or parameters in real life (languages, music, images, pressure, velocity, acceleration, brightness, temperature, glucose, odour, etc.) can be converted to electrical and/or optical signals by using analogue digital converters. In the case of analogue amplitude modulation, the strength of these signals corresponds to the strength of the analogue electrical signals (current in Ampere, voltage in Volt, power in Watt). In digital modulation, the amplitudes will be converted into certain codes, which consist of only binary bits 0 and 1.

Coming back to the analogue amplitudes, if the signal levels have a very large dynamic range, or change strongly from very weak to very strong signals, it is not very practical to use the power levels between the output power level P_{out}= 0.000001 W or 1 μW and the input power level P_{in}=1 W. Instead we can use the logarithm to comprehensively represent the ratios of the power levels in dB. In this example the ratio of the input power level to the output power level will be then

$$a = \frac{P_{in}}{P_{out}} = \frac{1}{0.000001} = 1000000 = 10^6 . \qquad (2.5)$$

As we will discuss in the next chapters, the electrical analogue signal is defined by the signal strength. Taking the power P we can define a ratio of the power at the input or power at the output of an arbitrary complex circuit in dB

$$a = 10 \cdot \log\left[\frac{P_{in}}{P_{out}}\right] = 10 \cdot \log\left[\frac{1}{0.000001}\right] = 10 \cdot \log\left[10^6\right] = 60 \text{ dB}. \qquad (2.6)$$

We can also define the relative power level with a reference to 1 mW as dBm (or dBmW), or with reference to W as dBW for the power level at input or output, which makes the values more compact

$$p_{in} = 10 \cdot \log\left[\frac{P_{in}}{1mW}\right] = 10 \cdot \log\left[\frac{1W}{0.001W}\right] = 10 \cdot \log\left[10^3\right] = 30 \text{ dBm}$$

$$p_{out} = 10 \cdot \log\left[\frac{P_{out}}{1mW}\right] = 10 \cdot \log\left[\frac{0.000001W}{0.001W}\right] = 10 \cdot \log\left[10^{-3}\right] = -30 \text{ dBm}$$

$$p_{in} = 10 \cdot \log\left[\frac{P_{in}}{1W}\right] = 10 \cdot \log\left[\frac{1W}{1W}\right] = 10 \cdot \log\left[10^0\right] = 0 \text{ dBW}$$

$$p_{out} = 10 \cdot \log\left[\frac{P_{out}}{1W}\right] = 10 \cdot \log\left[\frac{0.000001W}{1W}\right] = 10 \cdot \log\left[10^{-6}\right] = -60 \text{ dBW}.$$

If the output power is higher than the input power, the circuit behaves as an amplifier. In the other case, if the output power is lower than the input power, the power is reduced within the circuit by means such as absorption or scattering. Physically, absorption means also the conversion of electrical or optical power into thermal energy.

It is this part of the lost energy that acts as disturbing noise, which is called thermal noise, arising when atoms and molecules vibrate and the temperature is non-zero. The higher the temperature, the stronger the vibration of the atoms and molecules will be. This noise vanishes only when theoretically the temperature is absolutely zero Kelvin 0 K = -273 °C, since in this case all the atoms or molecules are frozen completely. The thermal noise is constant with frequency, which is a good approximation for frequencies much lower than $k \cdot T/h$ or approximately 6 THz at room temperature or for most of the cases which we investigate in the radio frequency mobile communication networks:

$$N = k \cdot T \cdot B \qquad (2.7)$$

with $k = 1.38 \cdot 10^{-23}$ Ws/K as the Boltzmann constant and $h = 6.6 \cdot 10^{-34}$ Ws^2 as the Planck constant. Depending on the spectral bandwidth B the total noise can be calculated.

Example:
Assuming a bandwidth of 20 MHz at a temperature of 300 K, the noise can be calculated as following

$$N = k \cdot T \cdot B = 1.38\ 10^{-23} Ws / K \cdot 300K \cdot 20 \cdot 10^{6} Hz = 8.28 \cdot 10^{-14}\ \text{W}.$$

With the above-mentioned output power P_{out}= 0.000001 W, and neglecting all other noise sources of other involved electronic circuits and interferences coming from other systems, we obtain the signal-to-noise ratio

$$\frac{S}{N} = \frac{P_{out}}{N} = \frac{10^{-6} W}{8.28 \cdot 10^{-14} W} = 1.208 \cdot 10^{7}$$

$$SNR = 10 \cdot \log\left[1 + 1.208 \cdot 10^{7}\right] = 70.8\ \text{dB}.$$

This leads to an estimated total channel capacity, or maximally achievable bit rate with the unit bps (bit per second), kbps, Mbps or Gbps

$$C = B \cdot \log_2 (1 + \frac{S}{N}) = 20 \cdot 10^{6} Hz \cdot \log_2 (1 + 1.208 \cdot 10^{7}) = 470 \cdot 10^{6}\ \text{bps}.$$

2.2 Multimedia Data

The final goal is to achieve the maximum channel capacity correspondingly to certain quality of service QoS requirements like delay, jitter, available spectral bandwidth, bit error rates or availability.

Generally all multimedia information sources can be represented by the binary information unit bits and bytes, in order to be transmitted by the digital communication networks and technologies:

Text: In data communications text is represented as a bit pattern, a sequence of 0s or 1s. Different sets of bit patterns are designed to represent text symbols. Each set is called a code and the process of representing the symbols is called coding. Famous examples for the coding methods are: Unicode and ASCII code, among others.

Numbers: Numbers are directly coded or converted to binary numbers.

Audio: Audio recording or broadcasting of sound or music can be done when the continuous audio signals are sampled in certain time steps, quantised in the amplitude and then converted into digital signal formats. For improving the efficiency, some compression techniques can be used. One of the most famous methods is MP3.

Images: Images consist of a matrix of pixels (small dots and picture elements). The resolution depends on the pixel size. Each pixel is assigned to a bit pattern. Several methods are used to represent the colour images, e.g. RGB (Red, Green, Blue) and YCM (Yellow, Cyan, Magenta).

Video: Videos, motion pictures, flash animations or movies consist of a certain number of subsequent images, for example 25 images per second. So generally speaking, the video recording and compression basically corresponds to the optimised processing of images in the corresponding time unit.

Advantages of digital information processing and transmission in comparison with the classical analogue techniques are:

- Lower costs with no need for expensive high-quality analogue circuits and components.

- Improved channel efficiency by utilising compression and multiplex techniques.
- Less impact from semiconductor component tolerances and interferences.
- Accuracy and resolution of digital circuits can be arbitrarily high.
- High transmission quality.
- High reliability by using sophisticated channel coding and FEC techniques.
- Long lifetime of the storage capability.
- Continuous reduction of the size and weight of the processors by continuous development of microelectronics, integrated circuits and computer technologies.
- Reduction of the access time to the storage and the signal transit time in the processor.

Disadvantages:

- High circuit complexity.
- High spectrum demand. This disadvantage can be partially compensated for by powerful compression techniques.

Driven by the large-scale integration in the integrated circuit IC design, down to the sub-micrometer or nanometer structure, the efficiencies of the ICs have been continuously improved in the last decades. Depending on the number of the basic semiconductor components (diodes and transistors) for the signal processing, the ICs can be characterised in the following categories:

SSI (Small Scale Integration):	<	100 components
MSI (Medium Scale Integration):	<	1.000 components
LSI (Large Scale Integration):	<	10.000 components
VLSI (Very Large Scale Integration):	<	100.000 components
ULSI (Ultra Large Scale Integration):	<	1.000.000 components
SLSI (Super Large Scale Integration):	<	10.000.000 components
ELSI (Extra Large Scale integration):	<	100.000.000 components
GLSI (Giant large scale integration):	>	100.000.000 components

Examples of digital information processing and signal transmission: analogue telephone technology PSTN has been almost completely replaced by digital wired and wireless telephone technologies such as ISDN (Integrated Service Digital Network), xDSL (ADSL Asymmetrical Digital Subscriber Lines, VDSL Very high bit rate Digital Subscriber Lines), GSM (Global System of Mobile Communications), UMTS (Universal Mobile Telecommunication System), LTE (Long Term Evolution) and 5G mobile communication technologies. Meanwhile even the first generation wired telephone network ISDN is being replaced by the IP-based telephony, Voice over IP, or by the so-called All-IP network. IP is the abbreviation of Internet Protocol.

In audio technology, the classical analogue audio storage system like magnetic tapes have already been replaced by CD (Compact Disc) or MP3 (MPEG 1 and MPEG 2 Audio Layer III). MPEG stands for Moving Picture Experts Group, the standardisation body for digital audio and video signal processing.

Analogue video signals have been replaced by digital video formats like DVB-x (DVB-S for satellite video broadcasting, DVB-T for terrestrial video broadcasting and DVB-C for cable video broadcasting by using hybrid fiber coaxial broadcasting networks), MPEG-2, MPEG-4 and HDTV. One recent successful example is H.264 (MPEG-4/AVC Advanced Video Coding), or even H.265 (HEVC High Efficiency Video Coding).

Generally all information parameters like amplitudes, brightness, colours, etc. can be converted from the analogue form to the digital form in a series of binary information units 0 and 1. As an example we have the signal in Figure 4.

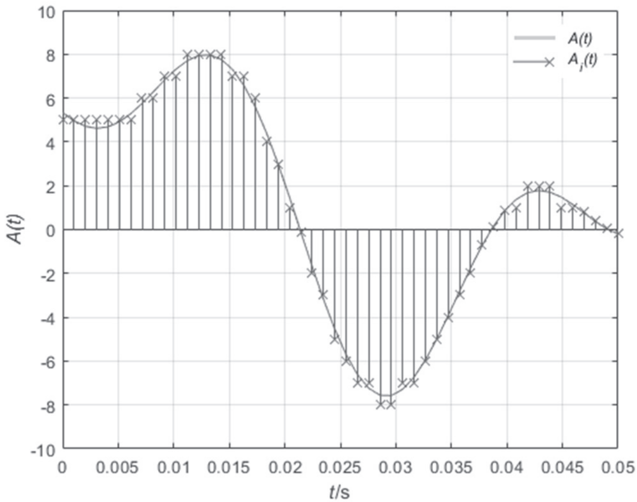


Figure 4 Sampling for analogue-digital conversion

The amplitude can be sampled with a period of ΔT, corresponding to a sampling frequency $f_{sampling}$, which is also called the sampling rate. Depending on the sampling rate, the temporal resolution or the sampling accuracy can be improved to an arbitrarily high value.

Shannon's sampling law (also known as Nyquist's criterion) requires the sampling frequency to be at least twice as high as the signal spectral bandwidth f_{Signal}, in order to fully represent the original analogue signal with no loss in information and to allow recovery of the original signal without deformation:

$$f_{Sampling} = \frac{1}{\Delta T} \geq 2 \cdot f_{Signal} . \tag{2.8}$$

By doing so, the continuous signal $A(t)$ will be sampled to discrete signal values $A(t_i)$.

The discrete, sampled amplitudes $A(t_i)$ will further be quantised in 2^n-1steps or intervals, with the integer number n, which is the number of the bits used to represent or approximate the different amplitudes between the maximum amplitude A_{max} and minimum amplitude A_{min}, and with the step size

$$\Delta A = \frac{A_{\max} - A_{\min}}{2^n - 1} .$$

The discrete amplitudes are calculated by

$$A(i) = A(0) + \Delta A \cdot j .$$

By using this stepwise quantisation, the amplitude $A(t)$ at the point in time t can be approximated by

$$A'(t_j) \cong A(j) \text{, if } \; A(j) \geq A \geq A(j-1) + \frac{\Delta A}{2}$$

$$A'(t_j) \cong A(j-1) \text{, if } \; A(j-1) + \frac{\Delta A}{2} \geq A(j) \geq A(j-1)$$

which corresponds to maximum quantisation error $\pm \Delta A / 2$

$$\left| A(t_j) - A'(t_j) \right| \leq \frac{\Delta A}{2} .$$

By doing so, all possible amplitudes between the allowable range of the maximum amplitude A_{max} and minimum amplitude A_{min} can then be approximated by $A'(t_j)$.

Again, the error for the amplitudes can be minimised to a wanted or required level depending on the choice of the number of the n bits to represent the amplitudes.

n bits	2^n-1 intervals	Error less than $\Delta A/2$	Relative to $(A_{max}-A_{min})$
1	1	$(A_{max}-A_{min})/(2^{(n+1)}-2) = (A_{max}-A_{min})/2$	0.500
2	3	$(A_{max}-A_{min})/(2^{(n+1)}-2) = (A_{max}-A_{min})/6$	0.167
3	7	$(A_{max}-A_{min})/(2^{(n+1)}-2) = (A_{max}-A_{min})/14$	0.071
4	15	$(A_{max}-A_{min})/(2^{(n+1)}-2) = (A_{max}-A_{min})/30$	0.033
5	30	$(A_{max}-A_{min})/(2^{(n+1)}-2) = (A_{max}-A_{min})60$	0.017
6	62	$(A_{max}-A_{min})/(2^{(n+1)}-2) = (A_{max}-A_{min})/124$	0.008
7	126	$(A_{max}-A_{min})/(2^{(n+1)}-2) = (A_{max}-A_{min})/252$	0.004
8	252	$(A_{max}-A_{min})/(2^{(n+1)}-2) = (A_{max}-A_{min})/504$	0.002

Table 1 Discretisation of the analogue signal for ADC

The amplitudes can then be described by a digital number or code, for example $A_{min} = A(0) \sim 0000$, $A_{max} = A(2^4) \sim 1111$, if $n = 4$ bits will be used, so that the analogue amplitudes as analogue numbers will be converted into a digital code, or series of binary bits, with a required accuracy for each amplitude value $A(t_j)$.

As we can see in the table above, the relative error in comparison with the maximum range of the possible amplitudes can also be minimised to a required, or theoretically an arbitrarily, low level, depending on the dynamic range of the amplitudes, albeit with increasing effort.

In the above example, the quantisation is linear and the step size ΔA is constant. In some cases the quantisation could also be non-linear, for example in case of voice signal quantisation for telephony. A-law is a companding algorithm for voice signal digital encoding used in Europe (in USA and Japan a similar μ-law is used), in order to modify the amplitude of the sampled voice signal, reduce the dynamic range and increase the signal-to-noise ratio. With A=87.6, the signal amplitude S is mapped to S′ (the signal amplitude S is chosen instead of A, in order to distinguish the A factor here):

$$S' = \frac{sign(S)}{\left(1+\ln(A)\right)} \cdot \begin{cases} A \cdot |S| & if\ |S| \le 1/A \\ 1+\ln(A \cdot |S|) & else \end{cases} . \qquad (2.9)$$

Example:

For the telephony application, the human voice can be bandpass-filtered in the spectrum from 300 Hz – 3400 Hz, without significantly degrading the voice quality and comprehensibility. By using the Shannon's sampling law (2.9) the sampling rate must be larger than

$$f_{Sampling} = \frac{1}{\Delta T} \ge 2 \cdot f_{Signal} = 6800 \text{ Hz}.$$

In fact, the sampling frequency or sampling rate is chosen with $f_{Sampling} = 8000$ Hz with a distance of 125 μs between the neighbouring samples, which will each be compressed to 8 bits representing the amplitude. At the end a data rate of 8 bits $\times$ 8000 Hz = 64 kbps for PCM encoding G.711 is achieved, which is used in the ISDN telephony.

A digital number is characterised by a certain number of bits. Each bit has only 2 states, either 0 or 1.

Examples for binary digital signals in electrical or optical information and communication technology:

- Electrical pulse is on or off.
- Electrical voltage is on or off.
- Optical transmitter, which is a laser diode, is switched on or off.
- Normally in electrical digital technologies, the 2 binary states are characterised by the voltage.
- For example the gates in the TTL logic present the logic states by the voltage level: 0V ("Low level", "L"), +5V ("High Level", "H").

In principle, the relation between the binary states L (Low) and H (High) to the logical states 0 and 1 could be arbitrary. Normally the so-called "Positive Logics" is applied, when:

The logic state "0" corresponds to the binary state "L" (low).
The logic state "1" corresponds to the binary state "H" (high).

The logic state "1" means in the logic design or Boolean logic "True" or "Correct", whereas the logic state "0" means in the logic design or Boolean logic "False" or "Wrong".

Of course, if required or meaningful for some applications, and, if the definition is used consequently, the "Negative Logics" can also be used, where the relation between the binary states L (Low) and H (High) to the logical states 1 and 0 are:

Logic state "1" corresponds to the binary state "L" (low).
Logic state "0" corresponds to the binary state "H" (high).

In this case, the logic state "0" means in the logic design or Boolean logic "True" or "Correct", whereas the logic state "1" means in the logic design or Boolean logic "False" or "Wrong". Normally the positive logic is used.

All these logic states can be controlled by the semiconductor integrated circuits based on the bipolar or MOSFET transistors and diodes.

2.3 Data Processing, Boolean Logics

The digital binary information of all multimedia data will be processed by Boolean logics. In the following, we first define the logic combinations "and", "or", "not" or negation. Values of the logic variables *A, B, C ... X, Y, Z* can be "0" or "1" [2].

$$Z = A \wedge B \quad \text{(AND)} \tag{2.10}$$
$$Z = A \vee B \quad \text{(OR)} \tag{2.11}$$
$$Z = \overline{A} = \neg A \quad \text{(NOT, negation)} \tag{2.12}$$

In Boolean algebra, these basic logics can also be combined:

$$Z = \overline{A \wedge B} = \overline{AB} \tag{2.13}$$
$$Z = \overline{A \vee B} \tag{2.14}$$
$$Z = (A \wedge \overline{B}) \vee (\overline{A} \wedge B) = A\overline{B} \vee \overline{A}B \tag{2.15}$$

AND	OR	NOT
$0 \wedge 0 = 0$	$0 \vee 0 = 0$	$\overline{1} = 0$
$0 \wedge 1 = 0$	$0 \vee 1 = 1$	$\overline{0} = 1$
$1 \wedge 0 = 0$	$1 \vee 0 = 1$	
$1 \wedge 1 = 1$	$1 \vee 1 = 1$	

Table 2 Boolean algebra

Similarly with the linear algebra, the Boolean algebra use also:

The Commutative laws:

$$A \wedge B \wedge C = C \wedge A \wedge B \quad (2.16)$$

$$A \vee B \vee C = C \vee A \vee B \quad (2.17)$$

The Associative laws:

$$A \wedge (B \wedge C) = (A \wedge B) \wedge C \quad (2.17)$$

$$A \vee (B \vee C) = (A \vee B) \vee C \quad (2.18)$$

The Distributive laws:

$$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C) \quad (2.19)$$

$$A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C) \quad (2.20)$$

The DeMorgan's laws:

$$Z = \overline{A \wedge B} = \overline{A} \vee \overline{B} \quad (2.21)$$

$$Z = \overline{A \vee B} = \overline{A} \wedge \overline{B} \quad (2.22)$$

Shannon's law of inversion:

$$\overline{f(A, B, C, ...; \vee, \wedge)} = f(\overline{A}, \overline{B}, \overline{C}, ...; \wedge, \vee) \quad (2.23)$$

The Absorption laws:

$$Z = A \wedge (\overline{A} \vee B) = A \wedge B \quad (2.24)$$

$$Z = A \vee (\overline{A} \wedge B) = A \vee B \quad (2.25)$$

$$Z = A \vee (A \wedge B) = A \quad (2.26)$$

$$Z = A \wedge (A \vee B) = A \quad (2.27)$$

By using these basic laws one can analyse logic circuits, or in a reverse way, design the logic circuit, in order to fulfil some tasks in engineering.

By using these basic Boolean logic or algebraic laws, computer chips, storage systems, sensors, analogue digital converters ADCs, digital analogue converters DACs, transmitters, receivers, source coders, channel coders and modulators can be built. Generally each of these information and communication networks, with a certain number of given input parameters A, B, C … etc., can be designed in order to process all the input data and to deliver a set of output parameters Z, Y, X …, etc.

The input parameters could be computer data, sensor data, user inputs, multimedia information to be transmitted to the receivers, and so on. The output data could correspondingly be machine control data or modulated digital data bit streams to be transmitted wirelessly or via wirelines.

Example:

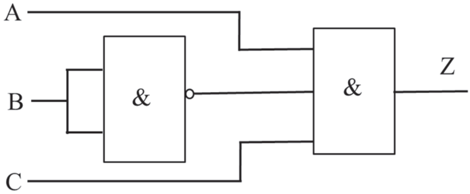
In order to demonstrate the logic design rules discussed above, a strongly simplified task is used [2] to design a digital circuit for safety purpose in such a way that the lift Z can lift up (Z=1) when, and only when:

1) The door contact sensor A shows that the lift door is closed (A=1). The other case is, the door is still open (A=0).
2) The overload sensor B shows that the lift is not overloaded (B=0). The other case is, the lift is overloaded B=1.
3) Someone pushes the button for one upper floor (C=1). If nobody pushes a button, then C=0.

This task can be translated into binary Boolean logic:

Case	C	B	A	Z
1	0	0	0	0
2	0	0	1	0
3	0	1	0	0
4	0	1	1	0
5	1	0	0	0
6	1	0	1	1
7	1	1	0	0
8	1	1	1	0

The control function Z for the lift will be then: $Z = A \wedge \overline{B} \wedge C$ which can be implemented with the digital logic gates, which in turn could be – for different applications and solutions – IC (mass production, consumer electronics), ASIC (application specific integrated circuit) or flexible FPGA (field programmable gate array, especially important for experimental setups and smaller technical problems):



2.4 Information Content, Entropy

In section 2.1 and section 2.2 we discussed channel capacity and multimedia data. Related to these topics is the information content. A multimedia message has certain information content. We will now discuss this information content in a more abstract or mathematical way.

Taking into account the different message units or symbols $m_1, m_2 \ldots, m_k \ldots m_n$ with the corresponding probabilities of occurrence $p_1, p_2 \ldots p_k \ldots p_n$ [1] [3] as an example. The symbols could be the alphabet of a language, used to express all the information contents. Shannon defined the amount of information of a symbol m_k as I_k with the binary information unit, or binary digit ("bit"), in a very strict technical sense, namely reciprocally proportional to the dual logarithm of the probability of occurrence.

$$I_k = \log_2 \frac{1}{p_k} . \tag{2.28}$$

Therefore the amount of the information each symbol carries directly depends on the probability of the occurence of this symbol in the message, which should be representative or large enough so that the probabilities of all symbols should be a stable average value.

If M symbols or messages are equally likely with $M = 2^N$, and independent from each other, the probability of each symbol will be $p_k=1/M$. The information content of each symbol is then

$$I_k = \log_2 \frac{1}{p_k} = \log_2 M = N . \tag{2.29}$$

If however the probabilities of the occurence of the symbols are different, so a long sequence of L messages will consist of different symbols m_k, each with the information I_k, then these L messages contain therefore the total information content

$$I_{Total} = p_1 L \log_2 \frac{1}{p_1} + p_2 L \log_2 \frac{1}{p_2} + \ldots + p_k L \log_2 \frac{1}{p_k} + \ldots + p_M L \log_2 \frac{1}{p_M} . \tag{2.30}$$

The average amount of the information of these L messages will then be defined as entropy $H= I_{Total} / L$

$$H = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + \ldots + p_k \log_2 \frac{1}{p_k} + \ldots + p_M \log_2 \frac{1}{p_M} \\ = \sum_{k=1}^{M} p_k \log_2 \frac{1}{p_k} \quad (2.31)$$

The maximum entropy can be achieved, if the distribution of the probabilities of the letters is equal. In this case the entropy will be

$$H_0 = \sum_{k=1}^{M} p_k \log_2 \frac{1}{p_k} = M \cdot \log_2 \frac{1}{1/M} = M \cdot \log_2 M \ . \quad (2.32)$$

Normally the maximum entropy cannot be achieved due to the unequal distribution of the probabilities of occurences. The difference of the maximum entropy H_0 and the realistic entropy H is called redundancy R.

$$R = H_0 - H \ . \quad (2.33)$$

As we will discuss later, the redundancy of an alphabet in a language or a code system is indispensable to allow for automatic FEC. This is also our daily experience. If somebody does not speak absolutely clearly, or if the cellular mobile phone reception is not very good, you may not understand every word of the conversation partner, but you still understand, at least you can imagine, what he means or what he could possibly say. The same is valid for written text. If someone writes a sentence with some missing or wrong letters, in many cases you can still understand the content of the text. This is exactly because each language has a lot of redundancy in the words and sentences, which are very helpful for the automatic error correction in daily communications. In information and communication technologies, redundancy is used to recognise and correct errors automatically in the similar way.

Taking the example of the English or German alphabet, the probabilities of different letters are totally different, leading to different amounts of information of each letter in an arbitrarily long text.

Example:

In Figure 5 and Figure 6, one example from our daily life is demonstrated by the probabilities of occurence of the letters in one article from the *The Independent* website as analysed by CrypTool [4]. In this article, all 26 English alphabet letters are used. A total of 1944 letters are contained in this text and different letters have different probabilities of occurence in this article. In German and English the letter used mostly is "e" with a probability of occurrence of approximately 14%.

By using the above-mentioned calculation, the entropy of this article is H = 4.14 with a maximally theoretically achievable H_0 =4.70, which could theoretically be valid, if all the letters would appear with the same probabilities. The redundancy of this English article will be R=0.56.

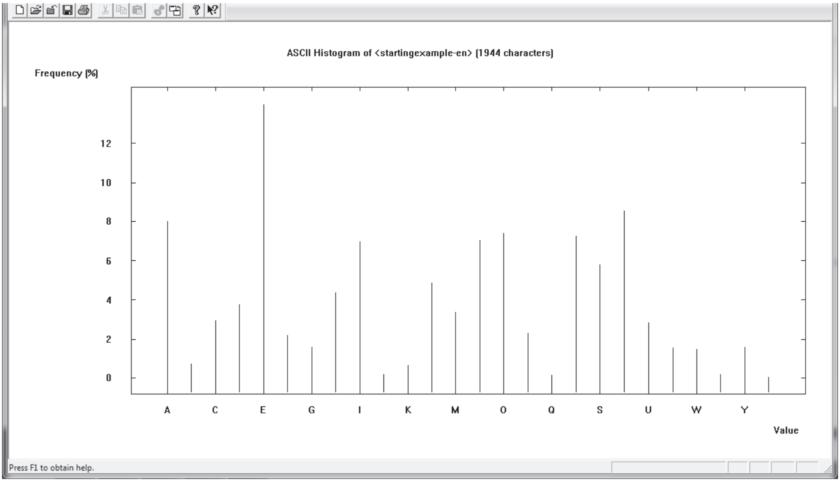
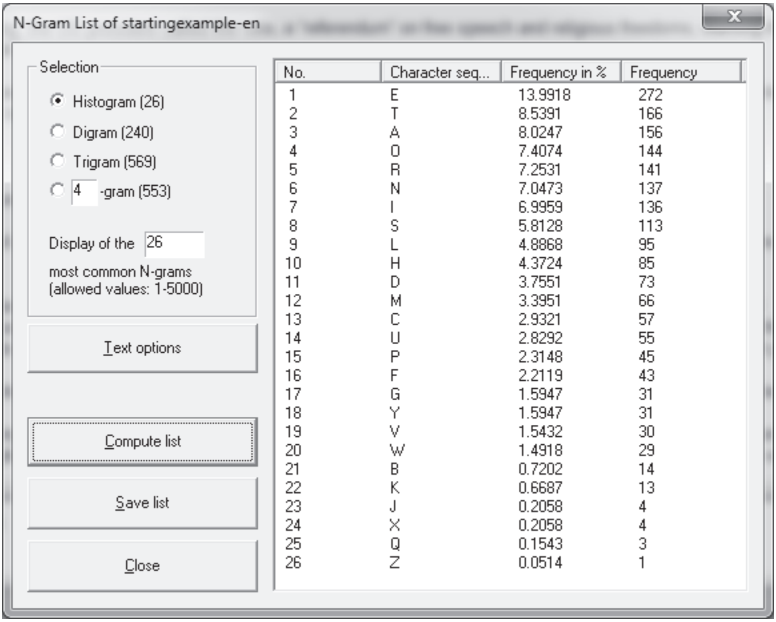


Figure 5 Occurence probability of the letters



No.	Character seq...	Frequency in %	Frequency
1	E	13.9918	272
2	T	8.5391	166
3	A	8.0247	156
4	O	7.4074	144
5	R	7.2531	141
6	N	7.0473	137
7	I	6.9959	136
8	S	5.8128	113
9	L	4.8868	95
10	H	4.3724	85
11	D	3.7551	73
12	M	3.3951	66
13	C	2.9321	57
14	U	2.8292	55
15	P	2.3148	45
16	F	2.2119	43
17	G	1.5947	31
18	Y	1.5947	31
19	V	1.5432	30
20	W	1.4918	29
21	B	0.7202	14
22	K	0.6687	13
23	J	0.2058	4
24	X	0.2058	4
25	Q	0.1543	3
26	Z	0.0514	1

Figure 6 Occurence probability of the letters

Since single letters appear with totally different probabilities, the redundancy will be increased. On one hand, the low redundancy means a relatively high efficiency to express the content. On the other hand, the errors will be hardly recognised and corrected.

Therefore the first question is: how we can code the source symbols in such an efficient way so as to reduce the redundancy, which is one basic task of the source coding. One efficient way is: regularly-used symbols should be coded by using short code words, and rarely-used symbols should be coded by longer code words. Code words should be binary, and coded at the transmitter and decoded by the receiver according to certain defined rules.

A famous example is the Morse alphabet (Fig.7). The most-used letter in English and German is the letter "e" which will be coded by ".", whereas a rarely used letter "q" will be coded by "- - · -". Comparing the Morse alphabet with the above-mentioned analysis of an arbitrarily chosen short text of the English newspaper *The Independent*, we can see some coincidence between the two examples. The numbers will be coded by five symbols consisting of "·" and "-".

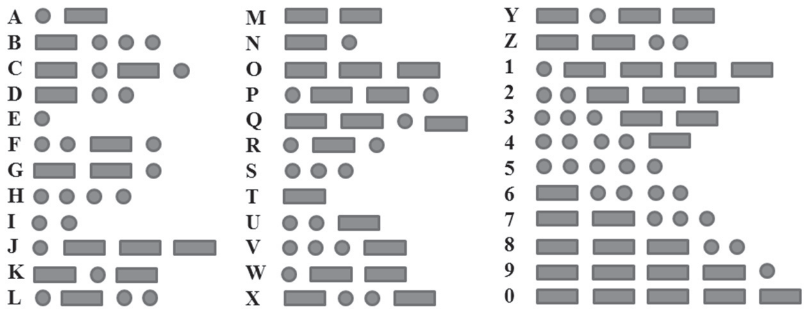


Figure 7 International Morse alphabet

Each dot has one-unit length and each dash has three-unit length. The separation of the dash and dot has one unit, whereas between the letters three units and between the words seven units will be needed, in order to separate them.

CHAPTER 3

SOURCE AND CHANNEL CODING

3.1 Source Coding

All multimedia information sources, i.e. audio, video and text data, contain redundant information which can be compressed to a minimum, in order to reduce the total requirement for transmission capacity, and to enable the efficient usage of the limited transmission network resources (Fig. 8). Also a multimedia information source contains irrelevant data like the noise of frequency components for voice which are not audible. The multimedia information sources like sound, music, still image and motion picture or video are first to be digitised or analogue-digital-converted (ADC analogue-digital conversion, sampling, quantising, source-coding, channel-coding), before they can be sent through the communication networks by modulating the source-coded data in an appropriate medium at a certain carrier frequency with a certain bandwidth.

According to the Nyquist criterion, the analogue signal with a bandwidth of B must be sampled with at least twice the sampling rate, in order to be able to reconstruct the digital signal back to the analogue signal without any deformation or loss of information. For example the human voice has a spectral width of approximately 10 kHz, though a limited spectrum of 3.4 kHz is sufficient to properly understand, so that the sampling should be at least 6,800 samples per second. The amplitude can be resolved or quantised by using a certain number of bits. For example, the voice signal can be quantized in $2^8 = 256$ logarithmic steps. Correspondingly, music and video can be analogue-digital-converted and processed.

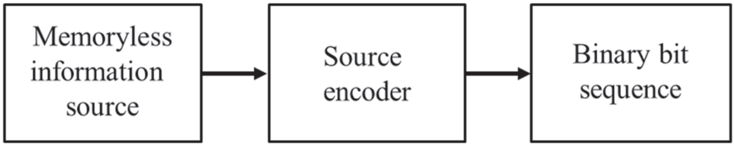


Figure 8 Source coding

Voice:
According to the Nyquist criterion, voice is sampled at 8,000 samples per second or 8 kHz, which is more than twice the signal bandwidth of 3.4 kHz. Each sample will be coded with 8 bits, meaning 256 amplitude steps for a good resolution (Note: these 256 steps are not equidistant, but follow a logarithm law, called A-law in Europe or μ-law in the USA). This results in a digital voice signal of 64 kbps for ISDN.

Music:
Music is often sampled at 44,000 samples per second with 16 bits per sample. This results in a digital signal of 705.6 kbps.

Video:
A colour video signal consists of 25 frames per second. Each frame is divided into small grids (pixels). Each pixel is represented by 24 bits (8 bits for red, 8 bits for green, 8 bits for blue). The lowest resolution of a colour frame consists of 1280×720×50 pixels per second, leading to 46.1 Mega pixels per second (Mpx/s). This means a minimum data rate of 1280×720×50 Hz×24 bits=1.106 Gbps**.** By using 50 Hz the so-called interlacing is used to improve the video flicker quality.

High Definition Television HDTV with the resolution of 1920×1080 will lead to approximately 2.5 Gbps.

In order to transmit the video or audio by the internet, and to efficiently use the scarce spectral bandwidth, the reduction of the total amount of the data is always necessary without degrading signal quality. This is possible, as we discussed in the last chapter, because the multimedia information sources always contain the so-called redundant information or redundancy *R*. This is also the main goal of source coding, removing the redundant information as much as possible, or to code the information more efficiently to achieve relatively high entropy.

The video signal consists of 50 images per second (50 Hz including interlacing) or 25 interlaced images per second. The compression will be done in two steps: spatial compression of the single image, and the removal of the redundant information of the subsequent temporal frames, where the redundant frames will be removed [5].

Spatial Compression:
The spatial compression of each video frame is done with JPEG (or a modification of JPEG). Each frame is a picture that can be compressed independently (Fig. 9). DCT (Discrete Cosine Transform) is used, because of its stable behaviour, prior to the quantisation and compression.

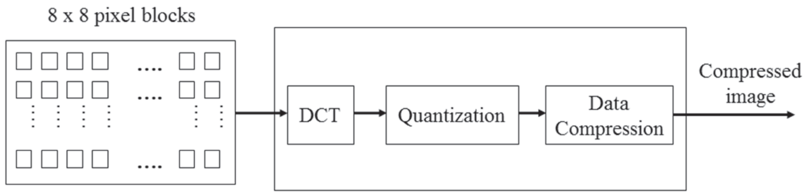


Figure 9 JPEG compression example

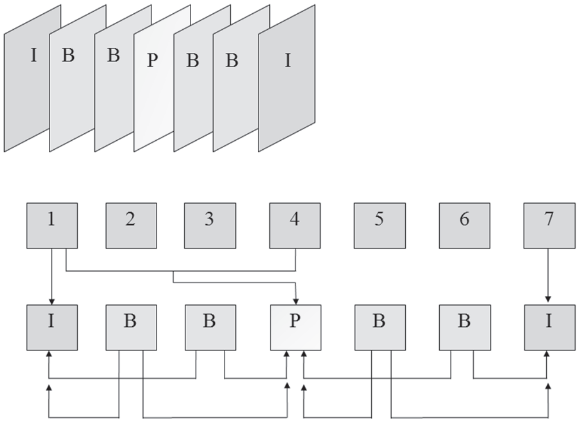


Figure 10 MPEG video compression

Temporal Compression:
In temporal compression, redundant frames are removed (Fig. 10). When we watch TV, we receive 50 frames per second. However, most of the consecutive frames are almost the same for most scenes. By considering this fact, only one out of six original frames will used, which is called the I-Frame. P-Frames and B-Frames can be derived from the preceding I-Frames and P-Frames.

I-Frames (Intracoded Frames):
The I-Frame is an independent frame or the original frame which is not related to any other frame.

P-Frames (Predicted Frames):
The P-Frame is related to the preceding I-Frame or P-Frame. It contains only changes from the preceding frame. The changes cannot cover a big segment.

B-Frames (Bidirectional Frames):
The B-Frame is related to the preceding and following I-frames or P-frames. A B-Frame is never related to other B-Frames.

For example by using the compression technique MPEG-2 (H.262) for 1080i50, a bit rate of only 27 Mbps is sufficient to transmit the video signal, instead of the above-mentioned total bit rate of 2.5 Gbps without compression.

Similarly, compression techniques for audio are MPEG1 for CD 1.5 Mbps and MPEG1 for Audio 100 – 400 kbps, where the original total audio data rate was 705.6 kbps. The video compression technique MPEG2 for DVD leads to a data rate of 2 – 6 Mbps.

3.2 Channel Coding

Whereas the source coding focuses on removing redundancy and irrelevance, therefore reducing the total data amount, channel coding adds intentionally redundant bits in a sophisticated way to recognise the transmission errors and to enable the so-called forward error correction FEC [3] [6].

As discussed in the last chapter, Shannon has shown that each communication channel is characterised by the SINR and bandwidth, defining the theoretical limit of the total channel capacity.

By using the appropriate channel coding technique, that is adding redundant bits to enable the automatic error correction, the total payload transmission capacity can be theoretically continuously improved to approach the channel capacity limit predicted by Shannon, as long as the signal processing performance is sufficient between transmitter and receiver, and the delay and jitter caused by the signal processing are tolerable.

The channel characteristics, i.e. available bandwidth, noise, signal power level or attenuation, as well as interferences due to neighbouring channels, do not only directly influence the quality parameters like bit error rate and delay, but in the end also the achievable throughput.

Shannon derived the channel capacity estimation formula which explicitly does not show the optimised methods for the channel coding.

Generally two procedures are possible to guarantee channel quality and error corrections. The first one is the simple ARQ (Automatic Repeat reQuest) technique, where the receiver recognises the possible transmission errors and sends the retransmission requests to the transmitter. This also means that an additional delay is caused by sending the requests and retransmitting the errored data packets. In order to avoid additional delay, the FEC technique is applied. In this case, the redundant bits are added in a proper manner to allow for recognising and correcting the errors automatically without ARQ.

In the following, some simple examples are used to demonstrate how the bit errors can be recognised and how these could be corrected. In reality, burst errors are possible, i.e. subsequent bits could be errored during the transmission from transmitter to receiver. In order to achieve single bit errors, the data stream can be prepared prior to transmission by using the so-called interleaving, where the neighbouring bits are reordered according to certain rules, so that these neighbouring errored bits will be separated. If a burst error occurs to the interleaved bit stream, then the bit stream can be de-interleaved at the receiver, so that the expected data bits will contain only single errored bits.

In order to enable easy error bit recognition and correction, normally the data bits are interleaved or separated prior to channel coding, and de-interleaved after the transmission and prior to decoding (Fig. 11). By doing so, possible burst bit errors (subsequent bit errors) caused by the channel noise and interference can be separated to single bit errors, which can be recognised and corrected simply.

Example 1: Interleaving

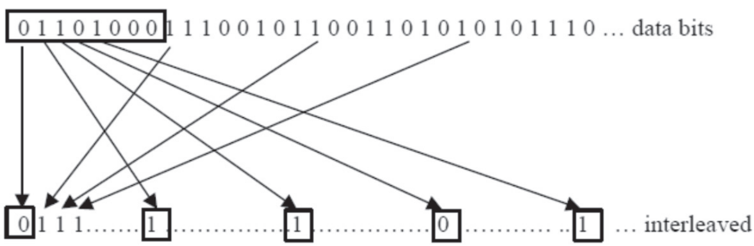


Figure 11 Interleaving example

Example 2: Redundant Bits

The basic idea for the error recognition and correction is to add redundant bits in such a way that single bit errors can be recognised or at least estimated. The simple method, which is by no means efficient, is to add redundant bits to each data bit, in order to recognise possible bit errors caused during the transmission. In a most simple case, the transmitter simply repeats the bits "0" and "1" two times, i.e. adding 200% redundancy, which helps to recognise, estimate and correct the errored bit, assuming only single bits could be errored.

Generally $0 \rightarrow 000$ and $1 \rightarrow 111$.

We assume that only a single bit is errored, not a burst of bits. Then the received bit sequences like 000, 100, 010, 001 could be assumed to be 0, whereas 111, 011, 101, 110 could be assumed to be 1.

0 1 0 1 0 0	Bits to be sent
000 111 000 111 000 000	Transmitter repeats the single bits
000 101 100 111 000 100	Receiver receives the errored data
0 1 0 1 0 0	Error correction

Example 3: Single Parity Check

Instead of redundant bits, a parity check bit can also be added to each block of data bits. In the following example, only a data bit stream of 7 bits are assumed. A single parity bit is used to ensure that the parity of the whole 8 bits (7 data bits + 1 parity bit) or that the number of 1s should be either even or odd. Again we assume single bit errors which could be enabled by interleaving the data bits prior to channel coding. The parity check can only recognise the errored data bit stream. In order to correct the errors, the ARQ will be sent to the transmitter to retransmit the data bit sequence.

Data bits	Parity bit
0011001	1
1101010	0
1011110	1
1010101	0
0111111	0
1001001	1
0011111	1

Example 4: Two-Dimensional Parity Check

Another method to check bit errors is to use the so-called two-dimensional or rectangular parity check (Fig. 12). The data bit stream is split into certain rows and columns. In each row and each column, consisting of 5 bits in this example, a parity bit is added vertically and horizontally, and chosen in such a way that the parity remains even. If we again assume a single bit error, after interleaving and de-interleaving, then the faulty bit can be localised and automatically corrected at the receiver site, without retransmission. This allows us to omit the ARQ process and reduce the delay caused by the ARQ request and retransmission. Of course the delay for adding the parity bits and checking the parity at receiver must be considered. This method is therefore usable if only single bit errors occur. Therefore, two-dimensional parity checks can be used to correct burst errors automatically in combination with interleaving.

Transmitter						Parity bit horizontal
	1	1	0	1	0	1
	1	0	0	1	1	1
	0	1	1	1	0	1
	0	1	0	1	1	1
	1	1	1	0	1	0
Parity bit vertical	1	0	0	0	1	0

Receiver						Parity bit horizontal
	1	1	0	1	0	1
	1	0	0	1	1	1
	0	1	0	1	0	1
	0	1	0	1	1	1
	1	1	1	0	1	0
Parity bit vertical	1	0	0	0	1	0

Figure 12: Rectangular parity check to find the errored bit

Example 5: Linear Block Codes
Linear Block Codes (n, k) consist of n code word bits and k message bits (see, for example, [3]). For a series of k message bits, an independent set of k vectors is needed to form the code vectors which are called generator matrix ***G***. ***G*** consists of a $(n\text{-}k)\times k$ parity check matrix ***P*** and an identity matrix ***I*** $(k \times k)$.

$$G = \begin{bmatrix} V_1 \\ V_2 \\ \cdots \\ V_k \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1,n-k} & 1 & 0 & \ldots & 0 \\ p_{21} & p_{22} & \cdots & p_{2,n-k} & 0 & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \ldots & \ldots & \ldots & \ldots \\ p_{k1} & p_{k2} & \cdots & p_{k,n-k} & 0 & 0 & \ldots & 1 \end{bmatrix}$$

(3.1)

Given the message vector m

$$m = \left(m_1, \quad m_2 \quad \ldots \quad m_k\right).$$

Then the code vector can be generated by product of m and ***G***.

$$\boldsymbol{U} = \boldsymbol{m}\,\boldsymbol{G}.$$

Additionally the parity check matrix ***H*** must be defined in order to check the received code vectors ***U***.

$$H = \left[I_{n-k} P^T\right] = \begin{bmatrix} 1 & 0 & \ldots & 0 & p_{11} & p_{21} & \cdots & p_{k,1} \\ 0 & 1 & \ldots & 0 & p_{12} & p_{22} & \cdots & p_{k,2} \\ \ldots & \ldots & \ldots & \ldots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & 1 & p_{1,(n-k)} & p_{2,(n-k)} & \cdots & p_{k,n-k} \end{bmatrix} \tag{3.4}$$

The received n bit sequence consists of the sent code vector ***U*** and possible transmission errors ***e.***

$$r = U + e\,.$$

Since the code vectors are independent, the correctly received code vectors without errors can be checked, if the following product, the so-called syndrome, is zero.

$$S = rH^T = (UH^T + eH^T) = eH^T .$$

Each syndrome therefore corresponds to one unique error pattern $\hat{e}$, therefore the error can then be corrected by

$$\hat{U} = r + \hat{e} .$$

The most famous block codes are Hamming codes and BCH codes (Bose-Chadhuri-Hocquenghem) (see detailed descriptions in [3]). The ratio k/n is called the code rate.

Example 6: Hamming Code

Bit Parity/Value	1 K_0	2 K_1	3 2^3	4 K_2	5 2^2	6 2^1	7 2^0
Decimal							
0	0	0	0	0	0	0	0
1	1	1	0	1	0	0	1
2	0	1	0	1	0	1	0
3	1	0	0	0	0	1	1
4	1	0	0	1	1	0	0
5	0	1	0	0	1	0	1
6	1	1	0	0	1	1	0
7	0	0	0	1	1	1	1
8	1	1	1	0	0	0	0
9	0	0	1	1	0	0	1

In this example, one of the regularly-used linear block codes, the Hamming code, is used to illustrate how efficiently the Hamming code can recognise and correct the bit errors [2] (Fig. 13). The seven bits consist of four information data bits (Bit3, Bit5, Bit6, Bit7) and three parity check bits (K_0, K_1, K_2). The code rate is therefore 4/7. Prior to transmitting the eight bits, the parity bits are grouped with two other information data bits (K_0, Bit3, Bit5, Bit7), (K_1, Bit3, Bit6, Bit7) and (K_2, Bit5, Bit6, Bit7), and are chosen in such a way, that the parity of each group will be even.

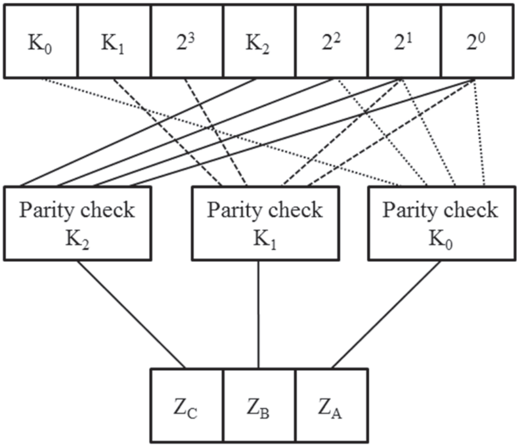


Figure 13 Hamming code parity check

At the receiver, each single bit error can be immediately recognised and corrected, as, for example, the results of the parity check (Z_C, Z_B, Z_A) shows exactly the position of the errored bit. For example, the binary triple (Z_C, Z_B, Z_A) = $(011)_2 = (3)_{10}$ means the third bit of the data bits has been errored during the transmission.

Odd parity in the groups	Results Z_C	Z_B	Z_A	Errored Bit
K_0	0	0	1	1
K_1	0	1	0	2
K_0 and K_1	0	1	1	3
K_2	1	0	0	4
K_0 and K_2	1	0	1	5
K_1 and K_2	1	1	0	6
K_0, K_1 and K_2	1	1	1	7

Similar linear block codes are the Extended Golay code and the BCH (Bose-Chadhuri-Hocquenghem) code (see details in [3]).

Example 7: Convolutional Code

A convolutional code is characterised by the parameters n, k, K, where k is the number of bits into the convolutional encoder, n the number of channel symbols output by the convolutional encoder in a given encoder cycle and K is the length of the shift register. Due to the shift register, the convolutional code is different from the linear block code, which is simply a logic combination of different data bits and parity bits, but temporally independent, or memory-less. The convolutional code depends temporally on the previous state of the data bits.

In Figure 14 a convolutional encoder example with a shift register constraint length of three is shown schematically [3]. On the left hand the message data bits are input, and will then be shifted step by step. The bits of the corresponding register are added by a modulo-2 adder. For each input symbol the register shifts the bits to the right. At the output the two code symbols will be sampled and an output symbol pair is obtained for each input symbol. Therefore the code rate k/n = ½.

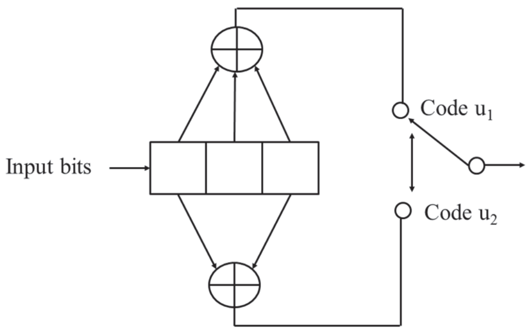


Figure 14 Convolutional encoder

Register content	Code symbol u_1	Code symbol u_2
<u>1</u>00	1	1
0<u>1</u>0	1	0
00<u>1</u>	1	1

In the above table, the single bit "1" is input from the left and shifted step by step to the right. Based on the register content, the output code symbol pair can be obtained by the modulo-2 adders.

Register content	Code symbol u_1	Code symbol u_2
<u>0</u>00	0	0
0<u>0</u>0	0	0
00<u>0</u>	0	0

In the above table, the single bit "0" is input from the left and shifted step by step to the right. Based on the register content, the output code symbol pair can be obtained by the modulo-2 adders.

If however a bit sequence of "101" is input (with a constraint length of *K=3* of the shift register) from the left and shifted to the right step by step, then the previous state of the register shifted to the right must be added by the modulo-2 adder, so that the sum will be 11 10 00 10 11.

Input data bit	u_1 u_2	u_1 u_2	u_1 u_2	u_1 u_2	u_1 u_2
1	11	10	11		
0		00	00	00	
1			11	10	11
Modulo-2-sum	11	10	00	10	11

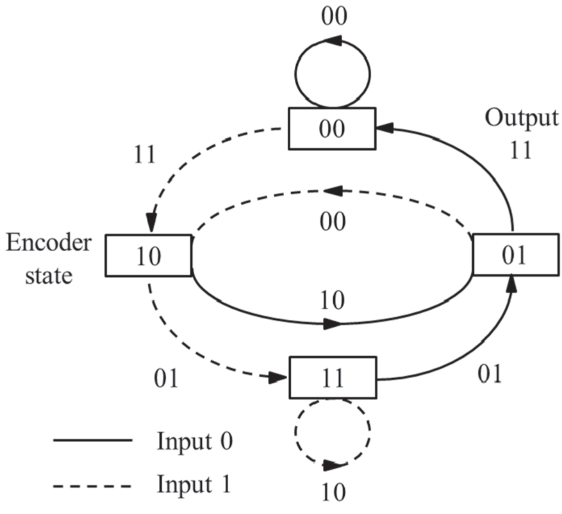


Figure 15 Convolutional encoder state diagram

Example 8: Viterbi Decoding

Viterbi decoding basically performs maximum likelihood decoding, by using the special trellis diagram to estimate the most probable code word. When two trellis paths merge, the Viterbi decoding algorithm must then decide to omit the trellis path which has the higher path metric: i.e. cumulated Hamming distances. In the example (explained in more detail in [3]) the input message data bits, coded transmitted codeword and received codeword are:

Input data bits	1	1	0	1	1
Transmitted codeword	11	01	01	00	01
Received codeword	11	01	01	10	01

By considering the convolutional state diagram in Figure 15, the possible state transitions can be drawn in the trellis diagram pair-wise. For each state transition, the Hamming distance, which is the number of different bits of the output codeword in the state diagram and the received codeword, can be depicted correspondingly to each transition. After certain time intervals, two paths merge to a common state. The surviving trellis path (the solid line in Fig. 16) with the minimum path metric will then be the most likely path.

The other path (the dashed line in Fig. 16) will be omitted for the following procedure. Since the output codeword of the convolutional encoder does not only depend on the input data bits, but also on the previous bits shifted in the register, therefore by using the efficient Viterbi decoding algorithm continuously, a number of the paths will be eliminated. By continuing this process for t_5, t_6, t_7 … etc., for all surviving paths, omitting all the paths with higher path metrics when merging at one certain state, the most likely output codewords or the path with the best metric will be chosen as the result.

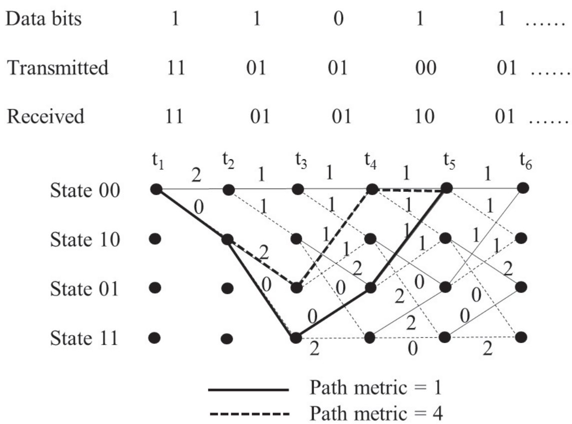


Figure 16 Trellis diagram with two merged paths

CHAPTER 4

MODULATION SCHEMES

Once the information is source- and channel-coded in the format of binary bit streams, these can then be transmitted by using certain carrier frequencies and different media like copper wires, coaxial cables, optical fibres or free space transmission. Thus millions of different signals can be transmitted at the same time, not interfering with each other, if appropriate multiple medium access schemes are used. This is very different from the early days of radio, when transmissions from and to ships were interfering with each other in a very small frequency spectrum, and signal bandwidth was used very inefficiently. In the case of free space transmission, either radio frequency bands (100 MHz – 100 GHz) or infrared/visible optical frequency bands (about 200 THz) can be utilised (see also the later chapter "Electromagnetic Waves").

Generally the spectrum in a certain frequency band, for example 2 GHz with the bandwidth of 5 MHz – 20 MHz in LTE or LTE Advanced technology, can be divided first into different, independent N sub bands with a bandwidth of, say, BW = 180 kHz. Each of these sub bands can now be modulated with different information signals $s(t)$.

These sub bands can be multiplexed to transmit independent information data sources. OFDM (Orthogonal Frequency Division Multiplex) is one method, used in different wireline and wireless technologies, where the sub bands are orthogonal, i.e. independent, to each other, and a certain spectral distribution of an envelope function at each central frequency f_i is used. Each of these carriers can then be modulated with a certain number of symbols by using a certain number of bits per symbol, in order to transmit the information.

We would like to emphasise that in this textbook we would not like to focus on the classical analogue modulation schemes AM (Amplitude Modulation), where the analogue electrical signals $s(t)$, containing the information, will be mapped or modulated to the change of the carrier signal amplitude $A(t)$, where the envelope of the carrier signal corresponds to the analogue signal. Similarly in FM (Frequency Modulation), the analogue signal changes or modulates the carrier signal frequency $f(t)$ in order to map the analogue signal to the carrier frequency, where the variable frequencies correspond to the analogue signal amplitudes.

Instead of these analogue modulation schemes, we begin right away with digital modulation schemes. First we would like to discuss how the information can be modulated to the radio frequency band by applying the digital modulation schemes BPSK (Binary Phase Shift Keying with M=2 symbols and each symbol 1 bit), QPSK (Quadature PSK with M=4 symbols and each symbol 2 bits), 8PSK (with M=8 symbols and each symbol 3 bits), 16QAM, 64QAM …. (Quadrature Amplitude Modulation with 16, 64 … and even higher numbers of symbols).

The radio frequency carrier signal $s_c(t)$ can be changed or modulated in different ways, in order to transmit the information signal, either in analogue or digital form, in amplitude $A(t)$, frequency $f(t)$, phase $\varphi(t)$ or a combination of some of these variables.

$$s_c(t) = A(t) \cdot \cos\left[\omega(t) \cdot t + \varphi(t)\right] = A(t) \cdot \cos\left[2\pi \cdot f(t) \cdot t + \varphi(t)\right]. \quad (4.1)$$

AM (Amplitude Modulation)

Classical analogue AM (Amplitude Modulation) varies the amplitude $A(t)$ to transmit the information signal $s(t)$, and maps $s(t)$ to the envelope of the oscillation at the fixed frequency $f(t) = f_c$.

$$s_c(t) = A(t) \cdot \cos\left[2\pi \cdot f_c \cdot t + \varphi\right]. \quad (4.2)$$

FM (Frequency Modulation)

Classical analogue FM (Frequency Modulation) varies the frequency $f(t)$ to transmit the information signal $s(t)$, and maps $s(t)$ to the frequency variation $f(t)$ with fixed amplitude A.

$$s_c(t) = A \cdot \cos\left[2\pi \cdot f(t) \cdot t + \varphi\right]. \quad (4.3)$$

PM (Phase Modulation)

Classical analogue PM (Phase Modulation) varies the phase $\varphi(t)$ to transmit the information signal $s(t)$, and maps $s(t)$ to the phase variation $\varphi(t)$ with fixed amplitude A and fixed frequency f.

$$s_c(t) = A \cdot \cos\left[2\pi \cdot f \cdot t + \varphi(t)\right]. \quad (4.4)$$

All analogue modulation schemes map the information signals to certain physical parameters like amplitude, frequency or phase. Even though the analogue modulation techniques are still used in different fields, for example AM in air traffic control and FM in radio broadcast, these analogue modulation schemes are being more and more replaced by digital techniques, for example DAB (Digital Audio Broadcasting) and DVB (Digital Video Broadcasting), basically because the digital modulation schemes are more efficient in terms of spectrum usage in most cases, and have the possibility of the forward error correction, so that higher throughput can be achieved.

In digital modulation schemes, both the amplitude *A(t)* and the phase $\varphi(t)$ can be varied simultaneously to modulate the information signal *s(t)* – in the simplest form a bit stream consisting of “0” and “1” like 11000100010011…000111. The baseband signals are the in-phase component *I(t)* and the quadrature component *Q(t)* (Fig. 17):

$$s(t) = I(t)\cdot\cos(2\pi\cdot f\cdot t) - Q(t)\cdot\sin(2\pi\cdot f\cdot t) =$$
$$= I(t)\cdot\cos(2\pi\cdot f\cdot t) + Q(t)\cdot\cos(2\pi\cdot f\cdot t+\frac{\pi}{2})\ . \qquad (4.5)$$

The phase and amplitude *s(t)* will form the so-called constellation diagram in the complex plane with the real part *I(t)* and the imaginary part Q*(t)*.

$$\underline{s}(t) = I(t) + j\cdot Q(t)\ . \qquad (4.6)$$

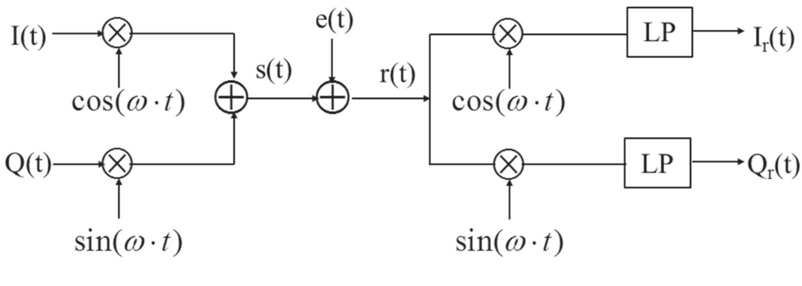


Figure 17 Complex quadrature carriers

BPSK

In BPSK *Q(t)* remains 0, whereas *I(t)* will be modulated between the two symbols representing “0” and “1” (Fig. 18). The constellation diagram contains only two symbols (2^1), representing one bit. The positions of the symbols are chosen in such a way, that the Hamming distance is always 1: i.e. only one bit is changed between the neighouring bit sets or symbols, in order to enable the FEC efficiently.

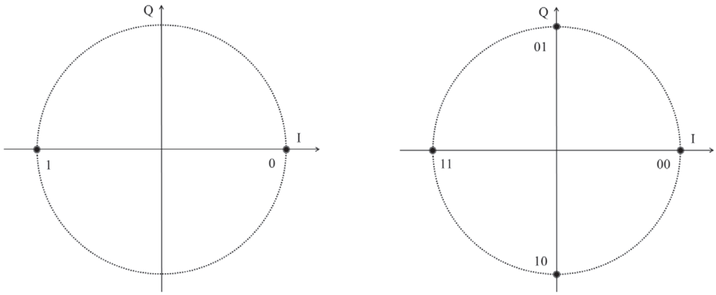


Figure 18 BPSK and QPSK

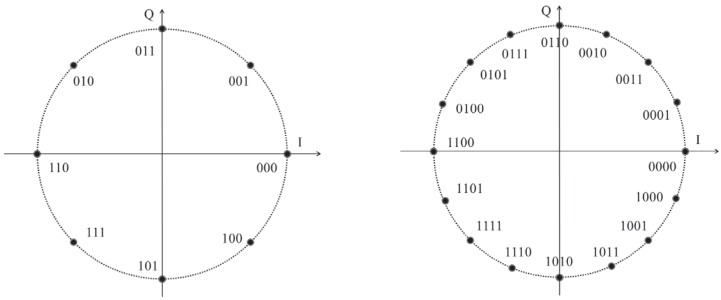


Figure 19 8PSK and 16 PSK

QPSK

In the case of QPSK (4PSK), *Q(t)* and *I(t)* together form four symbols (2^2) in the complex plane. Each symbol will represent one of the bit pairs “00”, “01”, “10”, “11” (Fig. 18). The positions of the symbols are chosen in such a way, that the Hamming distance is always 1: i.e. only one bit is changed between the neighouring bit sets or symbols, in order to enable the FEC efficiently.

8 PSK

In the case of 8PSK, *Q(t)* and *I(t)* together form eight symbols (2^3) in the complex plane. Each symbol will represent one of the three bit sets “000”, "001" … "110", "111" (Fig. 19). The amplitude remains the same, whereas for each symbol a different phase will be chosen. The positions of the symbols are chosen in such a way, that the Hamming distance is always 1: i.e. only one bit is changed between the neighouring bit sets or symbols, in order to enable the FEC efficiently.

16 PSK

In the case of 16 PSK, *Q(t)* and *I(t)* together form 16 symbols (2^4) in the complex plane. Each symbol will represent one of the four bit sets "0000", "0001" … "1110", "1111" (Fig. 19). The amplitude remains the same, whereas for each symbol a different phase will be chosen. The positions of the symbols are chosen in such a way, that the Hamming distance is always 1: i.e. only one bit is changed between the neighouring bit sets or symbols, in order to enable the FEC efficiently.

16 QAM

In the case of 16 QAM, *Q(t)* and *I(t)* together form 16 quadrature symbols (2^4) in the complex plane. Each symbol will represent the four bit sets "0000", "0001" … "1110", "1111" (Fig. 20). In this case, amplitude and phase change correspondingly, in order to form the constellation diagram.

32 QAM

In the case of 32 QAM, *Q(t)* and *I(t)* together form 32 quadrature symbols (2^5) in the complex plane. Each symbol will represent one of the five bit sets "00000", "00001" … "11110", "11111". In this case, amplitude and phase change correspondingly, in order to form the constellation diagram.

64 QAM

In the case of 64 QAM, *Q(t)* and *I(t)* together form 64 quadrature symbols (2^6) in the complex plane. Each symbol will represent one of the six bit sets "000000", "000001" … "111110", "111111" (Fig. 20). In this case, amplitude and phase change correspondingly, in order to form the constellation diagram.

256 QAM

In case of 256 QAM, *Q(t)* and *I(t)* together form 256 quadrature symbols (2^8) in the complex plane. Each symbol will represent one of the eight bit sets "00000000", "00000001" … "11111110", "11111111". In this case, amplitude and phase change correspondingly, in order to form the constellation diagram.

The positions of the QAM symbols are chosen in such a way, that the Hamming distance is always 1: i.e. only one bit is changed between the neighouring bit sets or symbols, in order to enable the FEC efficiently.

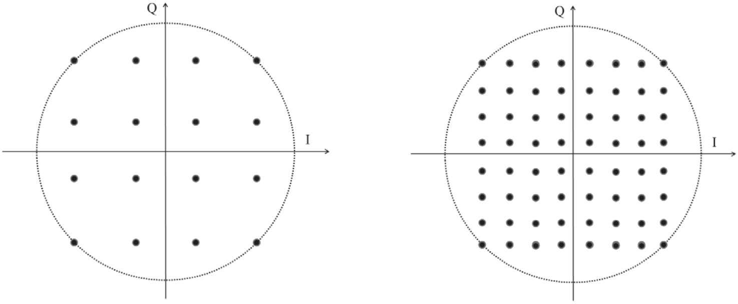


Figure 20 Constellation diagrams 16QAM and 64QAM

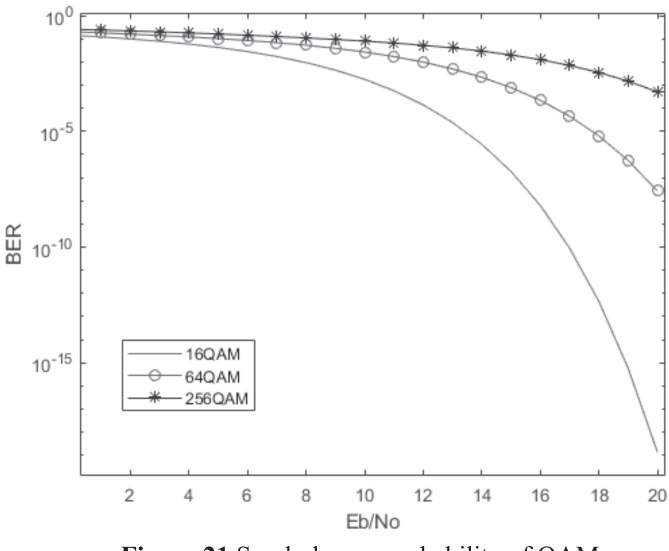


Figure 21 Symbol error probability of QAM

Figure 21 shows the symbol error probability of QAM depending on the signal-to-noise ratio. With increasing Eb/No, which is energy per bit to spectral noise density, corresponding directly to signal-to-noise ratio SNR, the bit error rate will decrease. Due to the relatively small distance in the constellation diagram for the higher order modulation scheme like

64QAM in comparison with 16QAM (see Fig. 19), the bit error probability BER will be correspondingly higher.

Depending on the modulation scheme and the symbol rate, the bit rate can be estimated. The so-called spectral efficiency or channel efficiency of different modulation schemes – which corresponds to the maximum number of the bits per symbol depending on Eb/No – are shown in Table 3.

Modulation scheme	Spectral efficiency b/s/Hz
BPSK	1
QPSK	2
8PSK	3
16PSK / 16QAM	4
64QAM	6
256QAM	8

Table 3 Channel efficiency of different modulation schemes

CHAPTER 5

INTERNET

Following the invention of telegraphs, the telephone was invented by Alexander Bell and Johann Philipp Reis by using twisted copper wires as transmission media. Then wireless communication was invented by Heinrich Hertz and Guglielmo Marconi by using electromagnetic waves in the nineteenth century, which will be treated in more detail in the next chapters. One of the very first applications was voice telephony, which was developed continuously for decades.

Voice telecommunication networks had a strongly hierarchical architecture. For a communication connection between the caller and the callee, a specific channel or circuit is set up, maintained during the conversation, and terminated after the conversation. This is called a CS (circuit-switched) network. All the traffic will be concentrated in the access and regional networks, and then be switched in the central offices in the wide area network nationwide or – in a limited manner also - worldwide. Text was transmitted by telegraphs. The central offices were and are normally located in the large metropolitan cities in each country.

After the invention of computers by Konrad Zuse (1945), John Presper Eckert and John William Mauchly (ENIAC, 1946), data processing and exchange required that data communication needed a specific network, since the narrow band, circuit-switched network for voice was not suitable and efficient enough.

During the cold war era, the DoD (Department of Defense) of the United States of America initiated the ARPA (Advanced Research Projects Agency), in order to establish an independent data communication network, primarily for sensitive military and government applications, to ensure the survivability of military operations in the case of atomic attacks [7]. The so-called ARPANET started on October 29, 1969 with NCP (Network Communication Protocol) and IMP (Internet Message Processor), Telnet and FTP Protocols, with 56 kbps telephone lines, wireless and satellite networks. Many computers were connected together. The interconnected computer network was born, as was the name Internetworking – or Internet.

During this project many useful protocols like TCP/IP (Transmission Control Protocol/Internet Protocol) and DNS (Domain Name Server) – the successor to NCP – were invented and used for the first time, as was the E-Mail application.

The internet remained basically a computer network for institutions and companies to exchange data for quite a long time, until the most famous invention WWW (world wide web) was developed by Tim Berners-Lee at CERN (Centre Européenne pour la Recherche Nucléaire), which was primarily used for exchange of the CERN research documents in terms of hyperlinks written in HTML (Hypertext Markup Language) in September 1991. But this very easy method to browse through the internet became much more interesting for private internet users, and as more and more applications were developed to meet the demands, complete new economies were born world-wide in the last 20 years, enabled by the internet and the world wide web. The most regularly-used internet protocols are summarised in the following:

MAC (Medium Access Control)
Each host, or each ethernet card, has a unique MAC address consisting of six octets (6×8 bits). Since this MAC address can be changed by software, it can not necessarily identify each hardware device or host without conflict. One of the most important applications is the assignment of an IP address by DHCP (Dynamic Host Configuration Protocol) to each host or hardware device.

PPP (Point to Point Protocol)
PPP is a protocol for data transport between the host computer and the ISP (Internet Service Provider). The details of the IP address and DNS server are exchanged. Therefore the most private internet connections are connected to the ISP via PPP.

PPTP (Point to Point Tunneling Protocol)
Proprietary protocol of Microsoft and 3COM for data transport by tunneling through PPP.

L2TP/L2FP (Layer 2 Tunneling/Forwarding Protocol)
L2TP enables the tunneling of the frames of layer 2 through routers between the L2TP Access Concentrator (LAC) and the L2TP Network Server (LNS) in an IP network.

ARP (Address Resolution Protocol) and RARP (Reverse Address Resolution Protocol)
ARP and RARP are responsible protocols which determine the assignment of IP addresses to a MAC address of a host or a hardware device.

DHCP (Dynamic Host Configuration Protocol)
DHCP assigns an IP address to a new host or hardware device in a LAN.

OSPF (Open Shortest Path First Routing)
OSPF is a routing protocol of IETF which is based on the link state shortest path algorithm of E. W. Dijkstra (IETF RFC 2328).

IGRP/EGRP (Interior/Exterior Gateway Routing Protocol)
IGRP and EGRP are CISCO distance-vector-protocols for the exchange of routing information in an interior network or between exterior networks. The metric considers available bandwidth, delay and traffic load.

BGP/GGP (Border Gateway/Gateway-to-Gateway Protocol)
BGP is a routing protocol between different or exterior autonomous networks. Therefore BGP is also called EGRP.

UDP (User Datagram Protocol)
UDP is a layer 4 transport protocol and adds a 16 bit port number to an IP address (OSI layer 3). UDP has a checksum, but does not contain sequence numbers in order to check the order of received packets. It also does not control the loss of packets or initiate the retransmission of lost packets. In comparison with TCP, UDP is also called an unreliable transport protocol.

TCP (Transmission Control Protocol)
TCP is a reliable transport protocol, which contains sequence numbers and acknowledgement numbers. The sequence numbers control the order of the received packets, and can correct the order of received packets, caused by jitter during the transmission. The acknowledgement number checks which packets do not arrive or are lost, and sends the retransmission request to the transmitter for the corresponding packet. By doing so, the number of lost packets can be strongly reduced. On the other hand, the checking of lost packets and the sending request and retransmission lead to additional delay, which is disadvantageous for the TSN (Time-Sensitive Network) or real-time applications.

SMTP (Simple Mail Transfer Protocol)
SMTP is TCP/IP (port 25) protocol for e-mail exchange (RFC 1855).

MIME (Mulitpurpose Internet Mail Extension)
MIME is a method to embed structured multimedia contents like images, texts, audio and video in simple e-mails.

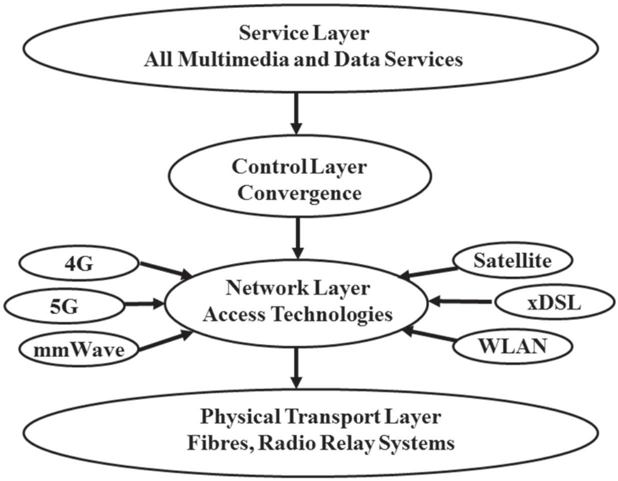


Figure 22 Integration of various services in an All-IP network

Depending on the coverage of the interconnected computers, we have:

- PAN: personal area network (Bluetooth, WiFi/WLAN stand-alone).
- LAN: local area network (Campus network, WiFi/WLAN infrastructure mode, fibre network).
- MAN: metropolitan area network (WiMax, partially meshed or ring fibre network).
- WAN: wide area network (statewide or nationwide network).
- GAN: global area network (international interconnected networks, trans-oceanic fibre network, satellite network).

Various applications and services must be via a convergence layer on top of different access technologies like 4G, 5G, WLAN, fibre, satellite, etc. (Fig. 22).

Figure 23 shows a classical transport network, whereas Figure 24 shows the All-IP transport network which is now being deployed by many network operators.

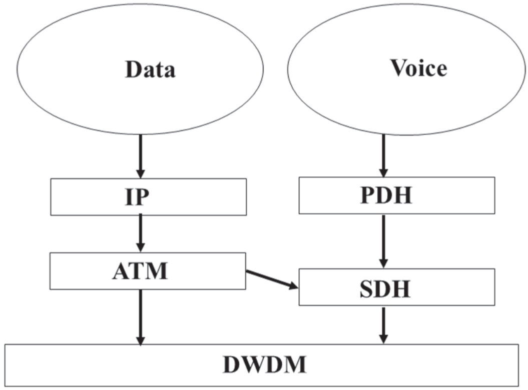


Figure 23 Classical Transport Network

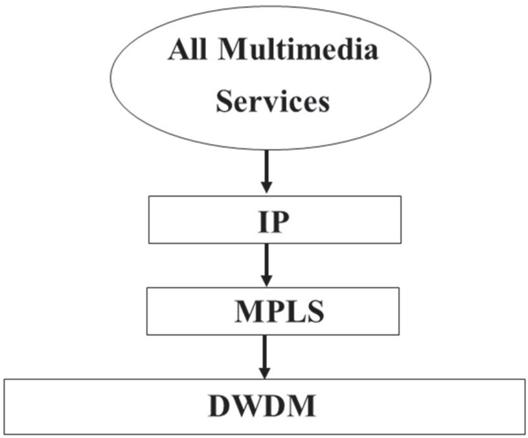


Figure 24 Future All-IP Transport Network

CHAPTER 6

ELECTROMAGNETIC WAVES

Maxwell´s equations are the governing equations to analyse all electromagnetic wave propagation problems, from RF radio waves used in the cellular mobile communications 1 GHz - 2 GHz, mm waves (mm waves in frequency range of 30 – 100 GHz) up to the optical wave propagation in the frequency range of 200 THz.

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} . \tag{6.1}$$

$$\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{J} . \tag{6.2}$$

$$\nabla \cdot \vec{D} = \rho . \tag{6.3}$$

$$\nabla \cdot \vec{B} = 0 . \tag{6.4}$$

Field	Symbol	Unit
Electric field	$\vec{E}$	V/m
Dielectric displacement / density	$\vec{D}$	As/m^2
Magnetic field	$\vec{H}$	A/m
Magnetic flux density / induction	$\vec{B}$	$T = Vs/m^2$
Current density	$\vec{J}$	A/m^2

Table 4 Basic electromagnetic field parameters and units

By using vector algebra, the above Maxwell´s equations can be derived in the different coordinate systems depending on the problems to be investigated. For example, for the rectangular metallic waveguides, the Cartesian coordinate system (x, y, z, t) will be meaningful, whereas for the problems of cylindrical metallic or quartz glass optical waveguides, the cylindrical coordinate system (r, ϕ, z, t) will be used. For the antenna problems again, the polar coordinate system (r, θ, ϕ, t) will be helpful.

In the following we will discuss briefly the vector algebra and the derivation of the Maxwell´s equations into the Helmholtz equations as a good approximation [8] [9] [10].

Generally a dynamic field is a physical quantity with different values and orientations at different positions and different points in time.

Vector fields

Vector fields that vary with different frequencies f have always characteristic values or magnitudes, and point into certain directions at certain positions. For example:

Electric field $\vec{E}(x, y, z, t)$ and electric displacement $\vec{D}(x, y, z, t)$.

Magnetic field $\vec{H}(x, y, z, t)$ and magnetic induction $\vec{B}(x, y, z, t)$.

Mechanical force $\vec{F}(x, y, z, t)$ and pressure $\vec{P}(x, y, z, t)$.

Velocity of solid body or fluid $\vec{v}(x, y, z, t)$.

Scalar fields

In comparison with vector fields, the scalar fields have only a value or a magnitude, and are direction-independent. For example:

Temperature field $T(x, y, z, t)$.

Potential $\varphi(x, y, z, t)$.

Electric charges $q(x, y, z, t)$.

Vector algebra definitions are the following:

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \quad \text{nabla operator} \quad (6.5)$$

$$\nabla \vec{E} = \left(\frac{\partial E_x}{\partial x}, \frac{\partial E_y}{\partial y}, \frac{\partial E_z}{\partial z} \right) = grad\vec{E} \quad \text{gradient} \quad (6.6)$$

$$\nabla \cdot \vec{E} = \nabla_x E_x + \nabla_y E_y + \nabla_z E_z \quad \text{divergence} \quad (6.7)$$

$$\nabla \times \vec{E} = rot\vec{E} = curl\vec{E} \quad \text{rotation, curl} \quad (6.8)$$

$$\left(\nabla \times \vec{E} \right)_z = \nabla_x E_y - \nabla_y E_x \quad (6.9)$$

$$\left(\nabla \times \vec{E} \right)_x = \nabla_y E_z - \nabla_z E_y \quad (6.10)$$

$$\left(\nabla\times\vec{E}\right)_y = \nabla_z E_x - \nabla_x E_z \tag{6.11}$$

Useful rules of the vector algebra are:

$$\nabla\cdot\left(\nabla T\right) = \nabla^2 T\ , \quad \text{a scalar field} \tag{6.12}$$

$$\nabla\times\left(\nabla T\right) = 0\ , \quad \text{curl of the gradient of a scalar field is 0} \tag{6.13}$$

$$\nabla\left(\nabla\cdot\vec{E}\right), \quad \text{a vector field} \tag{6.14}$$

$$\nabla\cdot(\nabla\times\vec{E}) = 0\ , \text{ divergence of the curl of a vector field is 0} \tag{6.15}$$

$$\nabla\times\left(\nabla\times\vec{E}\right) = \nabla\left(\nabla\cdot\vec{E}\right) - \nabla^2\vec{E} \tag{6.16}$$

$$\left(\nabla\cdot\nabla\right)\vec{E} = \nabla^2\vec{E} \quad \text{a vector field} \tag{6.17}$$

In the following tables the often-used IEEE frequency bands and ITU frequency bands with special abbreviations are shown.

Abbreviation	**Frequency band**	**Wavelength range**
HF	0.003 GHz-0.03 GHz	10 m – 100 m
VHF	0.03 GHz-0.3 GHz	1 m – 10 m
UHF	0.3 GHz – 1 GHz	300 mm –1000 mm
L Band	1 GHz - 2 GHz	150 mm – 300 mm
S Band	2 GHz- 4 GHz	75 mm – 150 mm
C Band	4 GHz- 8 GHz	37.5 mm – 75 mm
X Band	8 GHz- 12 GHz	25 mm – 37.5 mm
Ku Band	12 GHz- 18 GHz	16.7 mm – 25 mm
K Band	18 GHz- 27 GHz	11.1 mm – 17 mm
Ka Band	27 GHz- 40 GHz	7.5 mm – 11.1 mm
V Band	40 GHz- 75 GHz	4 mm – 7.5 mm
W Band	75 GHz- 110 GHz	2.7 mm – 4 mm
mm Waves	100 GHz- 300 GHz	1 mm – 2.7 mm
THz Band	*300 GHz- 3 THz*	*0.1 mm – 1 mm*
IR Band	*3 THz- 480 THz*	*0.625 μm –100 μm*
Visible Light	*480 THz- 680 THz*	*0.44 μm – 0.625 μm*

Table 5 IEEE frequency bands

The IEEE frequency bands and wavelength ranges of the most-used electromagnetic waves in the information and communication technologies

– together with the terahertz THz band, infrared IR-band and visible light – are shown in Table 5.

ITU No.	Band	Frequency Bands	Wavelengths
1	ELF	3–30 Hz	100,000–10,000 km
2	SLF	30–300 Hz	10,000–1,000 km
3	ULF	300–3,000 Hz	1,000–100 km
4	VLF	3–30 kHz	100–10 km
5	LF	30–300 kHz	10–1 km
6	MF	300–3,000 kHz	1,000–100 m
7	HF	3–30 MHz	100–10 m
8	VHF	30–300 MHz	10–1 m
9	UHF	300–3,000 MHz	1–0.1 m
10	SHF	3–30 GHz	100–10 mm
11	EHF	30–300 GHz	10–1 mm
12	THz, THF	300–3,000 GHz	1–0.1 mm
	Infrared	*3 THz- 480 THz*	*0.625 μm -100 μm*
	Visible Light	*480 THz- 680 THz*	*0.44 μm – 0.625 μm*

Table 6 ITU frequency bands

ITU frequency bands and wavelength ranges of the most-used electromagnetic waves in the information and communication technologies – together with infrared IR-band and visible light – are shown in Table 6. In Table 7 the following ITU and IEEE abbreviations are used:

ELF:	Extremely low frequency
SLF:	Super low frequency
ULF:	Ultra low frequency
VLF:	Very low frequency
LF:	Low frequency
MF:	Medium frequency
HF:	High frequency
VHF:	Very high frequency
UHF:	Ultra high frequency
SHF:	Super high frequency
EHF:	Extremely high frequency
THF:	Tremendously high frequency
L	Long wave
S	Short wave

C	Compromise between S and X
Ku	Kurz (“short”) under
K	Kurz (“short”)
Ka	Kurz (“short”) above

IEEE Standard		EU/NATO Bands	
L	1 - 2 GHz	D	1 - 2 GHz
S	2 - 4 GHz	E	2 - 3 GHz
C	4 - 8 GHz	F	3 - 4 GHz
X	8 - 12 GHz	G	4 - 6 GHz
Ku	12 - 18 GHz	H	6 - 8 GHz
K	18 - 27 GHz	I	8 - 10 GHz
Ka	27 - 40 GHz	J	10 - 20 GHz
V	40 - 75 GHz	K	20 - 40 GHz
W	75 - 110 GHz	L	40 - 60 GHz
mm	110 - 300 GHz	M	60- 100 GHz
Terahertz	0.3 - 3 THz	Terahertz	0.3 - 3 THz
Infrared	3 - 480 THz	Infrared	3 - 480 THz
Visible Light	480 - 680 THz	Visible Light	80-680 THz

Table 7 IEEE and EU/NATO Bands

Forces, electric and magnetic fields

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right). \tag{6.18}$$

Principle of superposition of fields in linear media

$$\vec{E} = \vec{E}_1 + \vec{E}_2 . \tag{6.19}$$

From equations (4.1) and (4.2) we see that electric and magnetic fields are always related to each other and form the electromagnetic waves transporting energy in a certain direction, as defined by the Poynting vector:

$$\vec{S} = \vec{E} \times \vec{H} . \tag{6.20}$$

Electromagnetic waves in a generally inhomogeneous dielectric isotropic medium require generally numerical methods, in order to solve the differential equations. In some cases, some simplified assumptions help to solve these differential equations analytically, as will be discussed later.

By considering the generally frequency-dependent characteristics of the electromagnetic waves in a generally inhomogeneous dielectric medium at the frequency f or angular frequency $\omega = 2\pi f$, we have

$$\underline{\vec{E}}(\vec{r},\omega) = \int_{t=-\infty}^{\infty} \vec{E}(\vec{r},t) \cdot e^{-j\omega t} \cdot dt \, . \qquad (6.21)$$

This equation is the Fourier transform of the electric field. It allows us to calculate the frequency spectrum from a given function versus time. Interpreting this equation, we see that a signal with a high data rate and where the signal amplitude changes rapidly with time has a wide frequency spectrum or needs a large bandwidth.

With this complex notation, equations (4.1) and (4.2) give:

$$\nabla \times \underline{\vec{H}} = j\omega \underline{\vec{D}} + \vec{J} \, . \qquad (6.22)$$

$$\nabla \times \underline{\vec{E}} = -j\omega \underline{\vec{B}} \, . \qquad (6.23)$$

In generally inhomogeneous media with complex dielectric permittivity coefficients and complex magnetic permeability coefficients, the relations between the electric field $\vec{E}$, the magnetic field $\vec{H}$, the dielectric displacement $\vec{D}$ and the magnetic flux density $\vec{B}$ can be described as follows:

$$\underline{\vec{D}}(\vec{r},\omega) = \left(\varepsilon'(\omega) - j\varepsilon''(\omega)\right) \underline{\vec{E}}(\vec{r},\omega) \, . \qquad (6.24)$$

$$\underline{\vec{B}}(\vec{r},\omega) = \left(\mu'(\omega) - j\mu''(\omega)\right) \underline{\vec{H}}(\vec{r},\omega) \qquad (6.25)$$

Whereas the real parts correspond to the so-called dispersion, the imaginary parts correspond to the losses caused by absorptions. Dispersion describes the frequency-dependent characteristics, i.e. different wave propagation velocities $\vec{v}$ at the different wavelengths.

These relations are described by the so-called Debye Dispersion Model which describes the delayed reaction of the molecular dipoles to the applied electromagnetic waves. See, for example, [11] with the relaxation time τ which describes the retardation of the molecular dipole in response to the excitation field.

$$\frac{\varepsilon'(\omega)}{\varepsilon_0} = \varepsilon_r(\omega) = \varepsilon_r(\infty) + \frac{\varepsilon_r(0) - \varepsilon_r(\infty)}{1 + (\omega\tau)^2}. \tag{6.26}$$

$$\frac{\varepsilon''(\omega)}{\varepsilon_0} = \omega\tau \frac{\varepsilon_r(0) - \varepsilon_r(\infty)}{1 + (\omega\tau)^2}. \tag{6.27}$$

Additionally the phase shift between $\vec{D}(\vec{r},t)$ and $\vec{E}(\vec{r},t)$ is caused by the absorption loss.

By considering the absorption losses ε'' of the dielectric medium and the Ohm's loss caused by the limited conductivity κ, the total power generated by the source $\vec{J}$ will be

$$-\frac{1}{2}\oiiint_V \underline{\vec{E}} \cdot \underline{\vec{J}}^* dV = \frac{1}{2}\oiint_A \underline{\vec{E}} \times \underline{\vec{H}}^* \cdot d\vec{A} + \frac{1}{2}\oiiint_V \kappa \left|\underline{\vec{E}}\right|^2 dV - \frac{1}{2}\oiiint_V (\underline{\vec{E}} \cdot \underline{\vec{D}}^* - \underline{\vec{B}} \cdot \underline{\vec{H}}^*) dV \tag{6.28}$$

The left hand term represents the generated power, for example by an dipole antenna with the excitation current $\vec{J}$, whereas the first term of the right hand expressions is the Poynting vector, or the radiated power from this antenna. The real part of the second and third terms represent the loss caused by the absorption loss of the dielectric medium and the limited conductivity that is directly related to ε'' or tan δ.

$$\underline{\varepsilon} = \varepsilon' - j\varepsilon'' - j\frac{\kappa}{\omega} = \varepsilon'(1 - j\tan\delta). \tag{6.29}$$

$$\tan\delta = \frac{\kappa + \omega\varepsilon''}{\omega\varepsilon'}. \tag{6.30}$$

For homogenous, time-invariant media (grad ε = 0):

$$\nabla \times (\nabla \times \vec{E}) = -\mu\varepsilon \frac{\partial^2 \underline{\vec{E}}}{\partial t^2} - \mu \frac{\partial \vec{J}}{\partial t} \tag{6.31}$$

$$\nabla\left(\nabla \cdot \vec{E}\right) - \nabla^2 \vec{E} = -\mu\varepsilon \frac{\partial^2 \vec{E}}{\partial t^2} - \mu\kappa \frac{\partial \vec{E}}{\partial t} \tag{6.32}$$

$$\nabla^2 \vec{E} - \mu\varepsilon \frac{\partial^2 \vec{E}}{\partial t^2} - \mu\kappa \frac{\partial \vec{E}}{\partial t} = 0 \tag{6.33}$$

$$\nabla^2 \underline{\vec{E}} + \omega^2 \mu\varepsilon \left(1 - j\tan\delta\right) \underline{\vec{E}} = \nabla^2 \underline{\vec{E}} + \omega^2 \mu\varepsilon \left(1 + \frac{\kappa + \omega\varepsilon''}{j\omega\varepsilon'}\right) \underline{\vec{E}} = 0 \quad (6.34)$$

we get the simplified Helmholtz equations by using the complex dielectric permittivity $\underline{\varepsilon}$:

$$\underline{\varepsilon} = \varepsilon \left(1 + \frac{\kappa + \omega\varepsilon''}{j\omega\varepsilon'}\right). \quad (6.35)$$

$$\nabla^2 \underline{\vec{E}} + \omega^2 \mu \underline{\varepsilon} \underline{\vec{E}} = 0 . \quad (6.36)$$

With the definition of the so-called wave number *k* and wave propagation velocity *v,* where *v* will be free space light velocity *c*, if the medium in the most simple case is vacuum ($\mu = \mu_r \mu_0 = \mu_0$, $\varepsilon = \varepsilon_r \varepsilon_0 = \varepsilon_0$). This simple case is also most important for wireless radio frequency communications

$$k = \omega\sqrt{\mu\varepsilon} = \frac{\omega}{v} . \quad (6.37)$$

For homogeneous anisotropic dielectric mediums the independent permittivity coefficients become dependent on the orientation in certain media, for example crystals, semi-conductors, etc. In this case the dielectric permittivity coefficients must be described by using a tensor matrix:

$$\left(\varepsilon_{ij}\right) = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix}. \quad (6.38)$$

The similar tensor matrix $\left(\mu_{ij}\right)$ is valid for the anisotropic magnetic medium.

For inhomogeneous dielectric $\varepsilon(x, y, z)$, isotropic, non-magnetic $\mu_r = 1$ and lossless mediums, the electric field component $\vec{E}$ of the electromagnetic waves travelling in an inhomogeneous, dielectric, isotropic, lossless medium, which is an ideal assumption of course, could be described by the differential equation

$$\nabla^2 \vec{E} = \mu_0 \varepsilon(x,y,z) \frac{\partial^2 \vec{E}}{\partial t^2} - \nabla\left(\frac{\nabla \varepsilon(x,y,z)}{\varepsilon(x,y,z)} \vec{E} \right). \quad (6.39)$$

Whereas the magnetic field component of the electromagnetic waves travelling in inhomogeneous, dielectric, isotropic, non-magnetic, lossless mediums could be similarly described by the differential equation

$$\nabla^2 \vec{H} = \mu_0 \varepsilon(x,y,z) \frac{\partial^2 \vec{H}}{\partial t^2} - \left(\frac{\nabla \varepsilon(x,y,z)}{\varepsilon(x,y,z)} \right) \times \left(\nabla \times \vec{H} \right). \quad (6.40)$$

Considering a homogenous media ($\varepsilon(x,y,z)$ =constant, $\nabla \varepsilon(x,y,z)$ =0), for example free space, or at least a partially homogenous region like the core or cladding region of the quartz glass optical fibre which will be discussed later, the above differential equations can be simplified to the so-called Helmholtz equations with the amplitude of the electric field $\hat{E}(\vec{r},t)$ and unit vector of the polarisation plane $\vec{u}$

$$\vec{E}(\vec{r},t) = \hat{E}(\vec{r},t) \cdot \vec{u} \qquad \text{and} \qquad \frac{\nabla \varepsilon(x,y,z)}{\varepsilon(x,y,z)} < 0.01 . \quad (6.41)$$

$$\Delta \hat{E} = \nabla^2 \hat{E} = \frac{\partial^2 \hat{E}}{\partial x^2} + \frac{\partial^2 \hat{E}}{\partial y^2} + \frac{\partial^2 \hat{E}}{\partial z^2} = \mu_0 \varepsilon(x,y,z) \frac{\partial^2 \hat{E}}{\partial t^2} . \quad (6.42)$$

Similarly, for the extremely small variation of the relative dielectric permittivity less than 1%, for example in case of dielectric waveguide like quartz optical fibre, the so-called “weekly guidance condition” is valid, where the relative gradient of the permittivity vanishes to zero, so that the Helmholtz equation is also valid. The general form of the solutions of the Helmholtz equation can be obtained for (6.42)

$$\hat{E}(\vec{r},t) = f\left(t - \frac{\vec{k} \cdot \vec{r}}{c} \right) \quad (6.43)$$

where $\vec{k}$ is the wave number vector pointing into the propagation direction of the electromagnetic waves and c is the vacuum speed of the light. The corresponding solution for $\hat{H}(\vec{r},t)$ and $\vec{H}$ can be found similarly.

The simplest solution for Maxwell´s equations or the Helmholtz equation could be the so-called transverse electromagnetic TEM waves. For example:

$$\vec{E}(\vec{r},t) = \hat{E}(\vec{r},t) \cdot \vec{x} \ . \qquad (6.44)$$

$$\vec{H}(\vec{r},t) = \hat{H}(\vec{r},t) \cdot \vec{y} \ . \qquad (6.45)$$

$$\vec{S}(\vec{r},t) = \left| \vec{E}(\vec{r},t) \times \vec{H}(\vec{r},t) \right| \cdot \vec{z} \ . \qquad (6.46)$$

The ratio between the perpendicular electrical and magnetic field $Z = \hat{E} / \hat{H}$ = 377 Ω =120π Ω becomes constant, and is defined as the characteristic impedance of the medium which could be air or a homogenous dielectric medium.

The governing Maxwell´s equations can be used to explain almost all the wave propagation mechanisms, both for free space wave propagation, antenna problems and optical waveguides.

CHAPTER 7

WAVE PROPAGATION

Depending on the problems to be investigated, different coordinate systems can be used, so that Maxwell´s equations can be transformed to the suitable differential equations which can be solved.

Wave propagation problems can be analysed in two groups: 1) guided waves by using metallic waveguides and dielectric or quartz glass fibre wave guides, and 2) wave radiation and propagation problems, for example antennas, free space propagation, diffraction and scattering.

Waveguides: coaxial waveguide, twisted pair copper wires, rectangular and circular waveguides and optical waveguides or fibres are shown schematically in Figure 25.

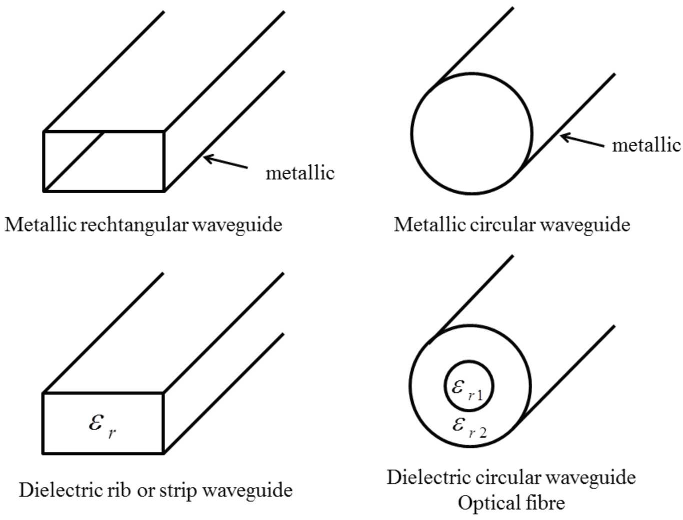


Figure 25 Waveguides

Antennas and wave radiation in free space: antenna problems, free space radio frequency and optical wave propagation, and multipath propagation or scattering problems can be analysed using Maxwell´s equations.

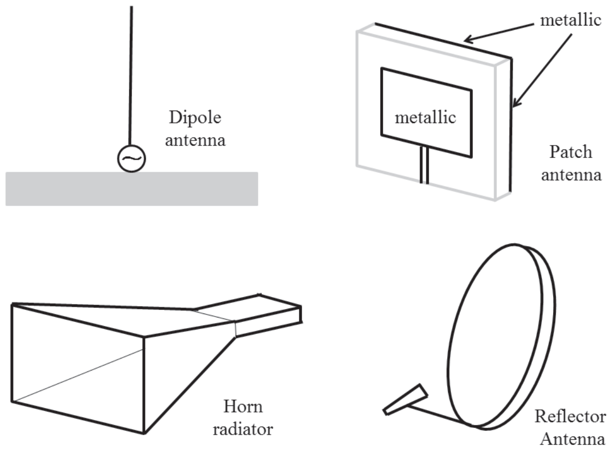


Figure 26 Antennas

7.1 Elementary Antennas

Antennas:
Typical and often-used antenna types (Fig. 26-Fig. 29) with special radiation characteristics are:

- Theoretical, hypothetic isotropic radiator
- Dipole
- Patch
- Horn
- Helix
- Parabolic antenna
- Cassegrain antenna
- Slot antenna, etc.

The performance and characteristics of the different types of antennas are characterised by the antenna gain *g* which describes the increase of the power density in the direction of the antenna radiation beam in comparison with the isotropic radiator (dBi). The higher the gain, the narrower the antenna radiation beam will be. The theoretical hypothetic isotropic antenna is considered as a reference.

Type	**Polarisation**	**Gain**
Isotropic antenna		0 dBi
Dipole antenna	linear	~1.5 - 2.3 dBi
Patch antenna	linear, circular	~ 7 - 10 dBi
Horn antenna	linear	~ 20 dBi
Helix antenna	circular	~ 10 - 18 dBi
Parabolic antenna		~ 20 - 70 dBi

Table 8 Typical antenna properties

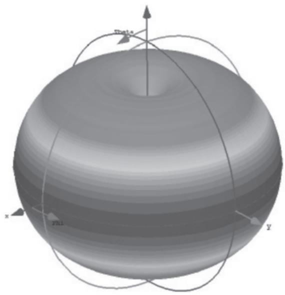


Figure 27 Radiation characteristics of a dipole antenna

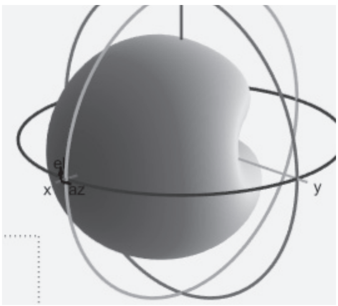


Figure 28 Radiation characteristics of a patch antenna

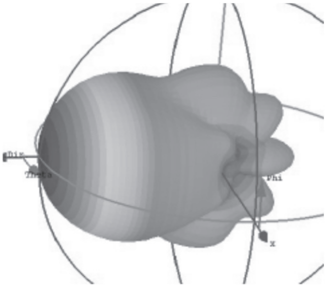


Figure 29 Radiation characteristics of a horn antenna

7.2 Wave Propagation Models

Free Space Wave Propagation Loss

Under the ideal assumption of free space wave propagation, with the transmitter assumed to be an isotropic radiator and there being no reflexion, no diffraction, no scattering and no multipath propagation, the power density of the radiated wave then will depend on distance only:

$$S = \frac{P_T}{4\pi l^2}. \tag{7.1}$$

This is only a strongly simplified model to describe the wave propagation, which is only possible for the wave propagation of an isotropic radiator in free space communications in a vacuum (no absorption, turbulence, scintillation, etc.) or if the transmitter and receiver are both far away from the earth's surface and there are absolutely no objects in the vicinity of them.

With the effective aperture of the receiver antenna, allocated at the receiver site in a distance of l to the transmitter site,

$$A = \frac{\lambda^2}{4\pi} \tag{7.2}$$

the received signal power will be:

$$P_R = \frac{P_T}{4\pi l^2} A \tag{7.3}$$

$$P_R = \frac{P_T}{4\pi l^2}\left(\frac{\lambda^2}{4\pi}\right) \tag{7.4}$$

$$\frac{P_R}{P_T} = \left(\frac{\lambda}{4\pi l}\right)^2 \quad (7.5)$$

with

$$\lambda = c/f \quad (7.6)$$

$$\frac{P_R}{P_T} = \left(\frac{c}{4\pi lf}\right)^2 \quad (7.7)$$

$$\frac{P_R}{P_T} = \left(\frac{c}{4\pi}\right)^2 l^{-2} f^{-2}. \quad (7.8)$$

The relative free space wave propagation loss, i.e. the ratio of the received signal power level to the transmitted signal power level will then be

$$L_{FSL} = \left(\frac{3\times10^8}{4\pi}\right)^2 l^{-2} f^{-2} \quad (7.9)$$

Expressed in dB, the free space loss will be

$$L_{FSL} = -32.44 - 20\log_{10}\left(\frac{l}{\text{km}}\right) - 20\log_{10}\left(\frac{f}{\text{MHz}}\right) \quad (7.10)$$

In reality, because of the limited height of the transmitter base station antenna, the limited height of the receiver and, in most cases of cellular mobile communications, the position of the mobile devices or smart phones, there will be additional losses, caused by atmospheric absorption, scintillation and turbulence, scattering, shadowing, reflexion from the earth surface and buildings, penetration loss through walls and other impairments. In different areas like densely-populated urban areas, suburban areas or rural areas, these impacts of the impairments are totally different, so that many different models have been developed in order to describe these effects appropriately. The most popular models are the Okumura model and the Hata model for frequencies below 2 GHz based on the measurement performance by Okumura.

Okumura model

The Okumura model is formally expressed as:

$$L = L_{FSL} + A_{mu} - G_{MG} - G_{BG} + C; \qquad (7.11)$$

where the parameters are:

L = the median path loss in dB.
L_{FSL} = the free space loss in dB.
A_{mu} = median attenuation in dB.
$G_{MG}(h_{re})$ = mobile antenna gain with effective antenna height h_{re}.
$G_{BG}(h_{te})$ = base station antenna gain with effective antenna height h_{te}.
C = correction factor gain (such as type of environment, water surfaces, isolated obstacles, etc.).

Okumura's model is one of the most widely-used models for signal prediction in urban areas. This model is applicable for frequencies in the range of 150–1920 MHz (although it can be typically extrapolated up to 3000 MHz) and distances of 1–100 km. It can be used for base-station antenna heights ranging from 30–1000 m.

Okumura developed a set of curves giving the median attenuation relative to free space A_{mu} in an urban area over a quasi-smooth terrain with a base station effective antenna height h_{te} of 200 m and a mobile antenna height h_{re} of 3 m. These curves were developed from extensive measurements using vertically polarised, omnidirectional antennas at both the base and mobile stations, and are plotted as a function of frequency in the range of 100–1920 MHz and as a function of distance from the base station in the range of 1–100 km.

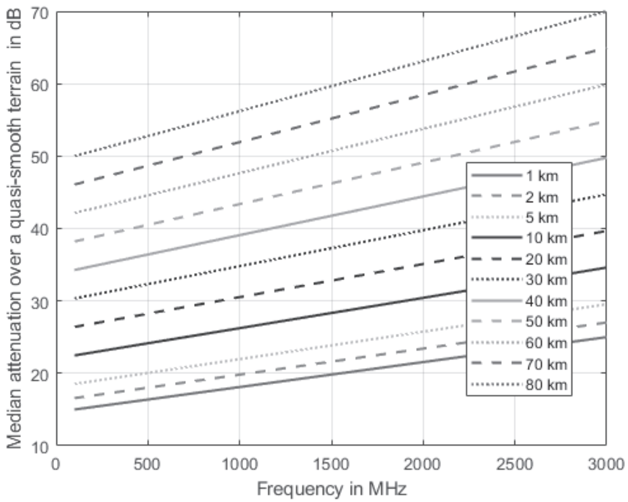


Figure 30 Approximation of median attenuation *Amu(f, d)*

To determine path loss, using Okumura's model, the free space path loss between the points of interest is first determined, and then the value of median attenuation over a quasi-smooth terrain $A_{mu}(f, d)$ depending on the operation frequency f and distance d (as read from the curves in Figure 30 which were approximated based on the Okumura measurements in [12]) is added to it along with correction factors to account for the type of terrain.

Note that the antenna height gains are strictly a function of height and have nothing to do with antenna patterns.

Furthermore, Okumura found that $G(h_{te})$ varies at a rate of 20 dB/decade and $G(h_{re})$ varies at a rate of 10 dB/decade for heights less than 3 m.

$$G(h_{te}) = 20 \log(h_{te}/200) \qquad 1000\ m > h_{te} > 30\ m \qquad (7.12)$$
$$G(h_{re}) = 10 \log(h_{re}/3) \qquad h_{re} <= 3\ m \qquad (7.13)$$
$$G(h_{re}) = 20 \log(h_{re}/3) \qquad 10\ m > h_{re} > 3\ m \qquad (7.14)$$

Other corrections may be applied to Okumura's model. Some of the important terrain-related parameters are the terrain undulation height, isolated ridge height, average slope of the terrain and the mixed land-sea parameter. Once the terrain-related parameters are calculated, the necessary correction factors can be added or subtracted as required. All these correction factors are also available as Okumura curves [12] [13].

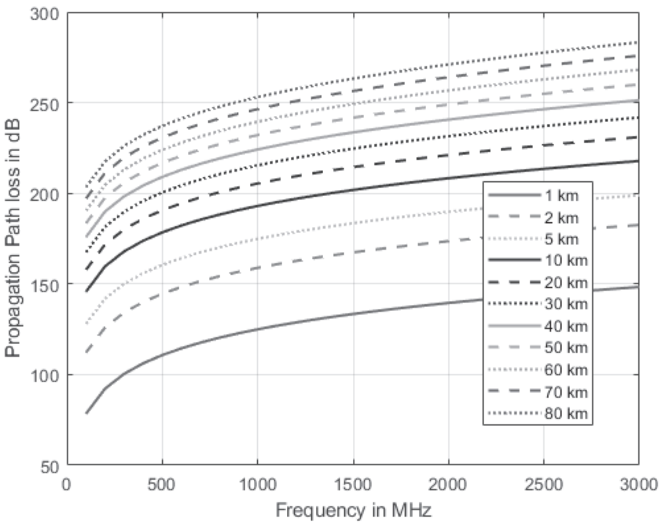


Figure 31 Okumura path losses depending on *d* and *f*

In irregular terrain, non-line-of-sight nLOS paths, caused by terrain obstacles, occur frequently. Okumura's path loss model (Fig. 31) includes a correction factor called the "Isolated Ridge" factor to account for obstacles. However, the applicability of this correction is only to obstacles conforming to that description, i.e. an isolated ridge. More complex terrain cannot be modeled by the isolated ridge correction factor. A number of more general models exist for calculating diffraction loss. However, none of these can be applied directly to Okumura's basic mean attenuation. Proprietary methods of doing so have been developed. However, none are known to be in the public domain.

Okumura's model is completely based on measured data and does not provide any analytical explanation. For many situations, extrapolations of the derived curves can be made to obtain values outside the measurement range, although the validity of such extrapolations depends on the circumstances and the smoothness of the curve in question.

Okumura's model is considered to be among the simplest and best in terms of accuracy in path loss prediction for mature cellular and land mobile radio systems in cluttered environments. It is very practical and has become a standard for wireless system planning in modern land mobile radio systems in Japan. The major disadvantage with the model is its slow response to rapid changes in the terrain. Therefore the model is fairly good in urban and suburban areas, but not as good in rural areas. Common standard deviations between predicted and measured path loss values are around 10 - 14 dB.

Hata model for urban environments

The Hata model for urban environments is based on Okumura's measurements [13] [14] , all in dB:

$$L_u = 69.5 + 26.6 \cdot \log_{10} f + 13.82 \cdot \log_{10} h_B - C_H + (44.9 - 6.5 \cdot \log_{10} h_B) \cdot \log_{10} d \tag{7.15}$$

for a small or medium-sized city:

$$C_H = 0.8 + (1.1 \cdot \log_{10} f - 0.7) \cdot h_M - 1.56 \cdot \log_{10} f \tag{7.16}$$

for large cities, and 150 MHz < f < 200 MHz:

$$C_H = 8.9 \cdot (\log_{10}(1.54 \cdot h_M))^2 - 1.11 \tag{7.17}$$

for large cities, and 200 MHz < f < 1500 MHz:

$$C_H = 3.2 \cdot (\log_{10}(11.75 \cdot h_M))^2 - 4.97 \tag{7.18}$$

where

L_u = path loss in urban areas (dB)
h_B = height of base station antenna (m)
h_M = height of mobile station antenna (m)
f = frequency of transmission (MHz)
C_H = antenna height correction factor
d = distance between the base and mobile stations (km).

Hata model for suburban environments

The Hata model for suburban environments is applicable to the radio links [15] in suburban areas where man-made structures are present, but not as high and dense as in cities. To be more precise, this model is suitable where buildings exist, but the mobile station does not have a significant variation of its height. It is formulated as:

$$L_{SU} = L_u - 2 \cdot (\log_{10}(f / 28))^2 - 5.4 \qquad (7.19)$$

where

L_{SU} =	path loss in suburban areas (dB)
L_u =	path loss in urban area (dB)
f =	frequency of transmission (MHz).

Hata model for rural and open environments

The Hata model for rural environments is applicable to transmissions in open areas where no obstructions block the radio link:

$$L_O = L_u - 4.78 \cdot (\log_{10}(f))^2 + 18.33 \cdot \log_{10} f - 40.94 \qquad (7.20)$$

where

L_O =	path loss in open areas (dB)
L_u =	average path loss in urban area (dB)
f =	frequency of transmission (MHz).

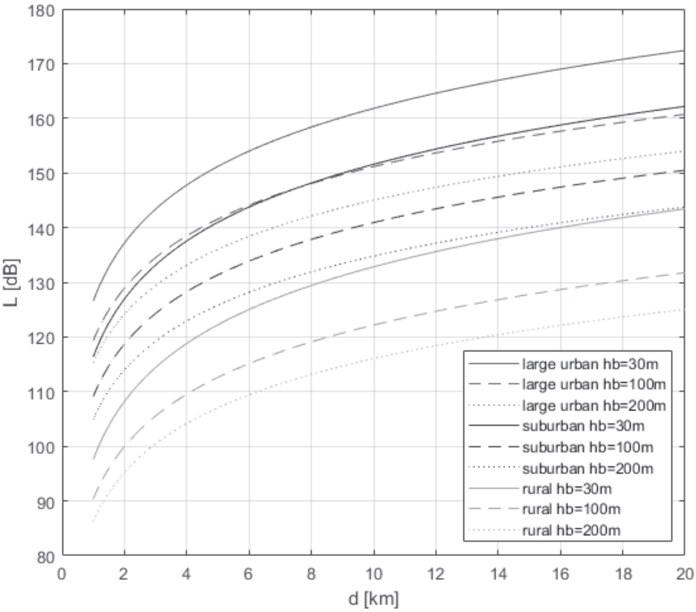


Figure 32 Okumura-Hata path losses

Figure 32 show the path losses calculated depending on the distance d with different height h_B parameters for urban, suburban and rural regions by using a Matlab program [16] for the operation frequency 1000 MHz.

COST Hata model [17]

The COST Hata model is also called the COST231 model, which is a European research cooperation project to extend the Okumura Hata model to 2000 MHz with the following range of parameters:

- Frequency: 1500–2000 MHz.
- Mobile station antenna height: 1–10 m.
- Base station antenna height: 30–200 m.
- Link distance: 1–20 km.

The COST Hata model is formulated as

$$L = 46.3 + 33.9 \cdot \log_{10} f - 13.82 \cdot \log_{10} h_B - C_H + (44.9 - 6.5 \cdot \log_{10} h_B) \cdot \log_{10} d \tag{7.21}$$

where

$L =$	the median path loss in dB
$f =$	the frequency of transmission in MHz
$h_B =$	height of base station antenna (m)
$h_M =$	height of mobile station antenna (m)
$d =$	distance between the base and mobile stations (km)
$C_H =$	correction factor gain (such as type of environment, water surfaces, isolated obstacles, etc.).

Young model [18]
The mathematical formulation for the Young model is an approximation for mobile communications planning valid for 150 MHz – 3700 MHz:

$$L = G_B G_M \frac{h_M h_B}{d^2} \beta \tag{7.22}$$

where

$L =$	path loss (dB)
$G_B =$	gain of base transmitter (dB)
$G_M =$	gain of mobile transmitter (dB)
$h_B =$	height of base station antenna (m)
$h_M =$	height of mobile station antenna (m)
$d =$	link distance (km)
$\beta =$	clutter factor and correction for all disturbances.

7.3 Antenna Array and Beam Focusing

By using an array of antenna elements with the same amplitude weights and the same phases, allocated in a line or in a planar or conformal surface, the antenna gain will increase and the beam width will be reduced, focusing the power of the electromagnetic waves more in the direction of the receiver ("beam focusing").

By using different amplitude weights and phases of these single antenna elements, however, the beam can be formed into a wanted pattern (“beam forming”) or steered into a wanted direction (“beam steering”), so that the antenna radiation direction can be tuned by Digital Signal Processing either in an analogue or digital manner, without rotating the antenna mechanically.

Depending on the allocation of the antenna elements on a line or a surface, we have the following typical antenna arrays, with normally $\lambda/2$ distance between the antenna elements:

- linear antenna arrays.
- planar antenna arrays.
- conformal antenna arrays.

For some applications in astronomy, large-scale thinned antenna arrays are also used with a much larger distance between the antenna elements in order to keep the system costs as low as possible and to achieve narrow beam width for the main lobe and moderate to relatively high antenna array gain with the trade-off of high side lobes.

MIMO: Multiple Input Multiple Output Antenna Systems

Generally both transmitter and receiver could be designed with antenna arrays consisting of multiple antenna elements. These systems are then called MIMO (Multiple Input Multiple Output Antenna Systems). If the number of the antenna elements is extremely large in comparison with the classical antenna arrays, the system is also called massive MIMO [19]. MIMO is the best method to form and steer the beams to point to a certain micro or pico station or “hot spot” within a macro mobile station cell or sector, enabling spatial diversity and high gain.

As proposed in the Nokia Bell Labs project “Future Cells” [20] massive MIMO can – in combination with solar panels – help to deploy the micro or pico base station in rural regions without additional power supply and wireline backhaul infrastructure. On the other hand, massive MIMO can also form the special beams to provide optimised, power-efficient multi-user coverage in one mobile base station cell. Massive MIMO Systems of future 5G mobile networks should be able to dynamically form multiple beams.

In order to achieve these goals, the channel measurements and channel estimations, which are the determination of the so-called precoding matrix of the amplitude weights and phases of the transmitter MIMO antenna elements, should be done prior to transmitting payload data. Generally there are two ways:

Open loop beam forming basically forms the MIMO beams without considering multipath propagation effects. If multipath propagation effects can be neglected in some environments, then the method is straightforward, and can be simply configured.

Closed loop beam forming considers the multipath propagation effects which will be first measured in all single frequency sub-bands by using the orthogonal pilot channels, by using sophisticated algorithms like Zero-Forcing to determine the precoding matrices.

To start with MIMO beam forming, we first investigate simple MIMO antenna arrays.

Antenna arrays are widely used in various applications such as mobile communications, synthetic aperture radar, medicine, sensing, imaging or radio astronomy [21] [22] [23] [24] [25] [26] [27] [28] [29] [30] [31] to enable fast and precise beam forming. Generally a high resolution is required for beam forming which leads to relatively large antenna arrays and complex signal processing systems. In some cases the systems should operate in the near field region.

The far field characteristics of antenna arrays have been thoroughly investigated [21] [22] [23] [24] [25]. The coupling effect between the array elements can be neglected for distances between the array elements larger than or equal to $\lambda/2$.

Far field and near field are commonly distinguished by the far field distance definition $r_{\min} = 2D^2 / \lambda$ with the largest dimension of the antenna array D and the wavelength λ. For the investigations for the number of the antenna array elements in x and z direction $M, N \leq 10$ and the spacings d_x, $d_z = \lambda/2$ (Fig. 33), the near field range (distance between the centre point of the array and the observation point) remains shorter than 32λ, even though the principles are valid for arbitrary antenna arrays with large numbers M and N [29].

For the investigation of the characteristics of the antenna array, the exact relation between the phase differences of the single antenna elements must be taken into account, which is also valid for the far field.

A planar antenna array (Fig. 33 and Fig. 34), which is a special case of conformal antenna arrays, is illustrated schematically in Fig. 33, where x_m and z_n are the x- and z-coordinates of the single array elements (x_m, y_{mn}, z_n) and the focal point $F\,(x_F\,, y_F\,, z_F\,)$. $\vec{r}_{mn}$ is the vector pointing to the array element from the coordinate origin (0, 0, 0) as the reference point of the antenna array. Similarly the vectors $\vec{r}_S$ and $\vec{r}$ can be defined for distances between the array element or (0, 0, 0), and any arbitrary observation point or focal point at 10λ. To be more precise, the beam steering angle is the angle with respect to the main lobe in y-direction. Neglecting the coupling effect between the array elements, the planar array can be further investigated by superimposing two orthogonal linear arrays parallel to the x axis and z axis [29] [32].

For $y_{mn} \neq 0$ the planar antenna arrays become then conformal antenna arrays (Fig. 35 and Fig. 36). The curvature of the conformal array profile, on which the single Hertzian dipoles or patch antennas are positioned, can be defined by a curvature radius r_c with corresponding centre points at y_c (concave) or $-y_c$ (convex).

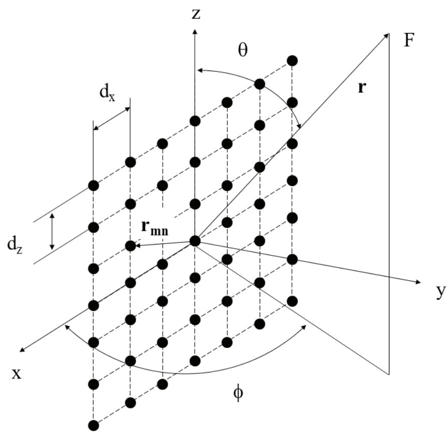


Figure 33 Planar antenna array with the focal point $F(x_F , y_F , z_F)$

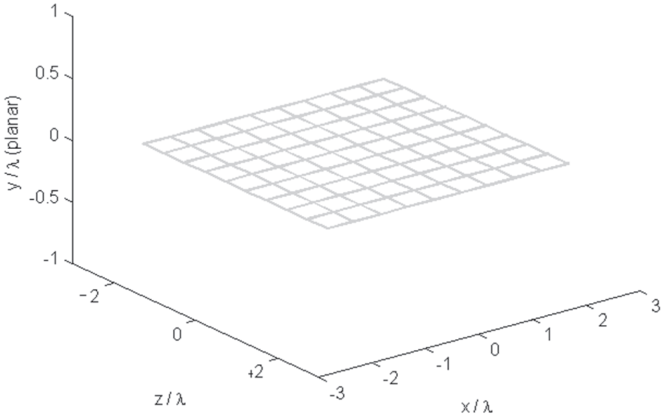


Figure 34 Planar array

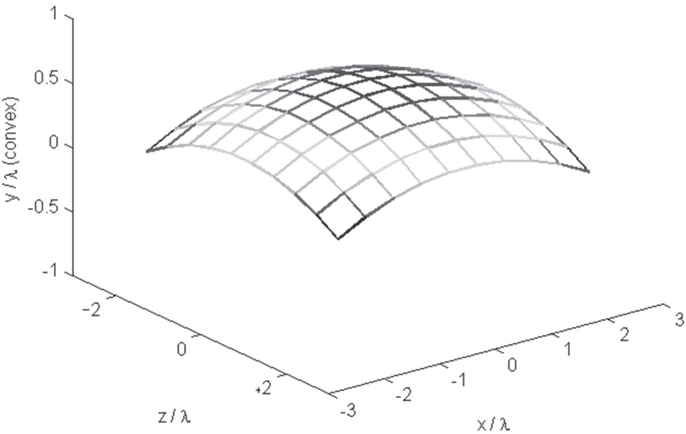


Figure 35 Convex array

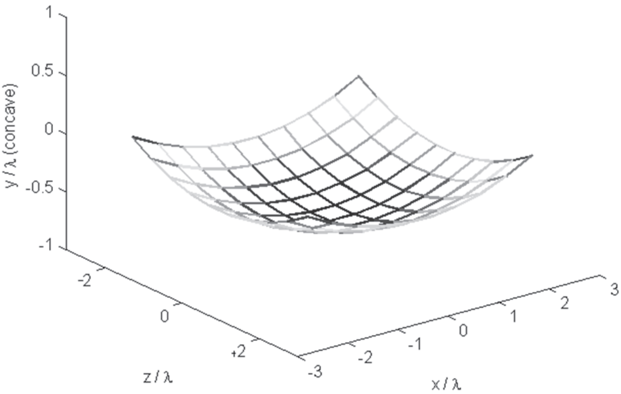


Figure 36 Concave array

$$y_c = \sqrt{r_c^2 - x_{m,\max}^2 - z_{n,\max}^2}\ . \qquad (7.23)$$

With the coordinate y_{mn} of an array element and the distance from each single array element to the observation point *P(x, y, z)* at $r = 10\ \lambda$, $\vec{r}_S = \vec{r} - \vec{r}_{mn}$ can be calculated

$$y_{mn} = \pm\left(y_c - \sqrt{r_c^2 - x_m^2 - z_n^2}\right) \qquad (7.24)$$

$$x_r = r \cdot \sin(\theta) \cdot \cos(\phi) \qquad (7.25)$$

$$y_r = r \cdot \sin(\theta) \cdot \sin(\phi) \qquad (7.26)$$

$$z_r = r \cdot \cos(\theta) \qquad (7.27)$$

$$r_S = \sqrt{(x_r - x_m)^2 - (y_r - y_{mn})^2 - (z_r - z_n)^2}\ . \qquad (7.28)$$

The near field characteristics of the conformal antenna arrays can be analysed either by using numerical methods or by using a simple, efficient analytical method by superimposing the radiation fields resulting from all the array elements located on the conformal array surface. The total radiation pattern of the arbitrary array at the focal point *F* or point of observation *P(x, y, z)* can be obtained

$$\vec{E}(\vec{r},\theta,\phi) = \sum_m^M \sum_n^N a_{mn} \cdot \vec{E}_{mn}(\vec{r},\theta,\phi) \cdot \exp(j \cdot \alpha_{mn}) \qquad (7.29)$$

with a_{mn} as amplitude and α_{mn} as the properly chosen phase of the corresponding antenna element located at (x_m, y_{mn}, z_n) to adjust the desired beam steering angle. The elements can be Hertzian dipoles, patches or other radiators. For focusing the radiation of the antenna array, the side lobes should be reduced to a minimum. The desired narrow beam width can be achieved by using a large number of array elements $M \times N$.

For the investigation of far field radiation characteristics some simplifications are possible, i.e. radiation characteristics are reciprocally proportional to the distance r between the source and the observation point P or focal point F in the first approximation, but for the near field case either the simplified isotropic radiators or the exact near field patterns of each single radiator (dipoles or patches) located at (x_m, y_{mn}, z_n) are to be considered.

For planar arrays the beam can be formed and steered by properly choosing the amplitude a_{mn} and phase α_{mn}. For simplicity we neglect the coupling effect between the array elements for a distance larger or equal to $\lambda/2$.

By doing so, the beam forming task is to achieve φ_0 and θ_0. This can be done for instance by properly choosing the amplitude parameters a_m and phases α_{mn}. To steer the beam to a desired direction in the near field, i.e. if the focus F is near the sources, say in a distance $< 30\lambda$, the connecting lines between the focus point and the array elements are no more parallel than in a far field case, so that the easy far field model leads to wrong estimations of the phases of the antenna elements. The phase shifts from the array elements to the focus point $F(x_F, y_F, z_F)$ must be calculated by considering the exact distance between the array element and the focal point, not only the steering angle like in the far field case. By using the equations (7.23) – (7.29) the vector $\boldsymbol{r}_F$ pointing to the focal point from the coordinate origin and then the phase differences α_{mn} can be calculated:

$$\alpha_{mnF} = k_0 \sqrt{(x_r - x_m)^2 - (y_r - y_{mn})^2 - (z_r - z_n)^2} \, . \qquad (7.30)$$

$$\alpha_{mn} = -(\alpha_{mnF} - k_0 \cdot r) \, . \qquad (7.31)$$

In all simulations the phases will be calculated by using this exact near field assumption.

There are many well-known steering techniques for beam forming and side lobe suppression, such as the Binomial and Chebyschev illumination coefficients, defined as weighting function $W(m)$ and $W(n)$ (e.g. [23]) . By using the Binomial distribution the side lobe level can be reduced without ripple, but the beam width is also increased at the same time. Chebyschev amplitude weights lead to better results both in terms of narrow beam width and reduced side lobe level which are eminently important for near field focusing.

The side lobe suppression is improved by using Chebyschev illumination coefficients for smaller steering angles, but the major side lobe increases for the beam steering angles of 15° to 45° [29] [31], especially for near field cases.

The significant increase of the side lobes is disadvantageous for signal transmission, sensing or imaging applications. These could lead to errors and interferences due to multipath propagation effect. In [29] a simple but effective way was proposed to reduce these side lobes further by using an additional asymmetrical amplitude illumination with the new coefficient s, which was also further improved in [31] for θ_0.

In order to simulate the near field focusing problems quickly, an analytical method is preferred. To ensure this, the analytical method was compared with commonly known sophisticated, but time-consuming numerical simulation techniques like CST [33] and HFSS [34] in [31]. The relatively small differences between these solutions confirm the low coupling effect for the parameter settings also used in this paper, and justify the proposed analytical method to investigate the near field focusing problems as good approximations.

Firstly, different 10×10 conformal Hertzian dipole arrays (concave, convex and planar) are analysed for different beam steering angles, i.e. $\theta_0 = 75°$ and $\varphi_0 = 60°$, with the focal point designed to be at a distance of $10\ \lambda$ to show the typical radiation characteristics, with homogeneous amplitude weights, Binomial, and with the asymmetrical weighting function proposed in [29] [31] in combination with Chebyschev to provide an optimum side lobe suppression, especially for the asymmetrical beam steering (Fig. 37 to Fig. 39).

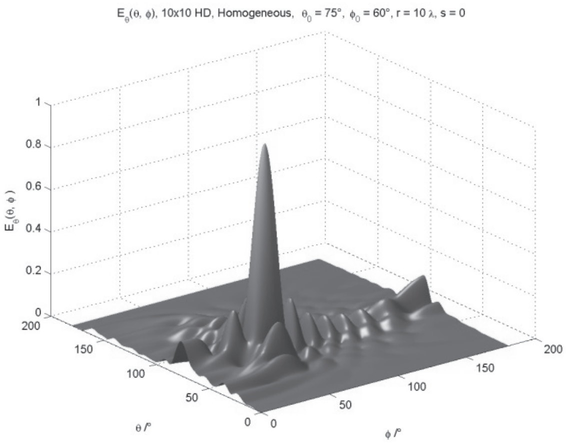


Figure 37 Near field radiation pattern, homogeneous

In Figure 37 we see the near field radiation pattern of a convex 10×10 Hertzian dipole array with homogeneous amplitude weights for a beam steering angle of 90°- θ_0=15° and 90° - φ_0 = 30°with the focal point at a distance of 10 λ, r_c=9 λ .

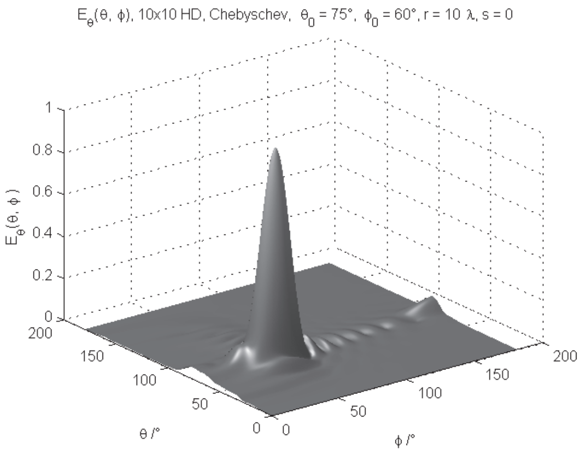


Figure 38 Near field radiation pattern, Chebyschev

In Figure 38 we see the near field radiation pattern of a convex 10×10 Hertzian dipole array with Chebyschev amplitude weights for a beam steering angle of 90°- θ_0=15° and 90° - φ_0 = 30°with the focal point at a distance of 10 λ, r_c=9 λ. The side lobes have been suppressed in comparison with Figure 37.

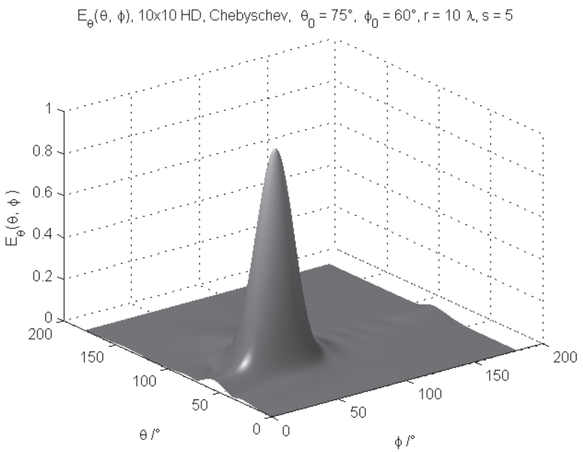


Figure 39 Chebyschev with additional asymmetrical weight

In order to further suppress the side lobes, the Chebyschev amplitude can be combined with additional asymmetrical amplitude weight s = 5. Figure 39 shows the improved side lobe suppression in comparison with Figure 38. At the same time the beam width became slightly broader.

Figure 40 illustrates the radiation pattern of planar, concave and convex antenna array with θ_0=90° and φ_0 = 90° for homogeneous amplitude weights in the azimuth φ.

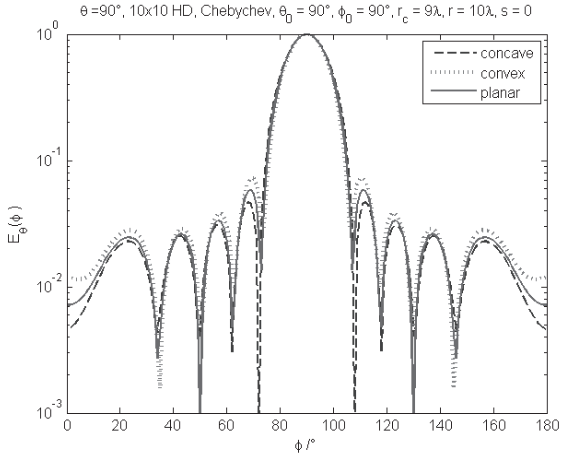


Figure 40 Comparison of different conformal arrays, azimuth

Figure 41 shows the radiation pattern of a planar, concave and convex antenna array with relatively large steering angles θ_0=75° and φ_0 = 60° in the elevation, leading to an increase of side lobes, even though the Chebyschev amplitude weights are applied.

In order to further optimise the side lobe suppression characteristics, the additional parameter *s* [31] can be determined individually in combination with different other antenna array design parameters, such as the number of elements and the near field focal point, and can be used to further optimise the overall side lobe suppression. The relatively high side lobe level of homogeneous amplitude weights can be reduced almost completely by Binomial amplitude weights, but the beam width is increased significantly, whereas by using Chebyschev with additional asymmetrical amplitude weights, the beam width remains narrow and the side lobes are suppressed efficiently. By doing so, the best near field focusing results can be achieved.

The side lobe suppression in Figure 42 is significantly improved, even though the beam width increases slightly. The desired beam width can be reached or further improved if, correspondingly, more array elements are used.

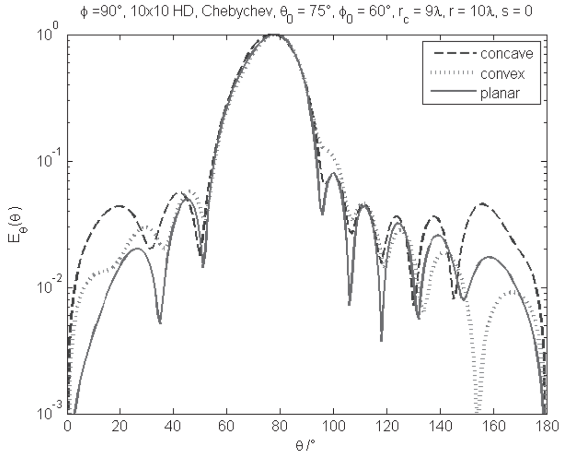


Figure 41 Comparison for different conformal arrays, elevation

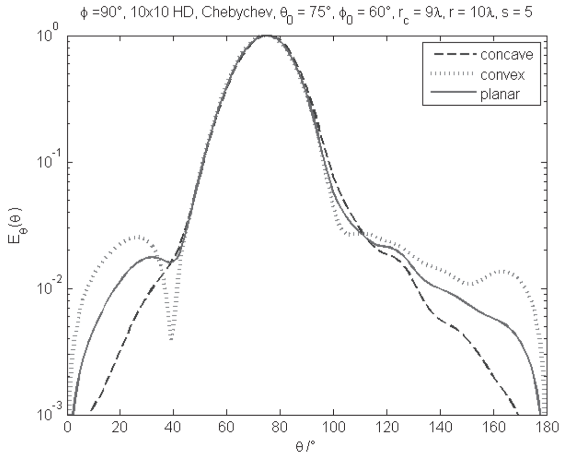


Figure 42 Comparison for different conformal arrays, elevation

The near field beam steering characteristics of conformal antenna arrays like planar, convex and concave antenna arrays are investigated and compared with each other by using an efficient analytical method.

Amplitude tapering by using the new asymmetrical weighting coefficients, in combination with the Chebyschev amplitude weights, efficiently reduces the side lobes over larger beam steering angles for the near field radiation characteristics. Depending on the beam forming requirement, this method can be combined with a well-known side lobe suppression technique (Binomial or Cheyschev) to achieve certain radiation characteristics and the desired narrow beam width by using a larger number of array elements. The method presented in this paper can be used to quickly provide the optimum parameters for dynamic beam forming problems for adaptive planar or conformal antenna arrays. The same method could be used to analyse generally volumetric antenna array problems in future projects.

7.4 MIMO and Beam Forming

Massive MIMO [19] is the most prominent technology to achieve the 5G requirements for high channel capacity and efficiency. 5G strives to increase the throughput of the wireless channel and in order to do that it requires the utilisation of all possible radio access technologies (Multiple RATs), and therefore all multiplexing techniques as earlier mentioned (FDMA, TDMA and SDMA). The modern communication technology LTE-A is able to achieve the best possible utilisation of TDMA and FDMA through OFDM combined with carrier aggregation and also managed to reduce the effect of ISI by using cyclic prefixes. By performing channel measurement and channel estimation with pilots, in combination with zero-forcing algorithm, the precoding matrix can be determined in a multi-path propagation environment. Channel estimation plays an important part in an OFDM system. It is used for increasing the capacity of orthogonal frequency division multiple access (OFDMA) systems by improving the system performance in terms of bit error rate (BER). To facilitate the estimation of the channel characteristics, LTE uses cell-specific reference signals (pilot symbols) inserted in both time and frequency. These pilot symbols provide an estimate of the channel at given locations within a sub-frame. Through interpolation, it is possible to estimate the channel across an arbitrary number of sub-frames. The pilot symbols in LTE-A are assigned positions within a sub-frame depending on the cell identification number and which transmit antenna is in use [35].

Normally the MIMO channel can be described by the model:

$$\mathbf{y} = \mathbf{G}\,\mathbf{x} + \mathbf{n}.$$

The received signal $\mathbf{y}$ is disturbed by noise and multipath propagation. If the multipath propagation effect can be neglected, then the open loop beam forming can be easily done, as discussed in the last chapter with amplitude weights and phases.

Generally the channels are measured by considering the interferences and noises. They are then estimated. The input signals **x** are pre-coded after the channel measurement and channel estimation, so that the received signals **y** will be exactly as wanted, i.e. the MIMO beams will exactly point at the user terminals (user 1, user 2 and user 3). Generally speaking the users could be single users or devices, or also pico-cell base stations which cover a hot-spot with many mobile users.

The channel estimation algorithm extracts the precoding matrix for a transmit/receive antenna pair from the received grid based on the channel measurement data. The least squares estimate of the channel transfer function at the pilot symbols can be calculated. One method is the zero-forcing ZF algorithm, the other is the maximum ratio algorithm [19]. Zero-forcing eliminates the interferences among the multiplexed signals, whereas the maximum ratio method tries to maximise the total channel capacity. The least squares error (LSE) are then averaged to reduce any unwanted noise from the pilot symbols since the mean of noise is zero. That is fast and easy to do in the case of SISO but once the system is expanded and becomes Multi User MIMO (Fig. 43) the channel estimation requires the adaptive filter to take many iterations to get a solid channel estimation. The coordination for the pilots also becomes more complicated to ensure that they remain orthogonal and unique to enable SDMA.

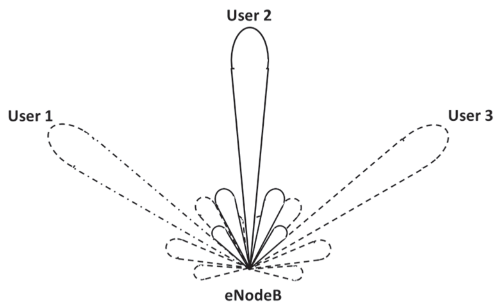


Figure 43 Multi-user MIMO system

The uplink and downlink capacity C^{UL} and C^{DL} can be estimated also by using Shannon’s formula where ρ_x is the SNR and G is the channel matrix [19].

It can be considered as multiple MISO in DL and multiple SIMO in UL where there is no correlation between the UEs. The extension to the Multi-user MIMO is the Massive MIMO and it is preferred since it saves the UE a lot of burdens and loads of the complexity of the eNodeBs.

Generally MU-MIMO can be implemented in two ways:

1) Complete channel measurement and channel estimation for multiple users, either pseudo-dynamically or statically, for example for some fixed pico-cells by using 4x16 arrays (4 rows and 16 columns are necessary, as described in detail in [19]).

2) Subarrays can be used to form the single beams, for example 4 times 4x16 or 4x64 antenna arrays, each for one single beam with sufficient side lobe suppression, in order to avoid any interference with the other beams.

Obviously the second method is more flexible, less complex and requires less computation time for the channel measurement and channel estimation. If open loop beam forming can be applied, then the different beams can be formed very quickly.

Figure 44 shows a large scale planar antenna array of 1×16 antenna elements. For the simulation, the 2.5 GHz ISM band is used for 1x16 elements.

In Figure 45 an example of open loop sub-array beam forming is illustrated, by using two sub-arrays of 2×16 antenna elements.

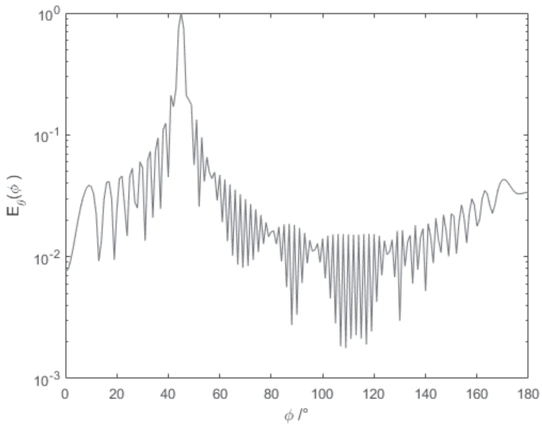


Figure 44 MIMO 1x16 antenna array (steering angle 45°)

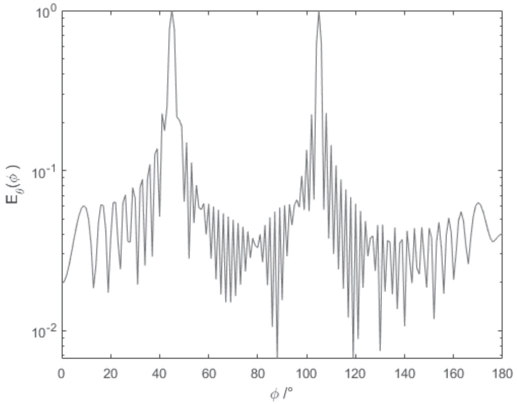


Figure 45 MIMO 2x16 sub-array beam forming (45°, 105°)

CHAPTER 8

WIRELESS TECHNOLOGIES

As discussed in previous chapters, the multimedia information contents can be digitised, coded and modulated to the physical transmission channel, either by wireline or wirelessly, and transmitted to the receiver (see for example [3] [36]). Many communication technologies utilise electromagnetic waves, guided dielectric waveguides or propagation in free space.

Depending on the frequencies used, transmission environments (including specific propagation conditions for indoor, outdoor, rural, suburban and urban regions), coverage area (Global Area Network GAN, Wide Area Network WAN, Metropolitan Area Network MAN and Local Area Network LAN) and other network planning preconditions, different technologies could be the proper choices, as will be discussed in this chapter.

8.1 WLAN, WiFi

WLAN (Wireless Local Area Network) or WiFi (Wireless Fidelity) [37] are different names for the same technologies, and use the license-free ISM band (Industry, Science and Medicine) at 2.4 GHz and 5.8 GHz, even though in different countries the spectrum usage and channel allocation are different. WLAN is basically used in the autonomous local area networks, like campus networks of enterprises, universities, airports and railway stations, but also at private homes for internet services.

WLAN can be very flexibly implemented within the defined operating area. Besides the infrastructure mode, with one or several central access points to which all mobile devices are connected to, the so-called ad-hoc networks without central access points and without previous planning are also possible.

In the infrastructure mode of WLAN, all terminal stations (STA) are connected wirelessly to the access point. Each of these groups connected to one single access point is called a Basic Service Set (BSS). The access point is connected via the distribution system and portal to the wired network or the internet gateway. The distribution system as interconnection network can form one logical EES (Extended Service Set Network) based on several BSS.

In the ad-hoc mode, two mobile devices can communicate directly with each other within a limited range. The terminal stations (STA) will establish a wireless connection without a central access point, and form the so-called Independent Basic Service Set (IBSS).

One of the major advantages of WLAN is the easy installation and rollout without cabling or wiring difficulties for the connection to user devices like Ethernet, and without difficult construction works in the existing buildings due to new cabling, and without affecting the existing firewalls which should not be modified normally.

Ad-hoc modes can be used during natural disasters like earthquakes or fires, where the telecommunications infrastructure could be destroyed. In this case the emergency services or fire fighters can use ad-hoc WLANs besides other technologies like ad-hoc TETRA or satellite phones.

Nevertheless, typically very narrow bandwidth is available compared to wireline networks. In the ISM band (industry, science and medicine), on one hand the frequency usage is license-free, on the other hand the frequencies are shared by many users, meaning also, that the total capacity for each single user or device may be strongly reduced depending on traffic load. There is no guaranteed quality of service.

There are many proprietary technical solutions, especially for higher bit-rates. These are specified and standardised by the IEEE 802.11 series. Besides the technical recommendations, the spectrum regulations and restrictions like Electromagnetic Compatibility EMC of different countries have also to be considered. For example the WLAN power limit is 100 mW in Europe, whereas in the USA 1 W is allowed. Low power consumption is also fundamental for battery lifetime.

At the moment, WLAN is not designed for global, seamless operation, even though WLAN technology, operated in license-free ISM bands at 2.4 GHz and 5.8 GHz or licensed frequency bands, will be part of the 5G Multiple Radio Access Technologies (Multiple RATs) in the future, together with LTE (4G), mm waves (low band up to 7 GHz or high band from 24.25 – 52.6 GHz) and GSM as the voice telephony backup.

By using the DMT (Discrete Multi-Tone) technique, the modulation scheme at each channel of the used WLAN spectrum will be automatically adapted, depending on the electromagnetic interferences of other transmitters or machines, as well as noise. That is the reason, why the WLAN transmission is robust, independently from the interferences.

Normally the wireless LAN is a combination with the xDSL (Digital Subscriber Lines), consisting of shielded or unshielded twisted pair copper wires, which have existed for about 100 years. Even though the twisted pair copper wire infrastructure will be replaced step by step by the significantly more powerful optical fibres (Fibre To The Home FTTH), in some applications, xDSL remains a sufficient technology, especially because this technology enables the protection of past and future communication network investments in an optimised manner.

Confidentiality and network security are very important issues in all areas of communication technologies, to guarantee the privacy of the communication partners, the encryption of data and the reliability of the transmission.

In the two bottom OSI-layers (cf. Chapter 1) PHY (physical layer) and MAC (medium access control layer), specific WLAN protocols for layer 1 (physical layer convergence protocol PLCP with physical medium dependent sublayer PMD) and layer 2 (logical link control LLC) address these issues.

The physical layer applies basically to three types of transmission techniques for WLAN: FHSS (Frequency Hopping Spread Spectrum) and DSSS (Direct Sequence Spread Spectrum) in the ISM bands and in the infrared band 850 nm – 950 nm. The maximum radiated power is limited to 1 W in the USA and 100 mW in the EU. In the MAC (Medium Access Layer) different traffic services can be provided: mandatory asynchronous data service based on best effort and optional time-bounded service using PCF (Point Coordination Function) – which is somewhat like CS (Circuit-Switched) services.

In order to allow collision free media access DFWMAC-DCF (Distributed Foundation Wireless Media Access Control) CSMA/CA (Carrier Sensing Multiple Access / Collision Avoidance) is specified by using different so-called IFS (Inter Frame Spacings) in order to differentiate the priorities of the traffic services and to avoid collisions (see detailed descriptions in [36] [37]):

- SIFS (Short Inter Frame Spacing): highest priority, for ACK, CTS, polling response and shortest IFS.
- PIFS (PCF Point Coordination Function IFS): medium IFS, medium priority, for time-bounded service using PCF.
- DIFS (DCF Distributed Coordination Function IFS): lowest priority, largest IFS, for asynchronous data service.

Collision avoidance is enabled when these different IFS are used to manage the waiting time of traffic services with different priorities. For traffic services with the same priorities, the randomised "back-off" mechanism will be used to assign different randomised numbers of unit waiting times to keep the media access order of the consecutive data packets.

Collision can also happen, if the mobile terminal stations can not directly detect each other due to the limited transmission power, so that both stations try to send the data packets at the same time ("hidden station problem"). In order to solve this problem and to avoid the collision of the hidden stations, some useful signaling packets are defined:

- ACK packet for acknowledgements (not for broadcasts).
- RTS ready to send.
- CTS confirm to send.

Each station which is ready to send data first starts sensing the medium (Carrier Sense based on Clear Channel Assessment CCA). If the medium is free for the duration of an Inter-Frame Space (IFS), depending on different service-type priorities, the station can start sending. If the medium is busy, the station has to wait for a free IFS and then wait for a random back-off time period (multiple of slot-time in order to avoid collision). If another station occupies the medium during the back-off time of the station, the back-off timer stops, so that the residual back-off time will be considered for the next contention period.

In order to solve the hidden station problems, the station first sends the RTS (Ready to Send) message with a reservation parameter after waiting for the DIFS (Distributed Coordination Function IFS). The reservation parameter determines and contains the amount of time the data packet needs for transmission in the medium. The acknowledgement message CTS of the receiver, after receiving the SIFS, confirms the readiness to receive. All stations store medium reservations distributed via RTS and CTS at the same time. Especially in the case of a hidden station situation, when some senders and receivers cannot receive the RF signal from each other, the collisions can be avoided. The sender can now send data immediately, with the successful reception acknowledged by an ACK message.

In order to guarantee the collision free media access control, the timing clock of all mobile terminal stations have to be synchronised with the central infrastructure clock.

The power management of mobile terminal devices is always a crucial issue in all wireless communications networks. In the case of WLAN, the sleep-mode helps to let the mobile terminals regularly reduce the power consumption and increase the power efficiency, i.e. periodically go sleep and wake up again without missing a message.

Many systems are available worldwide, integrated into laptops and mobile phones. And there are many vendors. The most-used WLAN standards are:

- IEEE 802.11b, bandwidth 22 MHz, 2400 MHz – 2483.5 MHz.
- IEEE 802.11g, bandwidth 20 MHz, 2400 MHz – 2483.5 MHz.
- IEEE 802.11n, bandwidth 20/40 MHz, 2400 MHz – 2483.5 MHz and 5150 MHz – 5725 MHz.
- IEEE 802.11a, bandwidth 20 MHz, 5150 MHz – 5725 MHz.
- IEEE 802.11ac, bandwidth 20/40/80/160 MHz, 5150 MHz – 5725 MHz.

With a channel spacing of 5 MHz for IEEE 802.11b/g/n (ISM band 2.4 GHz) the frequency for the channel k is numbered as:

$$f_k = 2412 \text{ MHz} + (k - 1) \times 5 \text{ MHz}. \tag{8.1}$$

In Europe channels 1 – 13 are used, whereas in the USA only channels 1 – 11 are used. Since the bandwidth of each channel is 22 MHz, the recommended non-overlapping channels are 1, 7, 13 in Europe, and 1, 6, 11 in the USA.

With a channel spacing of 20 MHz for IEEE 802.11a/n/ac (ISM band 5 GHz) the frequency for the channel k is numbered as:

$$f_k = 5000 \text{ MHz} + k \times 5 \text{ MHz}. \tag{8.2}$$

The operating, non-overlapping channels are 36, 40, 44, 48, 52, 60, 64 (5180 MHz – 5320 MHz) and 149, 153, 157, 161 (5745 MHz – 5805 MHz).

RSSI (dBm)	Data Rate for IEEE 802.11 g							
	SNR (dB)							
	4	5	6	7	8	9	10	11
-94	1	1	1	1	1	1	1	1
-91	1	1	2	2	2	2	2	2
-87	1	1	2	2	5.5	5.5	5.5	5.5
-86	1	1	2	2	6	9	9	18
-84	1	1	2	2	6	9	9	18
-82	1	1	2	2	6	9	11	24
-80	1	1	2	2	6	9	11	36
-75	1	1	2	2	6	9	11	48
-71	1	1	2	2	6	9	11	54

Table 9 SNR, RSSI, Data Rates

On the other hand, the license-free ISM band is shared by many users, so that interference cannot be avoided. Therefore, there is no guarantee of quality of service. The total available channel capacity does not depend on the received signal strength indicator RSSI in dBm, but also on the SNR (signal to noise ratio), or better SINR (signal to interference and noise ratio). One example for the achievable data rates depending on RSSI and SINR/SNR for IEEE 802.11g planning results is shown in Table 9 [38]. The data rates of WLAN standard IEEE 802.11a with a channel bandwidth of 16.6 MHz can achieve 6, 9, 12, 18, 24, 36, 48 or 54 Mbps by using an OFDM modulation scheme, again depending on the SNR or channel condition (Table 9). The coverage could be maximally 100 m outdoor and 10 m indoor, according to the calculation by using RF3D WiFi Planner [38].

The WLAN Standard IEEE 802.11g is an improved successor of the IEEE 802.11b standard. The data rates can be increased to 54 Mbps at 2.4 GHz by using OFDM.

IEEE 802.11n can achieve higher data rates above 100 Mbps by some optimisations of the PHY and MAC layers. By additionally using MIMO antennas (Multiple Input Multiple Output) up to 600 Mbps are achieved. The 5 GHz band can be used optionally, with 20 MHz and 40 MHz bandwidths and modulation schemes up to 64QAM.

The advanced WLAN standard IEEE 802.11ac also uses 5 GHz, with a much larger spectral bandwidth of 20 MHz, 40 MHz, 80 MHz or 160 MHz and higher modulation scheme up to 256QAM, in order to achieve bit rates up to 867 Mbps – or in combination with 8x8 MIMO even up to 6.9 Gbps.

The WLAN standard IEEE 802.11ad uses mm waves of 60 GHz for short distances up to 10 m. Therefore, a much higher spectral bandwidth of 4 – 7 GHz is made possible by using OFDM, as are much higher bit rates of 5.3 Gbps. 6.7 Gbps can be achieved by using 64QAM.

8.2 Bluetooth

One of the most famous WPAN (Wireless Personal Area Network) technologies is Bluetooth, standardised as the IEEE 802.15-1 recommendations [39].

The Bluetooth universal radio interface was initially proposed and developed by Ericsson. It enables ad-hoc wireless connections and is especially useful for connecting computers, mobile phones, printers, headsets and other mobile terminals. The coverage range is relatively small, less than 10 m, and correspondingly the power consumption is low.

IEEE 802.15-1:
The WPAN Bluetooth standard is defined by the IEEE 802.15-1. The data rate for synchronous voice service is connection-oriented 64 kbps, whereas asynchronous, connectionless data service could be up to 433.9 kbps bidirectional symmetrical, and 723.2 / 57.6 kbps asymmetric. The transmission range is limited to 10 m. With special transceivers with directional antennas, the coverage range can be increased to 100 m. Because Bluetooth utilises the license-free 2.4 GHz ISM band interference is possible and no quality of service can be guaranteed. Errors can be corrected by using ARQ (automatic retransmission request) and FEC (forward error correction).

Bluetooth devices have been implemented into many systems and products provided by almost all vendors for flexible wireless connections. The free ISM band, a simple system, ad-hoc networking and peer-to-peer, scatter networks make the Bluetooth technology very attractive for PAN (personal area network) applications, despite the possible interference in the ISM band and relatively large latency. In comparison with WLAN, Bluetooth does not apply inter frame spacing to avoid collisions. The following standards try to further develop the Bluetooth technology and improve the interference tolerance.

Bluetooth utilises the license-free 2.4 GHz ISM band. Even though the bit rate of about 1 Mbps is relatively low, Bluetooth can enable basic voice and data transmission, and replaces wired connections.

IEEE 802.15-1 specifies Bluetooth applications in the 2.4 GHz ISM band, i.e. 79 RF channels with 1 MHz carrier spacing, can be allocated from 2402 MHz (channel 0) to 2480 MHz (channel 78). The radio frequency transmit power is about 1 mW - 100 mW. Contrary to WLAN/WiFi, frequency hopping FHSS (Frequency Hopping Spread Spectrum) in combination with the TDD (Time Division Duplex – for transmitting and receiving) is used to transmit the data. Frequency hopping with 1600 hops/s between certain defined frequencies or a pseudo random hopping sequence is determined by a master.

Bluetooth services can be voice service links SCO (synchronous connection oriented or circuit oriented connection), supported by FEC (forward error correction) without retransmission, enabling 64 kbps duplex point-to-point bidirectional connections, typically for real time telephone conversations. Bluetooth data link ACL (asynchronous connectionless or packet oriented) also provides point-to-multipoint symmetrical 433.9 kbps or asymmetrical 723.2/57.6 kbps data connections.

Each group of Bluetooth devices connected in an ad-hoc manner is called a pico-net, with one Bluetooth device as master and the other devices as slaves or parked members. The master determines a pseudo random frequency hopping pattern or hopping sequence, and all the slaves have to synchronise to this sequence, in order to join this ad-hoc pico-net.

Each pico-net has one master and up to seven simultaneous slaves. Additionally more than 200 more could be parked. The master unit provides the clock and device ID which consists of 48 bits and also determines the hopping sequence. Together with the clock and the hopping sequence given by the master, the pico-net members can then be synchronised.

The different statuses of the members are coded by using a) 3 bits for active members and b) 8 bits for parked members. If two or more pico-nets are overlapping, then the Bluetooth devices can join one or another pico-net by using different hopping sequences and synchronise to the clocks of different pico-net masters.

Classical voice services will be supported by the SCO (Synchronous Connection Oriented) with synchronous, periodic single slot packet assignment to provide the 64 kbps full-duplex, point-to-point connections. The data services will be enabled by using the ACL (asynchronous connectionless) mode to transmit the variable packet sizes, by combining one, three or five slots. This is because of the typical nature of the asymmetrical bandwidth requirement of the data communications, for example data download, audio and video streaming. In some applications point-to-multipoint connections will be useful, like broadcast services.

The slow frequency hopping technique, with hopping patterns determined by a master, is especially advantageous in the case of interferences on certain frequencies. The different hoping patterns help to separate the different pico-nets.

In the case of ACL data communications the retransmission technique can be used to correct for the transmission errors. For the synchronous mission critical voice conversation services, retransmission will increase the latency, therefore FEC (as discussed in Chapter 3) will be more efficient, in order to correct for the transmission errors.

A Bluetooth pico-net can handle 8 active members AMA and more than 200 passive members. The passive standby members are parked without actively transmitting data, then sniff periodically – but not at each time slot – the pico-net. If necessary, these passive members search for wanted devices, inquire for access and if possible will be connected to a specific device of the pico-net. On the other hand, if active members end the data transmission, they will be parked, release the AMA (active member address) and get a PMA (passive member address), so that other passive members can join the pico-net actively. Obviously the passive members do not send and receive data packets in each time slot, so the power consumption will be much lower. The typical average power consumptions for a room temperature of 20°C for Bluetooth are the following [36]:

Active members:	
SCO connection HV3 (1s interval Sniff Mode) (Slave)	26.0 mA
SCO connection HV3 (1s interval Sniff Mode) (Master)	26.0 mA
SCO connection HV1 (Slave)	53.0 mA
SCO connection HV1 (Master)	53.0 mA
ACL data transfer 115.2kbps UART (Master)	15.5 mA
ACL data transfer 720kbps USB (Slave)	53.0 mA
ACL data transfer 720kbps USB (Master)	53.0 mA
Passive members:	
ACL, Sniff Mode 40ms interval, 38.4kbps UART	4.0 mA
ACL, Sniff Mode 1.28s interval, 38.4kbps UART	0.5 mA
Parked Slave, 1.28s beacon interval, 38.4kbps UART	0.6 mA
Standby Mode (Connected to host, no RF activity)	47.0 µA
Deep Sleep Mode	20.0 µA

Bluetooth defines an authentication and encryption process. The first step is the so-called pairing of two Bluetooth devices to initialise the connection by using a 16 byte PIN. After that the authentication is done in combination with the 128 bit link key. The 128 bit stream cypher encryption key is then produced, with which the payload data will be encrypted and transmitted on air. Encryption and authentication techniques will be discussed in Chapter 13.

8.3 Other Wireless PAN and WAN Technologies

IEEE 802.15-2 investigates the coexistence of Wireless Personal Area Networks (802.15) and Wireless Local Area Networks (802.11) in terms of interference in the ISM band 2.4 GHz. This issue has been discussed often, but up to now, there is no severe doubt that the two standards coexist pretty well.

IEEE 802.15-3 increases the data rate up to 11, 22, 33, 44, 55 Mbps. The quality of service for isochronous protocol for telephony and multimedia applications is investigated to replace the wired headsets. IEEE 802.15.3a is a further application especially for multimedia and picture transmission.

IEEE 802.15-3b further improves error correction and interoperability of MAC.

IEEE 802.15-3c adapts the Bluetooth technology to the higher frequency band 57-64 GHz to increase the data rate significantly above 2 Gbps.

IEEE 802.15-4 on the other hand tries to achieve extremely low power consumption for low data rate (less than 250 kbps) solutions like **ZigBee**, which is extremely attractive for sensors, smart badges, remote controls and home automation applications where replacing the battery would be expensive and time-consuming.

IEEE 802.16 deals with the specific applications of Broadband Wireless Access / WirelessMAN / WiMax Wireless distribution systems, also called Wireless MAN (metropolitan area network or WiMax) for the last mile, as an alternative to xDSL (digital subscriber lines). It provides 75 Mbps up to 50 km for LOS (line of sight) propagation conditions, and up to 10 km NLOS (non line of sight) propagation conditions by using the 2-66 GHz band. Even though the initial standards do not enable roaming or mobility support, IEEE 802.16e adds mobility support and allows roaming at a speed up to 150 km/h.

IEEE 802.20 or standard for Mobile Broadband Wireless Access (MBWA) uses the licensed bands < 3.5 GHz, especially optimised for IP traffic with a peak data rate above 1 Mbps per user, partially with mobility support up to speeds of 250 km/h with a coverage range up to 15 km.

IEEE 802.21 specifies the media independent handover interoperability between different 802.x and/or non-802 networks.

IEEE 802.22 defines the Wireless Regional Area Networks (WRAN), i.e. radio-based PHY/MAC, for use by license-exempt devices on a non-interfering basis in a spectrum that is allocated to TV broadcast services.

Nowadays, numerous Radio Frequency **RF Controllers** are being used in ISM bands, typically 27 MHz (EU, US), 315 MHz (US), 418 MHz (EU, possible Interference with TETRA), 426 MHz (Japan), 433 MHz (EU), 868 MHz (EU), 915 (US) MHz and 2.4 GHz. The focus is on low costs and industrial automation. Embedded RF controllers with a relatively low data rate of up to 115 kbps (serial interface) are used. The transmission range is lower than 5-100 m and the transmit power is 10-500 mW. There is no QoS guarantee. Additional processors are necessary for security functions.

RFID (Radio Frequency Identification) [40] is also very popular for asset management, warehouse and supermarket administration and access or entrance control systems. RFIDs are standardised by ISO/IEC 18000. RFIDs are designed for a relatively low amount of data (for example, 48 bit, 64kbit or 1 Mbit). Correspondingly, a low data rate of 9.6 – 115 kbps is necessary.

Generally there are two types of RFIDs: a) passive RFID tags do not contain a built-in battery, so that the sending request signal of the RFID base station or reader will be responded to by the RFID tags with the device ID of the device at the same or a different frequency with a coverage range of about 3 m; b) active RFID tags contain a battery which will respond to the device ID at a different frequency back to the RFID base station or reader and can therefore achieve a larger distance of 30-100 m.

Quasi simultaneous detection of up to 256 tags and scanning of 40 tags/s is possible. The frequencies 125 kHz, 13.56 MHz, 433 MHz, 2.4 GHz and 5.8 GHz can be utilised for the RFID applications. The connection set-up time is dependent on the medium access scheme chosen for the products; it takes typically 2 ms per device. Relative speeds of up to 300 km/h could be tolerated. RFIDs do not guarantee QoS because of the crowded ISM bands. Collisions can be avoided by using enhanced collision avoidance media access schemes. Additional data can also be transmitted. Because of the low frequencies typically used by RFIDs, no line-of sight is required. Unlike the IR or laser sensors, the RFID tags withstand difficult environmental conditions like sunlight, cold, frost, dirt, etc. Typical applications are therefore asset tracking during manufacturing, localisation of pallets and goods in transport logistics and warehouse management, customer loyalty cards for payment, automated toll collection, access control, animal identification, tracking of hazardous material, inventory control, localisation of disoriented patients and more.

Band	Regulations	Range	Data speed
120–150 kHz (LF)	Unregulated	10 cm	Low
13.56 MHz (HF)	ISM band worldwide	10 cm–1 m	Low to moderate
433 MHz (UHF)	Short range devices	1–100 m	Moderate
865–868 MHz (Europe)	ISM band	1–12 m	Moderate to high
902–928 MHz (America)			
2450–5800 MHz (microwave)	ISM band	1–2 m	High
3.1–10 GHz (microwave)	Ultra wide band	Up to 200 m	High

Table 10 Different RFID

In daily practice, the different types of RFID transponders are applied in different frequency ranges [41] (Table 10). RFID tags are made out of three parts: a microchip (an integrated circuit which stores and processes information and modulates and demodulates radio-frequency signals), an antenna for receiving and transmitting the signal and a substrate. RFID tags can be active, passive or battery-assisted passive devices. An active tag has an on-board battery and periodically transmits its ID signal. A battery-assisted passive device has a small battery on board and is activated when in the presence of an RFID reader.

Concerning security, denial-of-service attacks are always possible. Furthermore, interference from other wireless transmissions or shielding of RFID transceivers cannot be avoided. RFIDs are programmed during the manufacturing process of a product. For improved security, encryption techniques like RSA and AES can be used (cf. Chapter 13). In the near future, a totally automated asset management system will be the reality by using RFID technology in manufacturing, distribution, warehousing, transport and logistics chains. Also localisation and tracking of small children and elderly disoriented patients in hospitals can be very useful features enabled by RFID functionalities.

8.4 Satellite Communications

Satellites communications help to efficiently cover large areas of the earth´s surface almost seamlessly, from vessels on the oceans to mountains and deserts, everywhere where communications infrastructure cannot be set up efficiently and quickly, if at all. Depending on the applications, coverage area and distance between the satellites and the earth, different satellite systems with different frequencies will be used. Concerning the distances, the following major categories of satellite communication systems exist (see, for example, [36] [42]):

GEO (Geostationary Earth Orbit Satellite Communications) with an orbit at 36,000 km altitude, synchronous to the earth rotation. Due to the large distance to the earth's surface, the latency is extremely high.

MEO (Medium Earth Orbit Satellite Communications) with an orbit at about 10,000 km altitude.

LEO (Low Earth Orbit Satellite Communications) with an orbit at, for example, 100-1500 km. Due to the small distance to the earth's surface, the latency is relatively small in comparison with the GEO systems. At the same time, the connection to the LEO satellite must be adjusted continuously, because the LEO orbit is not synchronous to the earth's rotation.

Band	**Frequencies used**
HF-Band	1.8 - 30 MHz
VHF-Band	50 - 146 MHz
UHF-Band	0.43 - 1.3 GHz
L-Band (GEO, LEO)	1.53 - 2.7 GHz
S-Band (GEO, LEO)	2.7 - 3.5 GHz
C-Band	DL 7.25 - 7.745 GHz, UL 7.9 - 8.395 GHz
X-Band	DL 3.7 - 4.2 GHz, UL 5.925 - 6.425 GHz
Ku-Band (Europe) (GEO, LEO) (Broadband)	DL FSS 10.7-11.7 GHz, DL Telecom 12.5-12.75 GHz DBS 11.7 - 12.5 GHz, 17.3 - 18.1 GHz UL FSS and Telecom 14.0 - 14.8 GHz
Ku-Band (America) (GEO, LEO) (Broadband)	DL FSS 11.7 - 12.2 GHz DBS 12.2 - 12.7 GHz, 17.3 - 17.8 GHz UL FSS 14.0 - 14.5 GHz
Ka-Band (GEO, LEO)	18 - 31 GHz

Table 11 Frequency ranges for satellite communications

In Table 11 the abbreviations are:

DL: downlink.
UL: uplink.
DBS: direct broadcasting satellite.
FSS: fixed satellite service.

Besides the links to GEO, MEO and LEO, inter-satellite links also play an important role for the efficient utilisation of the uplink and downlink satellite communications channel, and in the reduction of the delay.

In Table 11 all popular satellite communication systems are illustrated with the above-mentioned applications, the corresponding radio frequency bands and the orbits [42]. Major applications of satellite communication are traditionally weather satellites, earth observation, radio and TV broadcast satellites, military satellites, satellites for navigation and localisation (e.g., GPS, Galileo, Glonass), telecommunication satellite for global telephone connections, backbones for global networks (especially for connections for communication in deserts or regions with underdeveloped telecommunications infrastructure), maritime fleet communications and extension or backup for the global cellular mobile communication systems like GSM, UMTS, LTE and 5G.

Due to the limited spectrum and large delay, especially concerning the GEO with more than 30,000 km propagation delay corresponding to one-way latency of about 100 ms or total uplink and downlink delay of some 200 ms, which is definitely larger than the ITU-T delay limit of 120 ms for terrestrial voice telephony applications, the satellite system extensions or backups are not primarily designed for broadband application or mission-critical applications with low-delay requirements, but rather for coverage in areas with poor or no infrastructure, like desert regions or maritime vessel communications. Therefore L-Band and S-Band are used for telephony services while Ku-Band and Ka-Band are better for TV broadcasting and broadband internet services because of their larger available bandwidth.

One solution to reduce the total delay is obviously to use LEO (lower earth orbit) satellites. The problems for LEO satellite systems for mission critical real-time services are, firstly, the large number of the LEO satellites needed and, secondly, the fact that non-synchronous satellites require a coordinated handover between satellites in order to have seamless operations, which increases the system complexity significantly.

To reduce the total delay of the user to user connection, especially for delay-sensitive real-time applications, inter satellite links (ISL) can be used to reduce the number of gateways needed and to forward the connections or data packets within the satellite network as long as possible. Only the uplink and the downlink from the end users are then needed for the connection of two mobile phone users.

The principle of satellite mobile communication systems is similar to terrestrial cellular mobile communication systems, such as GSM, UMTS and LTE as well as 5G. Gateways maintain registers' HLR (Home Location Register with the user data) and VLR (Visitor Location Register with the last known location of the mobile station) as well as SUMR (Satellite User Mapping Register). SUMR determines the satellite assigned to a mobile station by considering the positions of all available satellites. It is responsible for the registration of mobile stations, the localisation of the mobile station via the satellite's position, the requesting of user data from HLR and the updating of VLR and SUMR. By using these user data systems, the incoming and outgoing calls can be handled in the same manner by using HLR and VLR like cellular mobile networks, and the connection can be set up between the mobile users by using appropriate satellites, possibly in a combination of inter satellite links.

Handover between the satellite systems (inter-satellite handover) is also comparable with the interconnection between different mobile network operators, inter-MSC (mobile switching centre) handover and inter-BSC (Base Station Controller) handover.

For example, the so-called intra-satellite handover will change the one spot beam of the same satellite to another spot beam, so that the mobile station is still within the footprint of the satellite, but in another cell.

In the case of an inter-satellite handover, the connection of the mobile user is switched to another satellite, primarily due to the movement of the satellites or the mobile user being out of the coverage region.

In the case of an earth gateway handover, the mobile station is still in the footprint of a satellite, but the gateway leaves the footprint of the satellite.

Inter system handover is the handover from the satellite network to a terrestrial cellular network or fixed network like PSTN/ISDN, GSM, UMTS, LTE or 5G. As mentioned before, it is favorable to hand over the traffic of one mobile user to the terrestrial networks near the destination of the other mobile user, to achieve low latency and low system cost.

Satellite orbits

The satellites in circular orbits experience the attractive, gravitational force

$$F_g = m \cdot g \cdot \left(\frac{R}{r} \right)^2 \tag{8.3}$$

and the centrifugal force

$$F_c = m \cdot r \cdot \omega^2 \qquad (8.4)$$

where m is the mass of the satellite, R= 6370 km is radius of the earth, r is the distance to the centre of the earth, g is the acceleration of gravity (g = 9.81 m/s² at the earth surface) and ω is the angular velocity ($\omega = 2\ \pi f$ with f as the rotation frequency).

When gravitational force is equal to centrifugal force

$$F_g = F_c \qquad (8.5)$$

then a stable orbit with the radius to the earth´s centre point r will be obtained (Fig.46)

$$r = \sqrt[3]{\frac{g \cdot R^2}{\left(2\pi f\right)^2}}\ . \qquad (8.6)$$

Depending on the orbit or the radius to the earth centre mass point, the velocity and the satellite period can be determined, which are displayed in the following figure. The GEO satellite rotates with exactly the same period of the earth, i.e. (1-1/365)$\times$ 24 hours, so that the receiver antenna on the earth pointing to a GEO satellite can be fixed, and does not need to be permanently readjusted to keep the connection to the GEO satellite. Therefore GEO satellites are favourites for audio and video broadcasting services, i.e. DVB-S (Digital Video Broadcasting, Satellite). The disadvantage of the GEO satellite systems is the large delay, which is critical for real-time voice telephony, but not for broadcasting services.

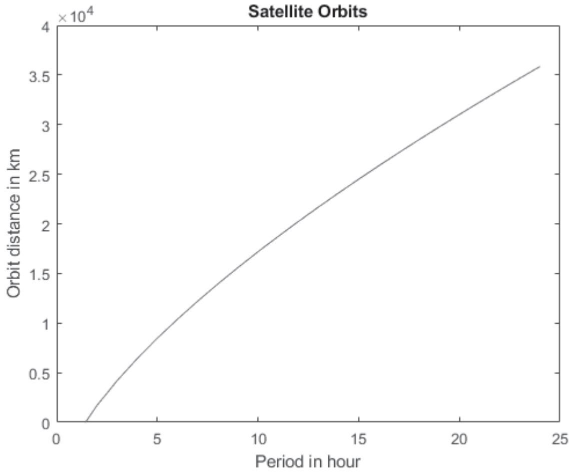


Figure 46 Satellite altitudes versus orbit period

Between the LEO and MEO, i.e. 2000 - 6000 km from the earth, the so-called Van-Allen-Belt with ionised particles does not permit satellite orbits in these regions.

As discussed before with regard to channel capacity, the parameters signal-to-noise ratio and the available spectral bandwidth determine the maximally achievable bit rates. In the free space propagation case or free LOS (line of sight), the signal-to-noise ratio again is basically determined by the transmit power of the transmitter, the gain of the transmitter antenna, the distance between transmitter and receiver or attenuation and the gain and sensitivity of the receiver. In the case of NLOS (non LOS) the received signal is disturbed additionally by multipath propagation, scattering and diffraction. Additional attenuation occurs in the case of rain, fog, snow and clouds for LOS and NLOS. These problems can be partially solved by satellite diversity, i.e. by using several visible satellites with LOS at the same time. Rain attenuation a in dB depends on the effective path length in the rain zone [42]:

$$a = \alpha \cdot L_R \tag{8.7}$$

whereas L_R is the effective path length in the rain zone depending on the rain height H_0 and the altitude of the earth station H_e as well as the earth station antenna elevation angle ε

$$L_R = \frac{H_0 - H_e}{\sin \varepsilon} \tag{8.8}$$

α is the specific attenuation in dB/km dependent on frequency and rain rate R (mm/h) in a rain zone, contained in the empirical tables for different rain climate regions worldwide (see, for example, [43]). By using a Matlab routine [44], this ITU-R model can be plotted for a frequency range from 1 – 1000 GHz (cf. Fig. 47).

The rain rates are chosen as R= 1, 5, 10 mm/h in the simulation in Figure 47. Normally the rain rates can range from 0.25 mm/h (drizzle), 1.0 mm/h (light rain), 4.0 mm/h (moderate rain), to 16.0 mm/h (strong rain) and 100 mm/h (very heavy rain). Also the attenuation due to fog and/or clouds is investigated [45]. Depending on the temperatures, the visibilities and attenuation values change (Fig. 47).

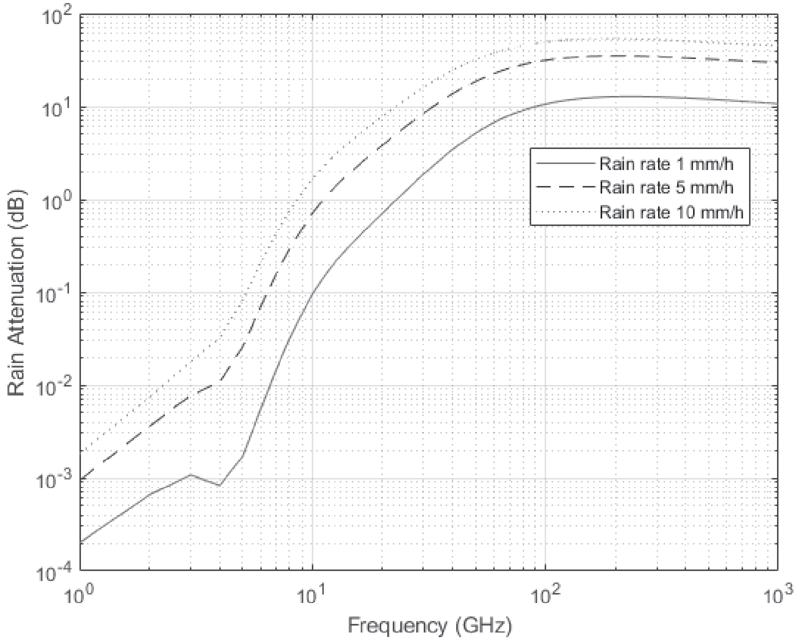


Figure 47 Rain attenuation distribution

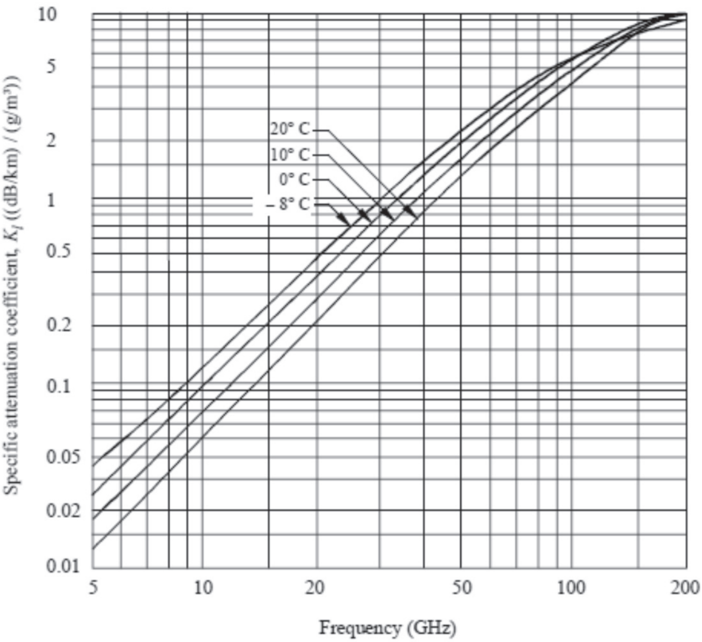


Figure 48 Attenuation due to clouds and fog, or water droplets [45]

High rain rates like in the tropical rain zones will interrupt service significantly, leading to outage and extremely high BER (bit error rates), and therefore low availability. High line-of-sight loss can be compensated for by high gain antennas with reasonable diameter.

Other factors to be considered are: gaseous absorption, cloud absorption and scintillation, leading to additional attenuation (Fig. 48).

For future Ka-Band applications, the rain attenuation was investigated in [46] and the following measures were recommended: uplink-power control (downlink not feasible), on-board processing (so that the uplink and downlink attenuation do not fully add up), adaptive data-rate allocation, adaptive power control for CDMA applications, adaptive automatic FEC, site diversity (in order to mitigate losses) and increase in the performance.

8.5 Broadcast Services

Popular applications are the unidirectional TV and radio broadcasting services, both terrestrial and via GEO satellites, in order to distribute the multimedia contents worldwide without setting up cable infrastructure everywhere, which is neither economically meaningful nor possible. In the past the analogue techniques used included NTSC, PAL and SECAM television. These are being replaced by digital formats, which typically show improved spectral efficiency and have more possibilities like automatic forward error correction by using sophisticated coding schemes and signal processing techniques. The digital video broadcasting DVB services like DVB-T (terrestrial), DVB-S (via satellites) and DVB-C (via cable, i.e. Hybrid Fibre Coaxial HFC) have been successfully implemented during the last decades. DVB-S is especially user-friendly for quick set-up at the user sites, even for maritime vessels. DVB-C is successful wherever the infrastructure allows the cable installation, for example in metropolitan areas, less so for rural or desert regions. Even though the DAB rollout is not yet completed, the digital radio DAB standard will more and more replace FM in the future [36].

DAB Digital Audio Broadcast

The DAB media access technique is the COFDM (Coded Orthogonal Frequency Division Multiplex). DAB is operated in a so-called SFN (Single Frequency Network) by using 192, 384, 768 and 1536 subcarriers within a 1.5 MHz frequency band. In the first phase, the VHF frequency bands 174 MHz – 230 MHz, one out of 32 frequency blocks for terrestrial TV channels 5 to 12 will be used, named as channel 5A - 12D. In the second phase, one out of nine frequency blocks in the L-band (1452 - 1467.5 MHz, LA - LI) will be used. The transmit power is 6.1 kW (VHF, with a coverage range of some 120 km) or 4 kW (L-band, with a coverage range of about 30 km), in order to provide data rates of 2.304 Mbps (net 1.2 – 1.536 Mbps). The modulation technique used is D-QPSK (Differential 4-phase modulation). Audio channels per frequency block are typically 6 channels of 192 kbps. So-called additional digital services can also be provided with 0.6 - 16 kbps for PAD (program-associated data services) and 24 kbps for NPAD (non-program-associated data services).

By using the Orthogonal Frequency Division Multiplex (OFDM), the payload data is transmitted on several orthogonal subcarriers with lower rates. In comparison with the classical analogue techniques, where tuning filters are necessary to select the channels with a certain bandwidth, here the orthogonality of the single carriers help to separate the channels without the necessity of filters, which means less hardware complexity and lower expenditures for DAB.

Each single subcarrier signal appears exactly at a frequency where all other subcarriers equal zero. In Figure 49 the subcarrier under consideration is marked as solid line.

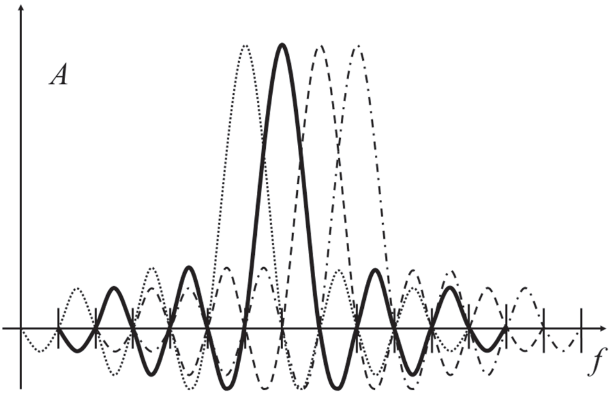


Figure 49 OFDM (Orthogonal Frequency Division Multiplex)

By using many carrier frequencies, each single subcarrier transmits only at a relatively low data rate. This reduces the probability of ISI (Inter-Symbol-Interference) due to the multipath. The interference from another carrier signal results only in interference of one subcarrier. The orthogonality allows for signal separation via inverse FFT at the receiver site. On the other hand, the precise synchronisation of transmitter and receiver is fundamentally important.

OFDM is used in a lot of modern digital transmission techniques like WLAN 802.11a, digital audio broadcasting DAB, digital video broadcasting DVB-T and 4G LTE mobile networks. This principle is also used in the digital subscriber lines ADSL in the residential access network, which is named DMT (Discrete Multi Tone) technique.

ISI of subsequent symbols can happen due to multipath propagation, in order to separate the payload data symbols correctly during the time frames T_{data}, a guard-interval (T_G = 31, 62, 123, 246 μs) precedes each payload data symbol (DAB: T_{data}= 1 ms; up to 1536 subcarriers).

For the transport mechanism, different channels are defined to transport control data and payload data: Main Service Channel, Fast Information Channel and Synchronisation Channel.

MSC (Main Service Channel) carries all user data (audio, multimedia) consisting of CIF (Common Interleaved Frames). Each CIF contains 55296 bit, all 24 ms, depending on transmission mode. CIT in turn contains the so-called CU (Capacity Units) with 64 bit.

FIC (Fast Information Channel) carries control information, consisting of FIB (Fast Information Block) with 256 bit (including 16 bit checksum), which defines the configuration and the content of the MSC payload content.

SC (Synchronisation Channel) is responsible for synchronisation.

Stream mode is used for transparent data transmission with a fixed bit rate, whereas packet mode transfers addressable packets.

DAB aims to achieve high-quality audio transmission almost with CD quality. It is robust against multipath propagation and the distortion of audio signals during signal fading.

Fully digital audio signal coding techniques (PCM, 16 Bit, 48 kHz, stereo) and MPEG compression of audio signals (compression ratio 1:10), will be used to achieve good spectral efficiency and at the same time a high quality of service, which will be enabled also by forward error correction. For this purpose, redundancy bits for error detection and correction are applied. Since the burst errors typically occur for radio transmissions, signal interleaving is used to correct single bit errors resulting from interference or transmission burst errors. By using low symbol-rate and long symbol sequences, separated by properly defined guard spaces, delayed symbols, e.g. by multipath propagation or reflection, still remain within the guard space without interfering with other symbols.

A DAB ensemble combines audio programs and data services with different requirements for transmission quality and bit rates. The DAB standard allows dynamic reconfiguration of the DAB multiplexing scheme during transmission. Data rates can be adaptively changed to transmit other services by using free capacities. Additional services can also come from different providers. In combination with the cellular mobile networks like GSM/GPRS, UMTS, LTE and future 5G mobile networks, backward feedback channels can also be used to enable communications between the broadcast customers by sending feedback information or requiring special information channels.

DVB (Digital Video Broadcast)
The pioneer standardisation works for the development of digital television in Europe were done by the ELG (European Launching Group), founded in 1991. This was renamed DVB (Digital Video Broadcasting), in order to introduce digital television technology based on different media like satellite, cable networks and terrestrial transmission. Correspondingly the technologies are named: DVB-T (Terrestrial); DVB-S (Satellite); DVB-C (CATV Cable Television, or HFC Hybrid Fiber and Cable Television).

The modulation and transmission technologies have been chosen to suit best each channel:

a) DVB-T is similar to DAB. The channel is characterised by multipath, so OFDM is the choice with many frequency carriers and a long symbol period.
b) DVB-S is restricted by relatively low received signal and highly efficient, nonlinear power amplifiers at the satellite. This is why a phase modulation with a constant signal amplitude, such as QPSK or 8PSK, was chosen.
c) DVB-C is impaired by the limited bandwidth of the cable channel. So a spectral efficient modulation is required. Channel dependent, 16-, 32-, 64-, 128- and 256-QAM are being used.

DVB uses video compression techniques like MPEG-2 or MPEG-4 to achieve high spectral efficiency and high flexibility for the transmission of different types of digital data information. DVB Service Information specifies the content of a container. The NIT (Network Information Table) lists the services of a provider and contains additional information for set-top boxes, whereas the SDT (Service Description Table) contains the list of names and parameters for each service within a MPEG multiplex channel. EIT (Event Information Table) contains the status information about the current transmission and additional information for set-top boxes. TDT (Time and Date Table) shows the update information for set-top boxes.

Due to characteristics of the broadcasting services, DVB, like DAB, is typically an asymmetric data exchange from the broadcasting station in a downlink to the DVB receivers. The data rates range per user from 6 to 38 Mbps. For the feedback or return channel in an uplink from the user or terminal station to the service provider or broadcasting stations, different transmission technologies can be involved: e.g. PSTN or ISDN with 64 kbps, xDSL with up to 50 Mbps or other access connection technologies like FTTH (Fiber to the Home).

CHAPTER 9

CELLULAR MOBILE NETWORKS

After the commercial beginnings of the first generation of analogue mobile wireless communication networks in the 1970s and 1980s, the really global, digital, cellular mobile communication was standardised as GSM (after the ETSI working group that was first called Groupe Spéciale Mobile but which has since been renamed the Global System of Mobile Communications) in the 1980s and implemented in the 1990s what is now commonly known as second generation mobile communication technology 2G.

By using digital transmission techniques like digital source coding, channel coding, FEC forward error correction, transmission quality in terms of efficient spectrum usage and a low error rate can be achieved. Also security was increased by using an air interface encryption for GSM.

At the same time, the really seamless roaming between network operators became possible, in order to support mobility. Location dependent services LBS could also be supported.

Along with research and development of microelectronics, i.e. integrated circuits, the performance of the mobile terminal devices in terms of battery operational time, digital signal processing, storage system, size and quality of the displays, graphic user interfaces and camera functionality has been continuously improved during the last 30 years. So much so that today the mobile terminal devices commonly known as smart phones have become integrated multi-media devices replacing many single, different devices.

The basic mobile network functionalities are: multimedia service, congestion and flow control, quality of service management or performance monitoring, addressing, routing, handover, authentication, media access control, multiplexing, modulation and encryption.

Mobile cellular networks provide seamless coverage in personal area networks (PAN), metropolitan area networks (MAN) and wide area networks (WAN), but also in the global area networks (GAN) by incorporating international roaming.

The cellular concept uses hexagonal cells to cover a certain, limited area up to 35 km, also called Space Division Multiple Access SDMA.

Each cell normally will be split into three sectors, in order to provide optimised radio coverage within each sector (Fig. 50). If the user moves from one sector or cell to another sector or cell, the connection will be handed over to the neighbouring sector or cell with better radio conditions.

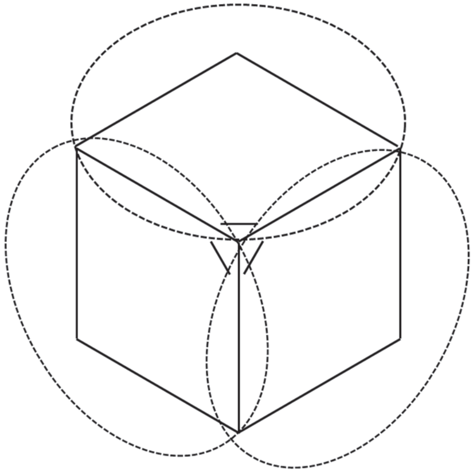


Figure 50 Three-sector cells of 2G/3G/4G (GSM/UMTS/LTE)

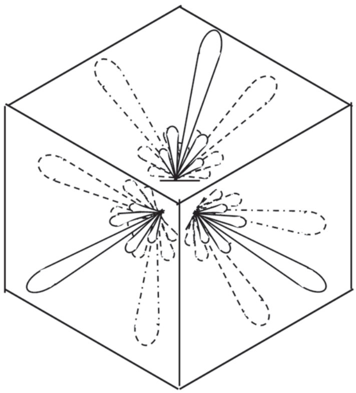


Figure 51 5G three-sector cells using MU-MIMO beam forming

In each cell or sector, several carrier frequencies can be used. In order to avoid interference with the neighbouring cells or sectors, the frequencies of the adjacent cells or sectors are to be planned carefully by the so-called reuse factor of the radio frequencies. Cell radii vary from some 100 m up to 35 km depending on user or traffic density, geography, transceiver power, etc. The technique with multiple carrier frequencies with certain spacing to each other (in GSM the spacing is 200 kHz) used in the cells and sectors is also called FDMA (Frequency Division Multiple Access). On each carrier frequency, the time domain is split into concatenated time slots, which will be reused by a certain mobile user in a planned manner. The multiple time slot reuse is called TDMA (Time Division Multiple Access).

Up to now, i.e. for 2G GSM, 3G UMTS and 4G LTE, inside one base station three-sector cell, the antenna covers approximately 120° (Fig. 50), independent from the location of the user terminals or some hot spots. This means that the spectrum efficiency is not optimised. In 5G a new improvement is proposed by using massive MIMO multi-user beam forming, so that the antenna beams really point to the important hot spots of single cells. By doing so, the unnecessary radiation of the transmit power into a region without any user terminals can be avoided, and spectrum and power efficiency of the transmission can be achieved (Fig. 51).

GSM uses a combination of SDMA, FDMA and TDMA. Later we will discuss how 3G UMTS which uses CDMA (Code Division Multiple Access) instead of FDMA and TDMA. Different generations of cellular mobile network technologies use different medium access technologies, in order to optimise the spectrum efficiency and to achieve the best possible available bit rate for a given spectrum width.

9.1 GSM (2G), UMTS (3G)

	GSM900	GSM1800	GSM1900
Uplink	890-915 MHz	1710-1785 MHz	1850-1910 MHz
Downlink	935-960 MHz	1805-1880 MHz	1930-1990 MHz
Channel spacing	200 kHz	200 kHz	200 kHz
No. of channels	124	374	299
Time multiplex	8 full-rate time slots, 16 half-rate time slots		
Time slot duration	577 μs		
Bits per time slot	114		
Frame length	4.615 ms		
Gross rate full-rate	22.8 kbps		
Gross rate half-rate	11.4 kbps		
Modulation scheme	GMSK		

Table 12 Typical GSM mobile network parameters

The 2G mobile communication system GSM uses SDMA (Space Division Multiple Access, i.e. cells and sectors), FDMA (Frequency Division Multiple Access, Carrier Frequency Spacing 200 kHz) and TDMA (Time Division Multiple Access, 8 Time Slots periodically) and provides basic voice telephony services and short message services.

In Figure 52 the GSM mobile network architecture is illustrated. The GSM mobile network consists of three basic subsystems:

1) **RSS/BSS:** The RSS (Radio Sub-System, Radio Access Network) or BSS (Base Station Sub-System) is responsible for radio propagation, coverage of the cells and sectors with the BTS (Base Transceiver Station) or BS (Base Station), MS (Mobile Stations, i.e. handhelds) and BSC (Base Station Controller) as the concentrator or hub-site for BTS to map the radio interfaces U_m to A for interconnection with the PSTN.

2) **NSS:** The NSS (Network Sub-System, Core Network) is responsible for call switching, mobility management, interconnection with other operators, MSC (Mobile Switching Centres), HLR (Home Location Register – data base containing all subscriber data) and VLR (Visitor Location Register – data base containing all user data of the visitors).

3) **OSS:** The OSS (Operation Sub-System) is responsible for operation and maintenance with all network management monitoring systems, AUC (Authentication Centre) and EIR (Equipment Identity Register).

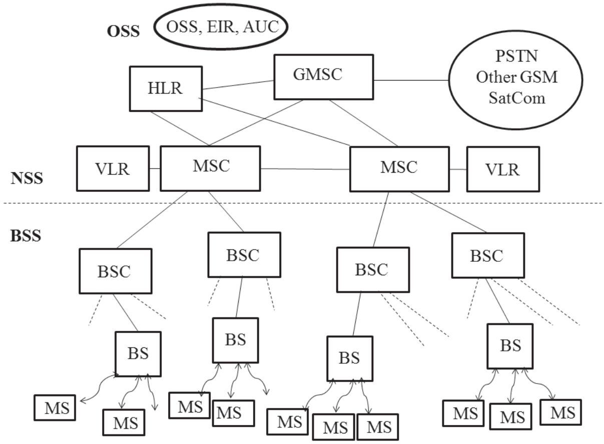


Figure 52 GSM mobile network architecture

The MSC (mobile switching centre) is the most important for switching functions, mobility support, management of network resources, interworking and interconnection functions via Gateway MSC (GMSC), integration of databases, paging and call forwarding, termination of SS7 or CCSS7 (common channel signaling system no. 7), mobility specific signaling, location registration and forwarding of location information and generation and forwarding of accounting and billing information.

The first improvement for 2G mobile communications was HSCSD (High Speed Circuit Switched Data), which proposed to use multiple GSM time slots in the CS (Circuit-Switched) manner in order to achieve higher bit rates, up to 57.6 kbps.

The other improvement for 2G mobile communications was GPRS (Generalized Packet Radio Service) to group up to eight time slots in PS (Packet-Switched) Manner, leading to a maximum bit rate of 171.2 kbps. Whereas HSCSD only needs a software update, GPRS requires hardware extensions SGSN (Serving GPRS Support Node) and GGSN (Gateway GPRS Support Node) in the GSM core network.

The further improvement and evolution of GSM is the so-called EDGE (Enhanced Data rates for GSM and TDMA/136 Evolution), which uses new modulation techniques for GSM-data services. GSM was initially specified to offer voice service with voice coding (260 bits) and channel coding (196 bits). GSM uses GMSK modulation, whereas EDGE uses 8-PSK (8 phases = 3 bits per phase), leading to three times more bits per burst compared to GMSK, about 70 kbps for user-data and channel coding.

EGPRS (Enhanced GPRS) uses new channel coding schemes in order to offer the highest possible bit rate at the current radio channel quality (C/I > 9 dB, BER ≤ 10^{-4}). Depending on the quality of the transmission channel, LA (link adaptation) can adapt within 100 ms with 9 different channel coding schemes:

MCS-1 (8.8 kbps, GMSK)
MCS-2 (11.2 kbps, GMSK)
MCS-3 (14.8 kbps, GMSK)
MCS-4 (17.6 kbps, 8-PSK)
MCS-5 (22.4 kbps, 8-PSK)
MCS-6 (29.6 kbps, 8-PSK)
MCS-7 (44.8 kbps, 8-PSK)
MCS-8 (54.4 kbps, 8-PSK)
MCS-9 (59.2 kbps, 8-PSK)

EGPRS2 (EGPRS2A/2B) goes further by using higher order modulation schemes 16QAM, 32QAM and Turbo Coding, in order to achieve even higher bit rates of 98.4 kbps (EGPRS2A) and 118.4 kbps (EGPRS2B).

The third generation 3G UMTS introduces a completely new medium access technology CDMA (Code Division Multiple Access), replacing the existing FDMA/TDMA of the 2G GSM. As discussed in the previous chapter, CDMA has a spectrum spread over the whole band. The single channels are separated by using an OVSF (orthogonal variable spreading factor) code in combination with scrambling codes. Therefore the complete UMTS terrestrial radio access network UTRAN has been introduced in parallel to the co-existing GSM Radio Access Network (Fig. 54).

UMTS base stations are named Node-B, whereas the radio network controllers, similar to BSCs (but much more powerful) are named RNC. The RNC is responsible for access and admission control, congestion control, radio resource control, channel coding, outer loop power control (FDD and TDD), handover, radio network configuration, radio channel encryption and system information broadcasting. Node-B is responsible for the channel quality measurements, inner loop power control, soft handover between different antennas and data transmission over the radio interface. The interface between the RNC and Node-B is called I_{ub}.

Slightly different in comparison with GSM, in UTRAN there is a new interface between the RNCs called I_{ur}.

Also there is a new concept for the cell mobility support: soft handover. In this case, instead of the "hard handover" in GSM, the connection of the UE to the previous Node-B is maintained while the new connection to the new Node-B is set up. This is a kind of redundant connection during the handover, providing a better stability and robustness. It is also called macro diversity. During the soft handover, the signals transmitted to the different Node-Bs contain identical data, but use different codes. UEs use a RAKE-receiver with multiple antenna elements which form the beams to certain base station antennas. Macro diversity does not only increase transmission robustness, it also reduces the multipath fading and shadowing effect for uplink and downlink.

The 3G UMTS mobile network is basically an extension of the Radio Access Network UTRAN (Fig. 53). The UMTS core network relies on the GSM voice and GPRS backbone, and consists of the GSM core including the connection interfaces to the UTRAN (Fig. 54). The Core Network (CN) has two logical domains:

1) **CSD** (Circuit Switched Domain) with the circuit switched services including signaling and resource reservation at connection set-up consists basically of GSM components (MSC, GMSC, VLR). The interface between CSD and UTRAN is the interface I_u-CS.

2) **PSD** (Packet Switched Domain) basically consists of GPRS components (SGSN, GGSN). The interface is I_u-PS.

One special characteristic of the UMTS CDMA technology is the effect of the breathing cells, i.e. the cell size is not only dependent on the transmitter power, but also depends on the neighbouring users in the same frequency band shared by all mobile users in the same cell in CDMA, representing the interference. Therefore the total cell capacity depends on the SINR (signal to interference and noise ratio). Since the transmitter power is limited, the decreased signal noise and interference ratio leads to drop out for users, especially at the edge of the cell. The cell seems to breathe, depending on the traffic or total cell capacity. Cell breathing is a great problem for network planning.

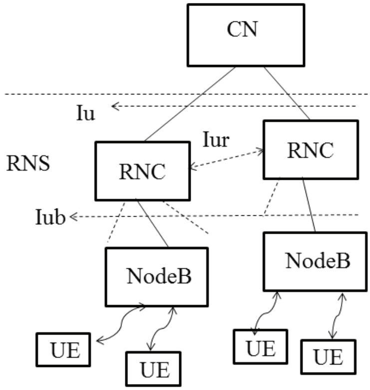


Figure 53 UMTS UTRAN, RNS

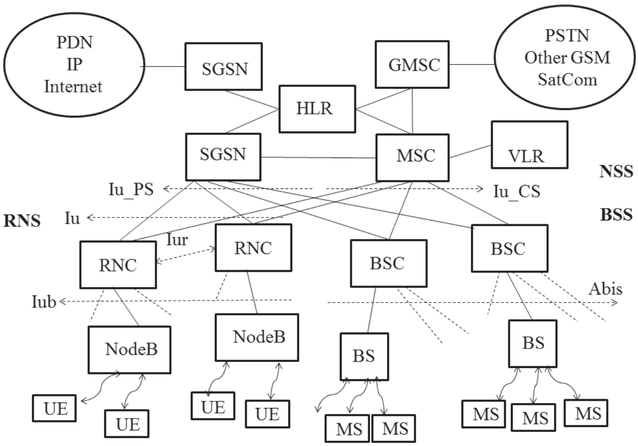


Figure 54 Integrated GSM/UMTS network

In UTRAN the radio network sub-system (like RSS or BSS in GSM) is called Radio Network Sub-system (RNS). It is responsible for mobility support, coverage and all radio specific tasks. The mobile terminal (similar to MS in 2G GSM) is called UE (user equipment). The core network (like NSS in GSM) is called CN (Core Network).

The next evolutionary step of UMTS is HSDPA (High Speed Downlink Packet Access), as specified in 3GPP Rel. 5, enabling data rates up to 14.4 Mbps with lower latency. HSDPA uses adaptive modulation and coding depending on the channel conditions, from 4-PSK to 16-QAM. The new feature MAC-HS (High Speed) in the MAC layer reduces the TTI (transmission time interval, or the encapsulation of data from a higher layer into frames for transmission on the radio link layer) to 2 ms.

With Release 7, Release 9 and Release 10, the use of Multiple Cell (Dual Cell, Triple Cell, Quad Cell, Hexa Cell, Octa Cell) and MIMO as well as a downlink modulation scheme up to 64QAM, HSDPA can achieve up to 168 Mbps. With 3GPP TS 25.306 Release 11, this can be even 337.5 Mbps. In the practical UMTS mobile networks 42 Mbps can be achieved.

HSUPA (High Speed Uplink Packet Access) allows 5.76 Mbps in category 6 and in category 9 of Release 9, even up to 23 Mbps.

9.2 LTE (4G)

4G LTE (Long Term Evolution) increases the spectral efficiency further in comparison with the HSPA rel. 6, reaching peak rates of 100 Mbps in downlink and 50 Mbps in uplink, and enabling round trip times of less than 10 ms. The purely packet switch mode enables the All over IP applications for voice (Voice over IP, VoIP) and data by using the IP Multimedia Sub-System (IMS) (Fig. 55). Also the mobility and security as well as terminal power efficiency have been further improved. Frequency bandwidth varies flexibly from 1.5 MHz up to 20 MHz.

Instead of FDMA/TDMA of 2G GSM and CDMA of 3G UMTS, LTE uses a new medium access technology OFDMA (Orthogonal Frequency-Division Multiple Access) for downlink, which is also used in WLAN/WiFi, WiMax (802.16), DAB, DVB-T, DVB-H and SC-FDMA (Single Carrier Frequency Division Multiple Access) for uplink. LTE shows relatively low peak-to-average transmit power ratio, which is favorable for the efficient terminal power amplifier.

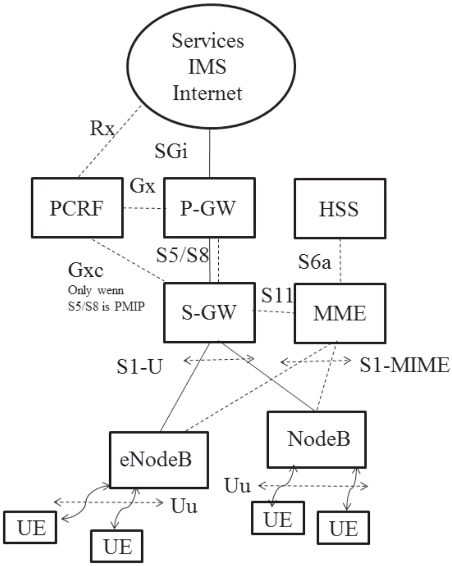


Figure 55 LTE network architecture

The LTE radio access network is called E-UTRAN. The base station is called eNodeB, whereas the core network is named EPC (Evolved Packet Core). UE (User Equipment), E-UTRAN and EPC together represent the internet protocol (IP) connectivity layer EPS (Evolved Packet System) [47] [48].

Mobility Management Entity (MME) is the main control element in the EPC. It is responsible for mobility support in the control plane CP. For example: authentication and security, mobility management, managing subscription profiles and service connectivity.

The Serving Gateway (S-GW) connects the E-UTRAN to the EPC, whereas the Packet Data Network Gateway (P-GW) is the edge router between the EPS and external packet data networks.

Policy and Charging Resource Function (PCRF) is basically responsible for the billing functionality.

Like HLR in GSM, Home Subscription Server (HSS) is the subscriber data base, which stores the subscriber profile and the location of the user in the visited network control node.

The services connectivity layer provides the users with IMS (IP Multimedia Services Subsystem) based operator services and Non-IMS based operator services as well as other services through the internet.

9.3 5G Mobile Networks

The 5G mobile network [49] [50] [51] [52] [53] [54] [55] [56] [57] [58] further improves the spectrum efficiency by using new technologies to significantly improve the QoS (Quality of Service) and QoE (Quality of Experience) in terms of high bit rate, low latency, redundancy, reliability, cost efficiency and energy efficiency.

Intelligent network management in the core network by using SDN (Software Defined Network) [49] and NFV (Network Functionality Virtualisation) [59] [60] enables not only implementation of network elements and sub-systems by software, but also better service differentiation depending on the QoS requirements (classical human-to-human communications, mission-critical services, time-sensitive networks TSN, URLLC (ultra reliable low latency communications), broadband services, massive machine type communications mMTC, etc. (Fig. 56). The so-called "network slicing" provides different "slices" within the existing physical transport network infrastructure with specific features, to exactly meet the QoS requirements concerning bit rates, latency and availability (Fig. 57).

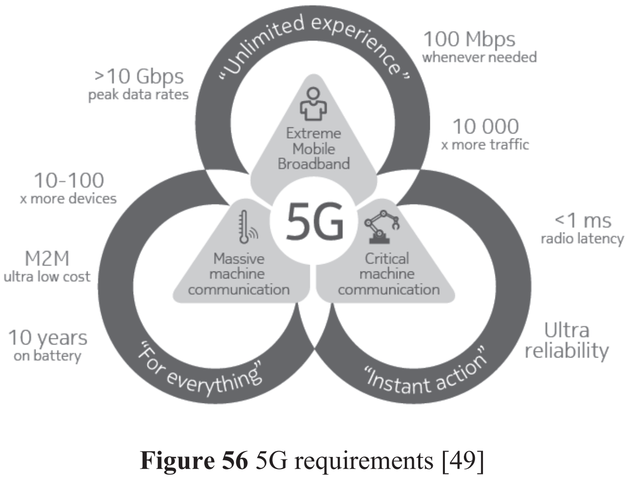


Figure 56 5G requirements [49]

By using different techniques including SDN (Software Defined Network) and NFV (Network Function Virtualisation) and based on the recommendations of standardisation bodies and previous research projects [61] [62] [63] [64] [65], partially already known in LTE-A, the total throughput will be significantly increased. Energy efficiency will be improved, and delay (1 ms) and latency can be strongly decreased, typically within one radio access technology. Besides the existing classical H2H (human-to-human) communications of the existing mobile communications networks 2G/3G/4G, as well as the rt-/nrt-NB/BB (real time and non-real time narrow band and broadband communications), more and more machine-to-machine communications will also be possible with 5G. The typical M2M massive Machine-Type Communications mMTC are sensor applications, which are typically nrt-NB (non-real time narrow band) services and used in industrial environments and smart homes (Fig. 58).

Mission-critical communications (ultra-reliable low latency communications URLLC or ultra-reliable machine type communications uMTC, industrial robot control, gaming, Vehicle-to-Vehicle/Car-to-Car communications V2V/C2C, Vehicle-to-Infrastructure communications V2I, autonomous driving, tele-surgery and health care) are real-time narrow-band and/or broadband communications rt-NB/BB. The basic service classes are summarised as follows [66]:

eMBB (enhanced Mobile Broad Band): focuses on services characterised by high data rates, such as the new enterprise services and applications along with the exploding consumption of multimedia and collaborative working and social communications, such as AR/VR and video in all its various forms and formats.

mMTC (massive Machine Type Communications): focuses on services that have high requirements for connection density, such as those typical for smart city and smart agriculture use cases. Smart homes, smart cities and smart factories, all containing billions of sensors, require access to a flexible and scalable infrastructure.

URLLC (Ultra-Reliable and Low Latency Communications) or uMTC: focuses on latency-sensitive services, such as self-driving, remote surgery or drone control. A large number of real-time applications will demand end-to-end network latency of single digit milliseconds to avoid perceivable lags in browsing or videos or to control drones and robots. In addition, critical machine communications require very high reliability and very low latency; for example, in public safety, autonomous vehicles on the highway and in telemedicine.

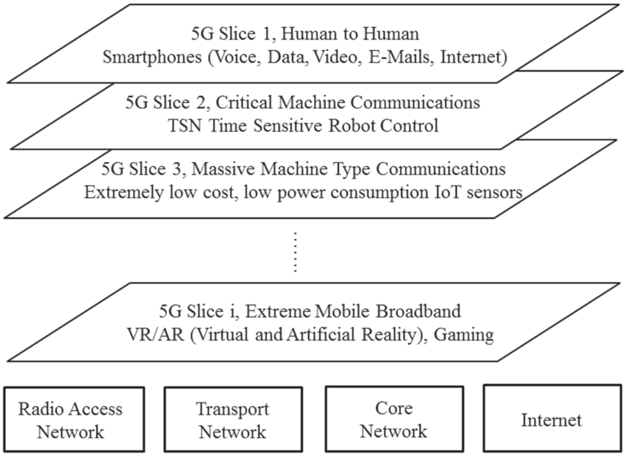


Figure 57 5G network slices enabled by SDN/NFV

One key success factor is the so-called 5G Multiple Radio Access Technology (M-RAT) (Fig. 59), which benefits from the coexistence and usage of various radio access technologies (4G LTE, including 2G GSM voice service backup, WLAN/WiFi and 5G NR (New Radio) from 3 GHz to cm-/mm-Waves). Network resources will be shared and selected by the mobile users and networks in an optimised manner.

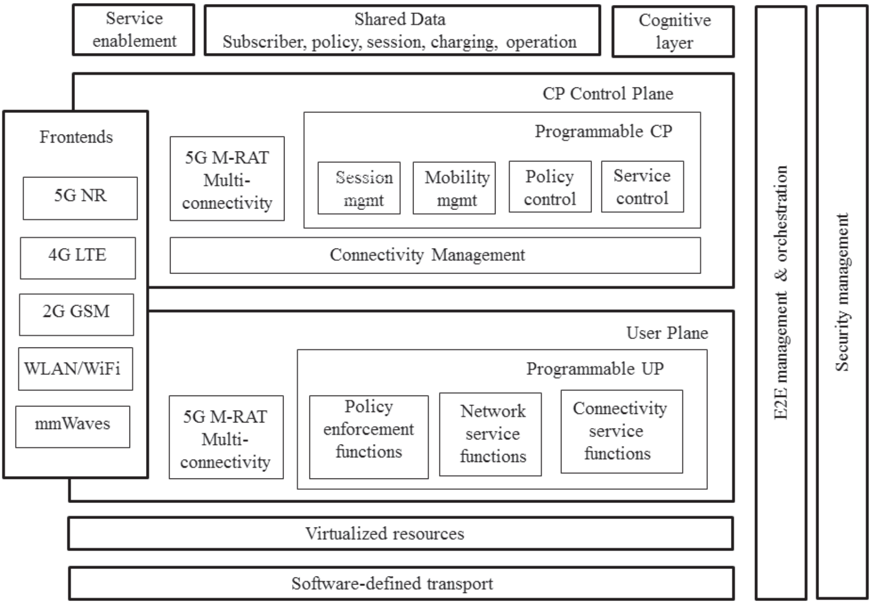


Figure 58 5G architecture [50]

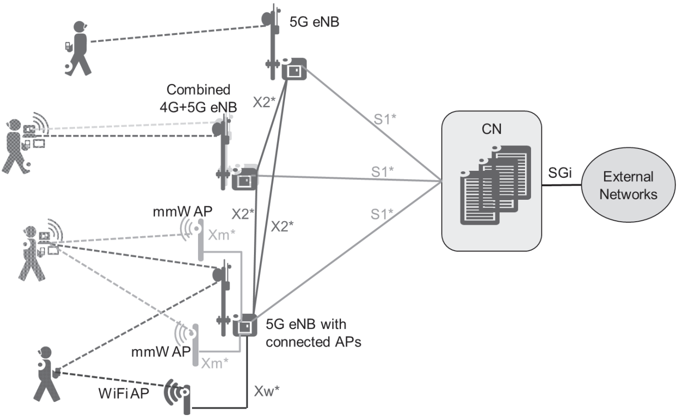


Figure 59 5G Multi-Radio Access Technologies (RAT)

Some Typical 5G Use Cases:

- eMBB for enterprises: Real-time (delay sensitive) rt and non-real-time nrt broadband services.
- Tele surgery (URLLC): Ultra reliable, low latency communications, but in most cases requires also a large bandwidth, in order to transmit high resolution live images or live streaming during the surgery with high reliability or low bit error rates as well as low delay and low jitter.
- Smart meters (mMTC, IoT Internet of Things): Narrow band, massive machine type communications, i.e. massive number of sensors or internet of things IoT devices, which normally do not necessarily require high bit rates, but do require high energy efficiency or low power consumption in order to operate without permanently changing the batteries of the devices and to enable a long operation lifetime.
- V2x communications: In autonomous driving, low latency will be fundamentally important in order to keep the reaction time as short as possible.

New 5G Core Functionalities:

- Control and User Plane Separation (CUPS).
- Network slicing (Virtual Network Function VNF corresponding to the QoS and performance requirements. SDN plays the crucial role in selecting the best path traversing through front haul and back haul in cooperation with the MRRM Multiple Radio Resource Management).
- Cloud native and micro-services (NFV-based deployment of services. The applications could be split into small, independent services, not a single language or single large code. Self-scaling, self-healing).
- Service-Based Architecture (SBA).
- Edge clouds, where possible and necessary, especially for time-sensitive, low latency-applications, in order to reduce the round trip time RTT to a minimum.

5G Security [67] [68] [69] [70] [71] [72] [71] [72] [73] [74] [48] [70] [75] [76]

- User devices UE, possibly the most attractive point for attackers (SMS/MMS-based DoS (denial of service) attacks and other mobile malware like worms, trojans, viruses, mobile botnets, etc.).

- Each interface of the RAN (Radio Access Networks), UE location tracking based on C-RNTI or on packet sequence number, IKEv2 attacks between HeNB (Home eNodeB, femtocell or small cell) and security gateway, user data and privacy, radio resource management.
- Each interface of the transport network.
- Each interface of the mobile core networks (IKEv2 attacks between HeNB and security gateway, amplification of signalling overhead, EPC HSS saturation or HLR overload).
- Each interface of core networks to external internets (enterprise networks).
- Interfaces between 5G to legacy 3G/4G.
- Unauthorised access or usage of assets.
- Weak slices isolation and connectivity.
- Traffic embezzlement due to recursive/additive virtualisation.
- Insufficient technology level readiness.
- Difficulties managing vertical SLA and regulation compliance.
- Slicing versus neutrality.
- Trust management complexity.
- Provisions to facilitate change of service provider domain lock-in.

5G Security Requirements

- 5G should be designed to provide more options beyond node-to-node and end-to-end security available in today's mobile systems.
- Design of security solutions (e.g. key exchange/derivation protocols upon handover or when interworking with other RATs) should provide better secrecy than 4G.
- Specific security design for use cases which require extremely low latency (including the latency of initiating communications).
- Improve resilience and availability of the network against signalling based threats, including overload.
- Improve system robustness against smart jamming attacks of the radio signals and channels.
- Improve security of 5G small cell nodes.

5G Multi-Radio Access Technologies

The Multiple Radio Access Technology M-RAT resource selection behaviour (Fig. 60) within one LTE macro cell with an arbitrary number of micro- and pico-cells (WLAN, mmWave) inside one macro-cell was investigated in [77] [78] [79] [80] [81].

	Carrier Frequency	Bandwidth	BS TX Power	Typical Bitrate	Cell Radius	Delay
LTE	0.9-3.5 GHz	1.25-20 MHz	46 dBm	2-50 Mbps	0.5-1.7 km	100 ms
WLAN 802.11a-n	2.4/5 GHz	20-80 MHz	23-30 dBm	100-400 Mbps	10-50 m	n. g.
WLAN 802.11ac	5 GHz	2-80 MHz	23-30 dBm	3.6 Gbps	10-50 m	n. g.
5G low band	0.8-6.0 GHz	GHz	15 dBm	1 Tbps	0.5 km	3 ms
5G high band	6.0-90 GHz	GHz	15 dBm	1 Tbps	0.5 km	3 ms

Table 13 5G Multi-RAT simulation parameter

	rt-NB	rt-NB (uMTC)	nrt-NB (mMTC)	rt-BB	nrt-BB
	Voice, Skype	V2V, Control	IoT, Sensors	Video Conversation	Streaming, Download
LTE	×		×		
WLAN/WiFi	×		×		×
5G low band	×	×	×	×	×
5G high band		×	×	×	×

Table 14 5G default resource allocation and MRAT-mapping

Interferences at the boundaries to the neighbouring LTE macro cells are considered in [81]. 2D antenna characteristics are applied. If necessary, the vertical beam width can be considered.

For mmWave (for example 28 GHz), two options were investigated: directional antenna and three-sector antenna like 3GPP LTE/UMTS [82] .

In the case of three-sector cells (LTE or mmWave), reuse factor (1 or 3) is investigated.

Three sector mmWave antennas are modeled by using $\cos^2$-characteristics (beam width 61° - 68°), whereas a directional mmWave antenna is modeled by using linear array (backfire attenuation 15 – 50 dB).

Traffic type: full buffer, i.e. mobile user or devices have always packets to transmit.

Traffic scheduling: round robin, i.e. available resources will be shared equally by all mobile users.

Multi-band mobile devices (2 GHz – 28 GHz) and SDN/NVF as well as MRRM are indispensable.

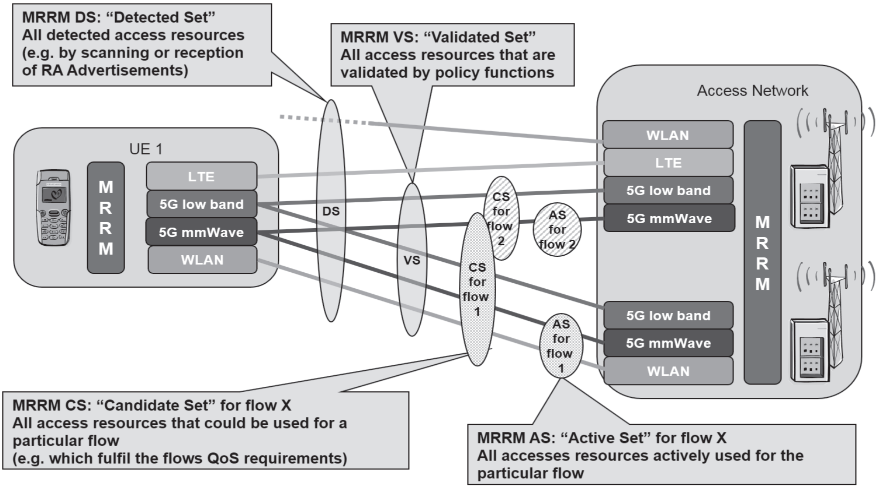


Figure 60 M-RAT by using MRRM access sets

9.4 Mobile Network Planning Aspects

Although generations of mobile networks use different media access technologies, different coding schemes and different operating frequency bands, there are some common, general planning aspects, which should be taken into account:

- Target mobile network coverage area and population coverage ratio.
- Densities of the population in certain areas.
- Provided services (voice, video, data and internet communications).
- Provided devices (mobile phones, computers, laptops, machine type communication MTC sensor applications and IoT devices).
- Required and demanded quality of service, i.e. blocking probability and call set-up time depending on the population density.
- Required and demanded reliability (access to services, outage time or availability).
- Required and demanded bit rates depending on the available spectrum.

Based on these primary service requirement parameters, license conditions for minimum area coverage and/or population coverage, the marketing strategy of the network operators and the capacity limitations of the mobile network systems, the mobile network planning procedures are described schematically in Figure 61.

Traffic Forecast, Traffic Planning

Based on the marketing strategy, the coverage planning will consider the different coverage areas, like urban and rural regions, highways, railways, shopping malls and business areas with many companies generating high traffic. Based on the prognosis for user distribution and traffic assumptions (such as traffic intensity per user for the entire service area), the so-called initial cell planning uses different sizes of the cells like macro-cells, micro-cells and pico-cells, to correspondingly enable the required traffic intensities.

The classical voice traffic intensity can be estimated in Erlang, which is the number of calls per hour (3600 seconds) with the average call holding time in seconds. Given the user density in an area, the traffic density can be expressed Erlang/km^2. If, on average, a user makes one call per hour for 120 seconds, then the traffic intensity will be 120 / 3600 = 33 mErlang.

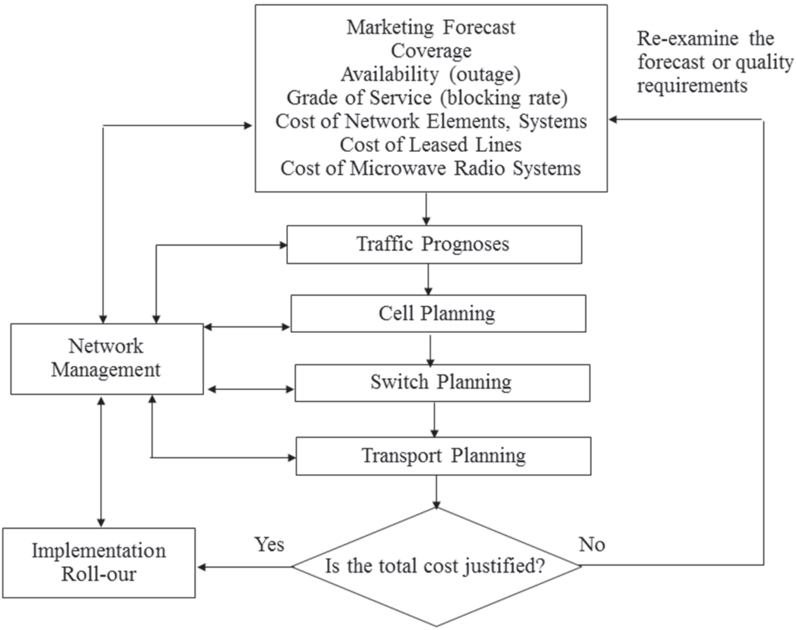


Figure 61 General mobile network planning procedures

The new generations of mobile communication technologies provide a large number of multi-media services. The traffic intensity then can be estimated by using the ETE (equivalent telephony Erlang), which is defined with relation to the transmission bit rates of the basic telephone calls. Data services traffic can be measured in Mbps/km^2.

Based on the traffic forecasts, nominal cell planning is performed with some radio network planning tools including realistic 3D geographic terrain data (forests, mountains, rivers and buildings of the region, etc.) in order to properly consider the line of sight (LOS) propagation of the base station antennas, reflection and multipath propagation, and the attenuation of walls, floors and ceilings in a building. Classical cellular network base station cells consist of three sectors. Each sector is served by one antenna system with a radiation beam width of approximately 120°. Normally iterative processes are needed to achieve an optimised cell structure in one region. At the end a theoretically optimised initial cell plan, also called a nominal cell plan, is obtained, also showing the potential base station sites, which enables the optimised radio coverage with the lowest interference between neighbouring cells. The interference can be minimised by choosing correspondingly optimised transmitter power levels of the base station sectors.

By using the nominal cell plan, radio site survey engineers will go to the interesting base station sites, negotiate with the landlords and building owners, and perform the site survey measurements to confirm the nominal cell planning results. If the sites cannot be used for the base station installation, or the radio propagation conditions are different from the predictions of the radio network planning tools, then the nominal cell plan (Fig. 62) must be modified. These processes can be repeated several times, until the radio network planning team can achieve a consolidated nominal cell plan.

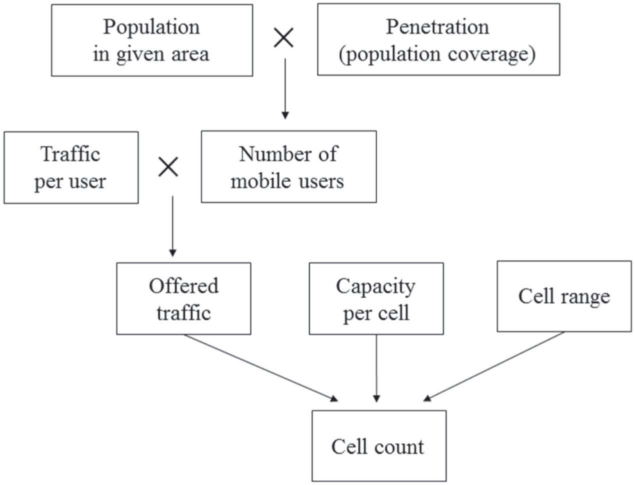


Figure 62 Cell count

At the same time, the transport network connections – i.e. the radio relay systems, optical fibres or leased lines – will be planned in parallel, in order to provide sufficient transport network capacities for the backhaul links.

Thereafter, the implementation process or construction work can start, sites and base station cabinets will be prepared, and antenna masts, antennas and base stations can be installed. In most cases, further adjustments or fine tuning steps will be necessary in order to achieve the optimised radio network coverage. Especially the drive tests during the trial phase after the installation will determine some coverage problems which must be solved before the base station really goes live in operation. Even after the start of the commercial operation of the base stations, the radio network must be continuously optimised, in order to get the best coverage everywhere in the cell or sectors, and to reduce the interference with the neighbouring cells to a minimum. Also the frequency reuse pattern, power control parameters of the transceivers and handover parameters are very important in this context. The following example will illustrate how many cells are needed to cover a certain area.

Example 1:
Assume a coverage area of 100 km^2 and a distribution density of the voice users of 5000/km^2, which means extremely high user density in a small area. This high traffic density area is often called a "hot spot". It is normally in an urban area, especially in city centres. According to the link budget calculation a maximum cell area of 23 km^2 will be the limit, in order to enable proper receive power levels. One base station cell can support 1040 users. Then the cell size and the number of the cells can be estimated: cell area = 1040 / 5000 = 0.2028 km^2; 100 / 0.2028 = 481 cells are needed.

Example 2:
On the other hand, if the user density in a rural region was 30 users/km^2 – an extremely low user density – then one cell could cover 37.4 km^2. Correspondingly the required number of cells would be only 100/23.32 = 5.

In the above example, a traffic load of 50% is considered. In network planning, the transmission capacity or circuits for voice traffic must be designed in such way, so that the relation between the available circuits and the traffic for a blocking probability p_B, also known as Grade of Service GoS, is determined by the Erlang B formula

$$p_B = \frac{E^M / M!}{\sum_{i=0}^{M} E^i / i!}. \tag{9.1}$$

The total traffic intensity can be estimated by $E = m \cdot H \cdot U$ (Erlang), with m the average call requests, H the average call duration per user, U the users and M the available channels.

Since the blocking probability p_B given in the Erlang B formula (9.1) can be only calculated numerically in a network planning tool or by using a computer, in daily practice the blocking probability is given in an Erlang B table. The blocking probability of 2% should never be exceeded in the network planning and dimensioning, and is considered as a limit, even for the busy hour traffic load. For this given limit, the corresponding number of the traffic channels TCHs must be estimated for a total traffic load. This estimation can be done easily by using the Erlang B tables for a given traffic and blocking probability. In mobile communication networks, the traffic load estimation is very dynamic in comparison with fixed networks, partially because of the handover possibility, that is, when the connection between the user with one base station changes to another base station. This can happen if the received power level from the first base station becomes, say, 3 dB lower (so-called handover margin) than the received power level of the second base station. This fact even accelerates the dynamic change of the traffic.

2G GSM uses FDMA/TDMA medium access scheme, so that for each sector in the cell the carrier frequency, with 200 kHz bandwidth, is used exclusively and the radio propagation characteristics are therefore deterministic and interference is limited, if the frequency reuse pattern between the neighbouring sectors and cells is planned properly.

3G UMTS instead uses the so-called CDMA medium access scheme, so that all users use the same shared, but broader frequency band with a bandwidth of 5 MHz. The users are distinguished by the orthogonal codes, a kind of keys. Here the traffic load of single users directly influences the signal noise interference ratio SINR, so that the interference increases with increasing number of users, and the cell coverage, especially at the cell edge, becomes worse. The cell size seems to shrink. The effect is called cell breathing in UMTS CDMA. For traffic loads over 75% the cell becomes unstable. Cell breathing also causes more handover failures and inter-cell interference.

Radio propagation can be calculated by using the simplified Okumura-Hata propagation model for an urban macro cell with a base station antenna height of 30 m, a mobile antenna height of 1.5 m and a carrier frequency of 1950 MHz:

$$L = 137.4 + 35.2 \cdot \log_{10} R \tag{9.2}$$

and for a suburban area

$$L = 129.4 + 35.2 \cdot \log_{10} R\,. \tag{9.3}$$

Once the cell range or cell radius R is determined, the site area can be approximated for hexagonal shape of the coverage area and will be calculated

$$A = 2.6 \cdot R^2\,. \tag{9.4}$$

Both the cell coverage range (limited by the radio propagation) and the cell capacity limit will be considered in the radio network planning, in order to determine the minimum cell count as shown in Figure 62.

For rural regions larger cells up to 35 km can be planned, since the traffic load is relatively low, whereas in a hot spot, like a city centre with large number of users, a high traffic load which pushes the cell capacity to the limit is generated, so that the cell must be planned correspondingly smaller.

Traffic increase can be handled either by reducing the cell size, leading to more necessary base station sites and higher total network investment, or by increasing the number of the sectors with corresponding narrow beam antennas. In the latter case, interference may increase due to the possible overlapping of sector antenna beams.

As mentioned before, multipath propagation can also happen, if the radio signal is reflected by buildings and other obstacles, leading to superimposed signals with different delays at the receiver. Due to multipath propagation, the receiver does not only receive the wanted LOS transmission from the transmitter, but also the multipath signals reflected by the obstacles. The total signal varies depending on the positions of the obstacles, and on the relative velocities of the transmitter and receivers to each other. Generally, transmitters and receivers could be mobile network base stations or mobile terminals. This effect is called the fading effect.

Multipath effects can be mitigated by using the so-called Rake receiver. This Rake receiver estimates the correlation of received signals with different phase delays. Then the phase of these fading signals can be shifted correspondingly, in order to regenerate the original transmitted signal by constructively superimposing these signals.

Besides a normal base station antenna, as, for example, used in 2G GSM, a MIMO (Multiple Input Multiple Output) antenna array can also be used to form the antenna beam flexibly, provide space redundancy or achieve the higher capacity.

Space redundancy improves the resistivity against fading. Beam forming can reduce the interference. It can also increase the level of the antenna beam focusing exactly to the receiver, therefore increasing the signal to interference and noise ratio, the channel capacity, the data rates and, finally, the spectral efficiency.

On the other hand, both fading compensation and beam forming or focusing require sophisticated DSP (Digital Signal Processing), channel measurement and channel estimations in order to compensate the multipath propagation effects.

In the following figures, SIMO (Single Input Multiple Output, Fig. 63), MIMO (Multiple Input Multiple Output, Fig. 64) are illustrated.

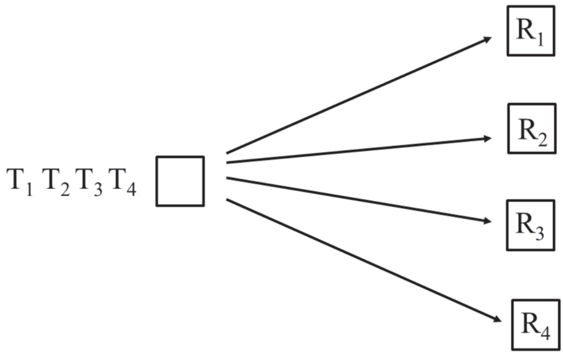


Figure 63 Transmit diversity, maximum gain (SIMO)

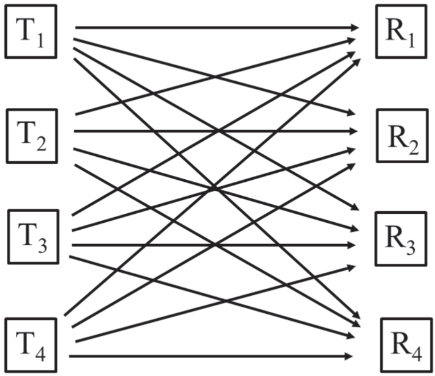


Figure 64 Parallel transmission, maximum capacity (MIMO)

In order to compare all the SISO with SIMO and MIMO, the channel capacities are calculated. For one channel, with no diversity:

$$C = \log_2\left(1 + SNR\right) \quad \text{in [b/s/Hz].} \tag{9.5}$$

MIMO with N transmit (TX) and M receiver (RX) antennas for unknown channel:

$$C = N \cdot \log_2\left(1 + SNR \cdot \frac{M}{N}\right) \quad \text{in [b/s/Hz].} \tag{9.6}$$

If $M = N$, then the total capacity C in (9.5) will be increased by a factor of N.

RX & TX diversity, i.e. the payload is transmitted in parallel via N TX to M RX antennas for a known channel, especially for SU (Single User) or MU (Multiple User) beam forming:

$$C = \log_2\left(1 + SNR \cdot \frac{M}{N}\right) \quad \text{in [b/s/Hz].} \qquad (9.7)$$

The backhaul and backbone transport networks are designed in such a way, that all mobile network traffic will be reliably transmitted between the radio access network (GSM BSS, UMTS UTRAN) and the core network, and between all the core network systems, sub-systems and elements.

The transmission media in an integrated BSS and UTRAN transmission network will be a mixture of PtP (Point-to-Point) radio relay systems (7 GHz, 23 GHz, 38 GHz, etc.), leased lines from the incumbent network operators or sometimes also optical fibres, either installed by the mobile network operators themselves or leased from large utility companies, gas pipeline operators or local power supply companies. Some communities build their own optical fibre communication networks, which can be also provided to the nationwide network operators, in order to complete their access transport network in certain regions.

The connection must be protected either by ring or partially meshed configurations, typically as an optical fibre communication backbone or redundant path protection (SNCP Sub-Netwok Connection Protection, MSP Multiplexer Section Protection), in order to avoid traffic loss in the case of outage of one section of the physical transmission media. Traffic overload can be flexibly rerouted in such ring configurations.

Reservation of the capacity from x% up to 100% can enable the so-called 1:1 protection or even 1+1 protection, by so-called trunk or circuit recovery. In the case of 1+1 protection, the primary path and redundant path will be configured and monitored in a mode of “hot standby”, so that the redundant path can take over the primary path traffic immediately, if the primary path suffers an outage or severe quality degradation. If properly designed, in the case of an outage of a single transmission link, the traffic will be recovered simultaneously to x% - 100%, i.e. no traffic overload or outage will be perceived by the mobile users. In the case of 1:1 protection with x% capacity reservation, only the highly prioritized traffic will be recovered by the x% reserved capacity.

CHAPTER 10

WIRELINES AND WAVEGUIDES

10.1 Copper Wire

The oldest wireline transmission technique is the twisted pair copper wire (Fig. 65), which is still used in the access network between the CO (central office) and the business or private subscribers or terminals via distribution box. It is also used in the digital subscriber line DSL. Different DSL techniques, also called xDSL, are used to provide the so-called last mile connections: ADSL (Asymmetrical Digital Subscriber Lines), HDSL (High bit rate Digital Subscriber Lines), SDSL (Symmetrical Digital Subscriber Lines) and VDSL (Very high bit rate Digital Subscriber Lines). The most popular and most important xDSL techniques nowadays are ADSL and VDSL.

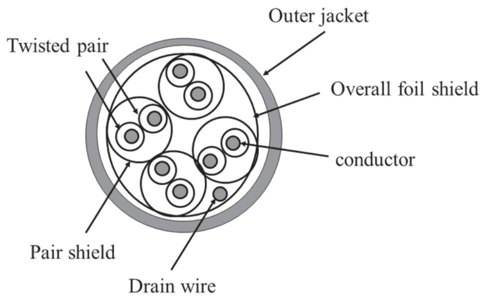


Figure 65 Twisted pair copper wires

ADSL supports downlink from a central office ATU-C (ADSL Termination Unit Central Office) to the customer's premises ATU-R (ADSL Termination Unit Remote) with a bit rate of about 16-50 Mbps and uplink (from the customer's premises to the central office) with about 1-10 Mbps. ATU-R is nothing but the ADSL router. Since the twisted pair copper wires suffer a strong specific attenuation (about 14 dB/km for a wire diameter of 0.35 mm, and 5.7 dB/km for a wire diameter of 0.8 mm) and interference during the transmission, especially for the unshielded twisted pair copper wires, the maximum distance between ATU-C and ATU-R is approximately 3 km for the above-mentioned bit rate of 16 Mbps. For data transmission the so-called DMT (Discrete Multi-tone Technique) in combination with adaptive modulation is used. The spectral band from 138 kHz – 2208 kHz is split into 256 sub-channels (multi tones) with a channel spacing of 4.315 kHz. Each channel will carry certain data modulated by mQAM (m-valued Quartenary Amplitude Modulation, for example 16QAM, 64QAM). Depending on the distance and received signal strength, noises and interferences in different sub-channels, the value m can be determined and adaptively used individually. This means, if certain sub-channels have very good *SINR* (signal to interference and noise ratio), the higher number of bits or higher order of modulation schemes can be used. If the attenuation and interferences are relatively high, i.e. the *SINR* value is relatively low, then only few bits, or in the worst case no bits, can be used to modulate the data in this specific sub-channel. This DMT technique is very similar to the OFDM principle used in the wireless systems (see Chapters 6 and 7) and enables flexible allocation of the payload data onto different sub-channels. It can also provide relatively high bit rates robustly, although the twisted pair infrastructure is sensitive to interferences.

10.2 Coaxial Cable

Coaxial cables (Fig. 66) have been used in the CATV (Community Area Television System/Cable Television) for decades. Classically the CATV network is a strongly asymmetrical TV program distribution network with downstream channels (80 MHz – 1 GHz). The upstream channels (5 – 65 MHz) are used to transmit the network management information from the subscriber to the cable network operator. Along with the digitalisation of the TV programs with DVB (Digital Video Broadcasting), the corresponding standards are DVB-C (C stands for cable) and DVB-C2. The other standards for terrestrial and satellite video broadcasting services are DVB-T and DVB-S. More and more the upstream channels are being used not only for pay-per-view services, but also for bidirectional internet services, making them a competitor for the xDSL and FTTH.

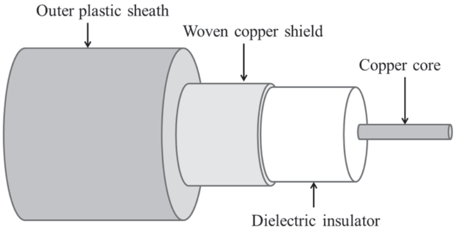


Figure 66 Coaxial cable

Many CATV network and service providers also use their HFC (Hybrid Fibre and Coaxial – a combination of optical fibre and coaxial cable) network to provide customers with high bit rates and bidirectional internet services as well as the classical TV programmes. Since the coaxial cables are shielded, interference can be eliminated and the SINR is higher and more stable in comparison with twisted pair copper wires.

10.3 Optical Fibre

Since the invention of quartz glass fibre for optical communication by Charles Kao [83], the optical fibre has been developed rapidly and is today the most important, most reliable and most indispensable backbone of the modern internet and global multimedia data transmission (Fig.67).

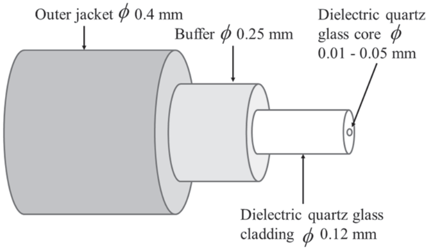


Figure 67 Dielectric quartz glass optical fibre

Optical fibre has no interference or EMC problems and almost unlimited bandwidth, so that each single optical channel can achieve up to 1 Tbps bit rate. Especially by considering DWDM (Dense Wavelength Division Multiplex) technologies, where 40 – 100 optical channels can be multiplexed, this is a tremendous capacity that can be transmitted in comparison with other wireline techniques and classical microwave propagation. Optical fibre will likely remain the unrivalled transport technology, not only used in the LAN (Local Area Network) and MAN (Metropolitan Area Network), but also in the nationwide wide area network WAN and the global area network GAN consisting of transoceanic optical fibre networks connecting all the continents with each other.

The technical details and applications will be discussed in Chapter 11.

10.4 FTTC, FTTH and FTTBS

If one likes to provide higher bit rates than 16 Mbps, then the *SINR* has to be increased. This can be done by decreasing the attenuation of twisted pair copper wires and/or by reducing the length of the copper wires. One promising solution is to use optical fibres, which show extremely low attenuation of 0.24 dB/km (see also Chapter 11) and also better EMC (electromagnetic compatibility) against the interferences produced by radio signals from various types of transmitter antennas, electric machines, switches and other electric pulses that affect the twisted pair copper wires in the corresponding spectral frequency band. In this case the optical fibre will be extended from the backbone network very closely to the customer's premises, i.e. from 3 km to less than 300 m, so that not only high SINR values can be achieved in the band from 138 kHz – 2208 kHz, but one can even extend the band to 12 MHz for VDSL or 30 MHz for VDSL2 to achieve 25 Mbps DL and 5 Mbps UL (VDSL) or even 50 Mbps DL and 10 Mbps UL (VDSL2).

In the case of a distance of 100 – 300 m from the optical fibre termination unit to the customer's premises, we talk about FTTC (Fibre to the curb). If however the fibre is directly installed into the home or office floor, the solution is called FTTH (Fibre to the home), which will be technically definitely the perfect solution. However, replacing the existing, almost-100-year-old twisted pair copper wires completely with fibres requires tremendous construction and installation work as well as extremely high expenditures.

At the time being, the copper wire infrastructure therefore remains the indispensable last mile technique to support xDSL. Full FTTH (Fibre to the Home) coverage will definitely take decades. xDSL will remain the last mile technique for quite a long time in rural regions, even though the FTTH and FTTC will be rolled out gradually. This is because it is economically viable only in urban areas with high population densities in the first roll out phase, promising cost-efficiency for the relatively expensive construction and installation work for the optical fibre cable ducts.

Along with the rapid increase of the traffic demands due to the new 5G and later 6G technologies as well as new real-time and non-real-time-broadband services, the backhaul connections between the mobile base stations and the backbone networks can be only enabled by the optical fibre connection, so that the FTTBS (fibre to the base stations) will be indispensable.

CHAPTER 11

OPTICAL FIBRE COMMUNICATIONS

11.1 Optical Fibre

A quartz glass fibre consists of a concentric cylindrical core and cladding regions (Fig. 67 and Fig. 68). Based on the refractive index difference between core and cladding of about 1 %, the optical waves in the wavelength band of 1.3 – 1.6 µm can be predominantly guided in the core region. The fibre cladding has a diameter of about 120 µm, whereas the core diameter ranges from about 9 µm – 50 µm. Depending on the size of the core, the fibres are categorised into single mode fibres with a core diameter of 9 µm and multi-mode fibres with a core diameter of 50 µm (Fig. 68). Analogue and digital data can be modulated and transmitted [8] [10] on these optical carriers efficiently.

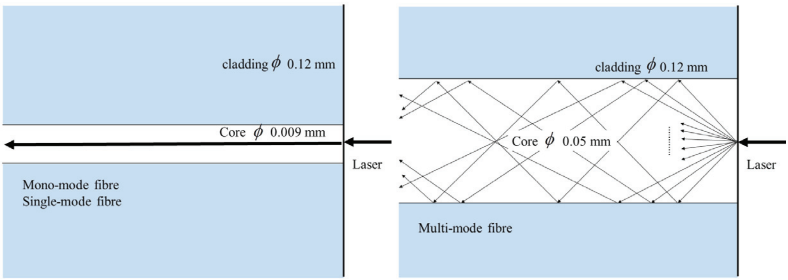


Figure 68 Single-mode and multi-mode fibre

In Figure 68 the wave guidance is illustrated by geometrical optics; that is an approximation in the form of total reflexion at the boundaries between the core and cladding regions. It should be pointed out that this geometrical optics approximation only gives a simplified explanation of the wave guidance. Only Maxwell's equations can give the exact solutions for only a certain number of waves which could be guided and are called modes. The number of the guided modes basically depends on the diameter of the core region $d = 2a$ (9 – 50 μm) and the difference between the refractive index in the core region n_c and the refractive index in the cladding region n_{cl}, or the so-called numerical aperture NA between core and cladding by using the normalised frequency V and normalised propagation constant B, where β is the propagation constant and λ the wavelength.

$$V = k \cdot a \cdot \sqrt{n_c^2 - n_{cl}^2} = \frac{2\pi}{\lambda} \cdot a \cdot \sqrt{n_c^2 - n_{cl}^2} = k \cdot a \cdot NA \qquad (11.1)$$

$$B = \frac{(\beta / k)^2 - n_{cl}^2}{n_c^2 - n_{cl}^2} \qquad \text{with} \qquad n_{cl} \le \beta / k \le n_c \qquad (11.2)$$

with λ as wavelength, $k = 2 \cdot \pi / \lambda$ the so-called wave number, a as the radius of the core region and n_{cl} and n_c as the refractive indices of the cladding and core regions of the quartz glass optical fibre. n_c is about 1.45 and the relative refractive index difference $(n_c - n_{cl}) / n_{cl}$ is less than 1%. All modes with the propagation constants $b / k < n_{cl}$ will not be guided by the fibre, but radiated through the cladding. These are called radiation modes and cannot be used for optical signal transmission.

The difference of the refractive indices in the core and cladding regions can be obtained when pure silica quartz glass is doped with P_2O_5, GeO_2, Na_2O and B_2O_3.

The higher the normalised frequency V, the more modes will be guided. For each mode or guided wave, the higher the value V, the stronger the guidance will be, i.e. the value β will be higher. The maximum propagation constant β of the guided modes is lower than $n_c k$.

If $V = 2.405$ is chosen, then the fibre will guide only one fundamental mode in one polarisation plane. Then we have the so-called single mode or mono-mode operation. The corresponding wavelength λ_c for $V = 2.405$ is called the cut-off wavelength for the single mode operation. The single mode operation shows much better behaviour in terms of dispersion characteristics (pulse spreading) in comparison to the multi-mode operation, and is very important to achieve high bit rates, as will be discussed in the following sections.

11.2 Attenuation

Silica SiO_2 quartz glass shows a minimum attenuation of 0.24 dB/km at 1.55 μm and enables optical transmissions of about 70 km – 120 km without a so-called repeater. 3R (Reshaping, Retiming, Re-amplification) repeaters are needed in order to re-amplify the optical power, regenerate or refresh the optical pulse form, and retime the rising pulse edges before the pulses become overlapped due to dispersion or pulse spreading effect and the transmitted bits become errored. Therefore, the above-mentioned distance of 70 – 120 km is also called the repeater distance, which is an important planning parameter in the design and implementation of an optical fibre communications system.

Longer transmission distance can be achieved by using the 3R repeaters after the repeater distance to amplify the power, compensate for the so-called chromatic dispersion or pulse spreading effects, and refresh the pulse form with respect to the clock – also called retiming. The specific length-dependent loss coefficient α in dB/km is defined by the ratio of the input optical power P_i to the output optical power P_o along a fibre of the length L:

$$\alpha = \frac{10}{L} \cdot \log_{10}\left(\frac{P_i}{P_o}\right). \tag{11.3}$$

The loss of quartz glass optical fibres is basically caused by two factors: a) intrinsic losses and b) extrinsic losses. Intrinsic losses are fundamental losses of the optical waves in the form of absorption and scattering of photons through the silica quartz glass, interacting with silica quartz glass molecules and electrons in the form of photon absorptions. These absorptions are called IR (infra-red) absorption and UV (ultra-violet) absorption. The so-called Rayleigh scattering is caused by inhomogeneous silica quartz glass, where the spatial variation is much smaller than the wavelength.

Generally, the energy lost by the optical wave scattering and absorbed by the quartz glass molecules and electrons leads to a slight increase of the energy in the molecules and electrons. On the one hand, the infra-red IR absorption of photons leads to molecular vibrations of the pure silica glass, and the ultra-violet UV absorption of photons leads to higher energy levels of the electrons.

The so-called Rayleigh scattering is the third dominating intrinsic loss mechanism in the low wavelength range; it is caused by inhomogeneities on a small scale comparable with the wavelength of the optical waves. The inhomogeneities mean slight refractive index fluctuations which arise from the inhomogeneous distribution of the quartz glass within the optical fibre core region. Optical waves will be scattered in all directions proportional to λ^{-4}. The Rayleigh scattering decreases with wavelength (Fig. 69).

$$\alpha_{Rayleigh} \propto \lambda^{-4}. \tag{11.4}$$

The Rayleigh scattering formula can explain why the clear sky is normally coloured blue ($\lambda \sim 0.4$ μm), whereas the sunset through the clouds in the evening looks more red ($\lambda \sim 0.6$ μm). The sun light consists of the whole spectrum of different colours from red to blue and violet, even though not at same level. The red light of the sunshine suffers less scattering through the clouds, whereas the blue light is more strongly scattered in the sky.

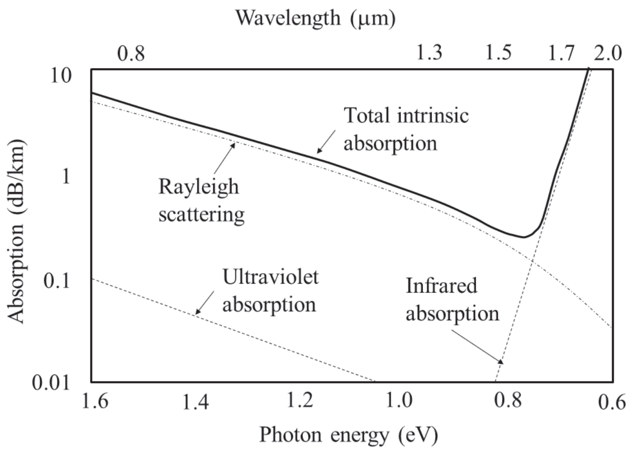


Figure 69 Intrinsic losses of the quartz glass optical fibre

The extrinsic losses are caused by impurities during the fibre manufacturing process, for example the OH ions (humidity or water) and metallic ions Cu, Cr, Fe, etc. in the production environment. In order to reduce the extrinsic losses, the production environment for the fibre preform should be extremely clean and dry.

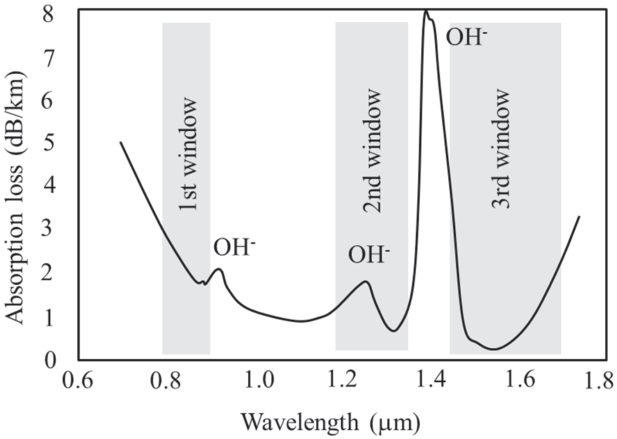


Figure 70 Total loss with extrinsic loss due to the OH-ions

Figure 70 shows the total loss including the intrinsic losses like Rayleigh scattering, IR absorption and UV absorption and the extrinsic loss caused by the OH ions. Due to strong OH ion absorption peaks at certain OH ion resonance frequencies, especially at 1.4 µm, one can use some wavelength ranges with local minimum absorption for optical signal transmission. These wavelength ranges are called optical windows at 800 nm, 1300 nm and 1550 nm. The absolute absorption minimum is at 1550 nm with 0.24 dB/km, which enables a repeater distance of approximately 100 km.

The modern fibre fabrication technology is able to reduce the OH ion to an absolute minimum, so that the OH absorption peaks vanish almost completely and the complete spectrum around 1.5 µm can be used for DWDM applications with 40 – 100 channels.

11.3 Modal Dispersion

In the case of a large core diameter of up to 50 µm of multi-mode fibres (Fig. 68), the waveguide allows the propagation of different waves or solutions of Maxwell´s equations, which are called “modes”. Due to different propagation velocities of these different modes, the original laser pulses are spread, limiting the transmission bit rates and repeater distance, since after a certain distance the timely subsequent pulses representing the independent bits 0s or 1s will overlap with each other if the bit rates are chosen relatively high. This or other pulse spreading effects are called “dispersion”.

This type of optical fibre with different modes, which have different propagation velocities, is called “multimode fibre” and the effect is called “intermodal dispersion” or simply “modal dispersion”. This is the most disturbing factor for the optical signal transmission, in comparison with other dispersion effects, which will be discussed later. Modal dispersion causes the largest delay difference, therefore leading to extremely low bit rates. This is also the reason, why the multimode fibre was only used in the LAN (local area networks), not in the WAN (wide area networks) or long distance backbone optical transport networks. By using multimode fibre only a few Mbps can be transmitted within a distance less than 10 km.

In the intercontinental or transoceanic optical fibre transmission networks only single mode fibres are used. For example, the standard single-mode fibre standardised in the ITU-T recommendation ITU-T G.652, or other dispersion optimised fibres like ITU-T G.653 (DSF dispersion shifted fibre) and ITU-T G.655 (NZDSF non-zero dispersion shifted fibre). As discussed at the beginning of this chapter, if the normalised frequency $V = 2.405$ is chosen, then only one single mode can propagate, so that the modal dispersion can be completely eliminated.

Generally the achievable bit rate B (unit bps) can be estimated by considering the pulse spreading factor $\Delta\tau$ (unit seconds per length, e.g. ps/km) and the length of the fibre L (for example in km) or the pulse spreading ΔT (unit second):

$$B \leq \frac{1}{2 \cdot \Delta\tau \cdot L} = \frac{1}{2 \cdot \Delta T}. \tag{11.5}$$

One of the planning parameters is the so-called BLP (Bit rate-Length Product), which can be used to characterise the performance of the optical fibre transmission system more comprehensively:

$$B \cdot L \leq \frac{1}{2 \cdot \Delta\tau}. \tag{11.6}$$

11.4 Chromatic Dispersion

If the normalised frequency $V = 2.405$ is chosen, corresponding to a diameter of about 9 μm, there exists only one single possible solution of the Maxwell´s equation in the cylindrical optical fibre, so that the modal dispersion disappears completely, leading to an increase in the possible bit rates.

But even by using the single mode fibre and by eliminating the modal dispersion of the multimode fibre, there are still other effects which spread the optical signal or optical pulses.

Optical waves propagate in the optical fibre ($\varepsilon_r = n_c^2(\lambda) = 1.45^2 \sim 2.1$) with the velocity $v(\lambda)$. $v(\lambda)$ will be reduced by n_c in comparison with c as light velocity in a vacuum or free space ($\varepsilon_r = 1$):

$$v(\lambda) = \frac{c}{n_c(\lambda)}. \qquad (11.7)$$

Figure 71 shows the wavelength-dependent refractive index of silica glass in the corresponding wavelength range from 0.6 – 2.0 μm. The optical transmitter, consisting of a laser diode with a central operation wavelength λ_o, e.g. 1.55 μm, has always a limited spectral linewidth $\Delta\lambda$ (definition: FWHM full width of half maximum). The linewidth can be reduced by using sophisticated techniques, but can never be completely reduced to zero. Therefore the optical pulses, on which the information bits will be modulated, also always consist of different wavelengths, the fast wavelength $\lambda_f = 1.56$ μm and the slow wavelength $\lambda_s = 1.54$ μm, when the linewidth is 20 nm. This means also, parts of the optical pulse will propagate with different velocities correspondingly, leading to pulse spreading, and lower achievable bit rate.

Example:

Taking the above example, then the photons of 1.56 μm will travel with slightly higher velocity than the photons of 1.54 μm. At the end after L = 50 km, the pulse spreads to ΔT:

$$\Delta T = \frac{L}{v_s} - \frac{L}{v_f} = \frac{L}{c}\left(n_c(\lambda_s) - n_c(\lambda_f)\right) \approx \frac{5\cdot 10^4 m \cdot 2\cdot 10^{-4}}{3\cdot 10^8 m/s} = \frac{1}{3}\cdot 10^{-7} s. \qquad (11.8)$$

Correspondingly the maximum bit rate of this optical transmission system can be roughly estimated to be less than

$$B \le \frac{1}{2\cdot\Delta T} = \frac{1}{2\cdot\frac{1}{3}\cdot 10^{-7} s} = 15 Mbps. \qquad (11.9)$$

This simple example shows that even after elimination of the most critical modal dispersion, the chromatic dispersion still limits the bit rates.

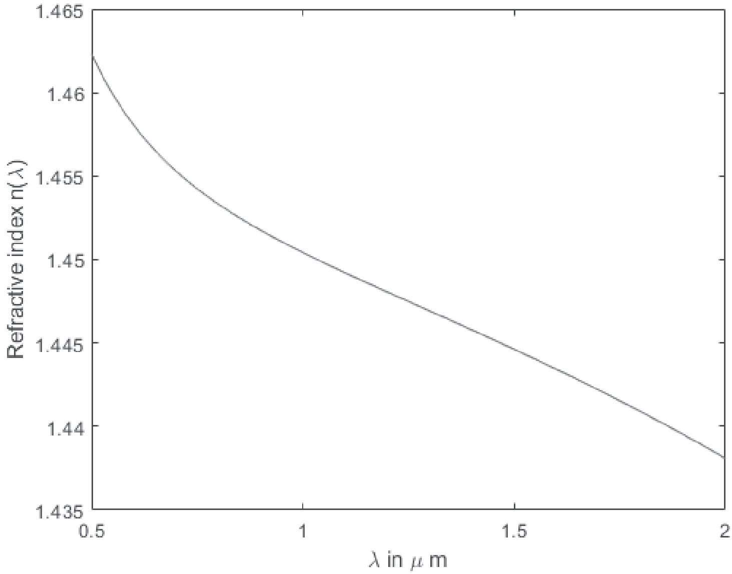


Figure 71 Refractive index *n* of the silica quartz glass

Due to the limited linewidth of the optical laser diodes and the different propagation velocities of the different optical wavelengths, pulse spreading is caused by "chromatic dispersion", consisting of "material dispersion", basically caused by the wavelength-dependence of the refractive index *n* (Fig. 71), and "waveguide dispersion", even though these effects are much smaller than the modal dispersion.

The phase delay time / distance is given by

$$\tau_{\varphi} = \frac{k}{\omega} = \frac{n}{c_0} \tag{11.10}$$

and the group delay time / distance by

$$\tau_g = \frac{dk}{d\omega} \approx \frac{\Delta k}{\Delta \omega}\Big|_{\Delta\omega \to 0} . \tag{11.11}$$

The group delay describes the propagation of optical pulses through the dielectric waveguide. Comprehensively speaking, the group delay describes the propagation of the pulse or the energy packet. The group delay basically determines the dispersion characteristics or the pulse spreading.

Generally the medium has no dispersion if $\tau_g = \tau_\varphi$, normal dispersion if $\tau_g > \tau_\varphi$, and abnormal dispersion if $\tau_g < \tau_\varphi$.

The group delay $\frac{d\tau_g}{d\lambda}$ can be calculated by**:**

$$\tau_g = \frac{dk}{d\omega} = \frac{1}{c}\left(n(\omega) + \omega\frac{dn}{d\omega}\right) = \frac{n}{c}\left(1 + \frac{\omega}{n}\frac{dn}{d\omega}\right) \qquad (11.12)$$

with $k(\omega) = \frac{\omega}{c} \cdot n(\omega)$. (11.13)

Using $c = \lambda \cdot f$ then $\frac{df}{d\lambda} = -\frac{f}{\lambda}$ (11.14)

$$\tau_g = \tau_\varphi\left[1 - \frac{\lambda}{n}\frac{dn}{d\lambda}\right] = \frac{1}{c}\left[n - \lambda\frac{dn}{d\lambda}\right] = \frac{N}{c} \qquad (11.15)$$

where *N* is called group (refractive) index. Compare also with (11.10).

$$\frac{d\tau_g}{d\lambda} = \frac{1}{c}\left[\frac{dn}{d\lambda} - \frac{dn}{d\lambda} - \lambda\frac{d^2n}{d\lambda^2}\right] = -\frac{\lambda}{c}\frac{d^2n}{d\lambda^2} = M(\lambda). \qquad (11.16)$$

This term is defined as material dispersion, measured in $ps/(nm \cdot km)$. If however the material dispersion $M(\lambda) = \frac{d\tau_g}{d\lambda} = 0$, then the second derivative $\frac{d^2n}{d\lambda^2} = 0$ will be zero correspondingly. Mathematically it means the material dispersion will be zero at the turning point of the curve $n = f(\lambda)$ at wavelength λ = 1.273 µm. In this region we will have almost no material dispersion effects if the spectral linewidth Δλ of the laser diode is very small. The effect of material dispersion near to the turning point, where $M(\lambda) \neq 0$, is then influenced mainly by the spectral width of the laser diode.

So far, we have analysed the propagation of optical pulses in a homogeneous medium with infinite dimension. Taking the geometrical limits into consideration, which are given by the fibre structure, we find the chromatic dispersion for pulses travelling through the fibre. There, the propagation is characterised by β, instead of k as in the homogenous space. Correspondingly, the derivative of τ_g is defined as chromatic dispersion of the waveguide [10].

$$\frac{d\tau_g}{d\lambda} = C(\lambda). \tag{11.17}$$

$$\frac{d\beta}{d\omega} = \tau_g = \frac{d\beta}{dk_c} \cdot \frac{dk_c}{dk} \cdot \frac{dk}{d\omega}. \tag{11.18}$$

We want to analyse these three terms separately.

$$\frac{dk_c}{dk} = \frac{d(k \cdot n_c)}{dk} = n_c + k\frac{dn_c}{dk} = n_c - \lambda\frac{dn_c}{d\lambda} = N_c \tag{11.19}$$

where N_c = group index of core material;

$$\frac{dk}{d\omega} = \frac{d\left(\frac{\omega}{c}\right)}{d\omega} = \frac{1}{c}. \tag{11.20}$$

Using the results (11.16) and (11.17), we find

$$\tau_g = \frac{N_c}{c} \cdot \frac{d\beta}{dk_0} = \frac{1}{c}\left(n_c - \lambda\frac{dn}{d\lambda}\right) \cdot \frac{d\beta}{dk_c}. \tag{11.21}$$

Comparing with equation (11.15):

$\frac{N_c}{c}$ corresponds to group delay time of a homogeneous wave in the homogeneous core material. $\frac{d\beta}{dk_c}$ is important for analysing fibres by taking into consideration the geometry.

$$\frac{d\tau_g}{d\lambda} = \frac{d}{d\lambda}\left[\frac{1}{c}\left(n_c - \lambda\frac{dn_c}{d\lambda}\right)\frac{d\beta}{dk_c}\right] = -\frac{\lambda}{c}\frac{d^2 n_c}{d\lambda^2}\cdot\frac{d\beta}{dk_c} + \frac{N_c}{c}\cdot\frac{d^2\beta}{dk_c d\lambda}$$
(11.22)

which results to

$$C(\lambda) = -\frac{\lambda}{c}\frac{d^2 n_c}{d\lambda^2}\cdot\frac{d\beta}{dk_c} - \frac{k}{c\cdot\lambda}N_c^{\ 2}\cdot\frac{d^2\beta}{dk_c^{\ 2}} \qquad (11.23)$$

Equation (11.23) corresponds to the complete mathematical correlation for the chromatic dispersion. The first term corresponds to material dispersion $M(\lambda)$ of the core material multiplied with the term $d\beta / dk_c$., which can be approximated by $(1+\tau)$. The second part describes the waveguide dispersion $W(\lambda)$

$$W(\lambda) = -\frac{k}{c\cdot\lambda}\cdot N_c^{\ 2}\frac{d^2\beta}{dk_c^{\ 2}} = \frac{1}{c}N_c\frac{d\tau}{d\lambda} \qquad (11.24)$$

with the distribution of energy travelling through the fibre, influenced by the core and cladding material and the fibre geometry.

$$C(\lambda) = M(\lambda)\cdot(1+\tau) + W(\lambda). \qquad (11.25)$$

For $\tau << 1$, which is valid, since $n_{cl}\,k < \beta < n_c\,k$ corresponds to the relative refractive index difference of 1%, we then have

$$C(\lambda) \approx M(\lambda) + W(\lambda). \qquad (11.26)$$

This formula is most important. It indicates that the material dispersion can be compensated for by the waveguide dispersion, if the waveguide structure and parameters are properly chosen. This leads to the design of DCF (Dispersion Compensating Fibre), DSF (Dispersion Shifted Fibre, where the zero dispersion wavelength will be shifted to a desired wavelength) and DFF (Dispersion Flattened Fibre, where a larger wavelength region shows – not necessarily zero – but reduced dispersion, and which is best for the DWDM applications), by using for instance a triangle fibre core refractive index profile or a multi-layer W-profile.

11.5 Polarisation Modal Dispersion

Up to now, we have tried to optimise the dispersion characteristics step by step, firstly eliminating the modal dispersion by choosing a single mode fibre, then by compensating the material dispersion by using a suitable waveguide dispersion. Also the circular cross section is assumed to be absolutely perfect. But in real life the cross section is never perfect, leading to the so called "birefringence", i.e. in the x-polarisation plane the wave is propagating with a slightly different velocity than in the y-polarisation plane. Again, the polarisation plane is defined as the plane defined by the electric field vector.

There are different possibilities of orientation of the electrical field component, also described by the polarisation of light. Sunlight is unpolarised, the electrical field vector is orthogonal to the z-axis, the direction of the travelling wave, and the x and y components are changing randomly.

In contrary to sunlight, the light of a laser diode is highly polarised. For many applications, it is important to study the influence of different elements (fibre, filter, polariser, etc.) to the changing polarisation states of the wave travelling through them. Generally we have 1) linear polarisation; 2) circular polarisation; 3) elliptical polarisation.

Some materials have absorption characteristics which depend on the different states of polarisation. The absorption depends on the orientation of the electrical field vector with respect to the optical axis of the crystal, the main axis, leading to the dichroism.

If the internal structure of the material is not completely symmetric in every orientation, this material is anisotropic or birefringent.

The polarisation characteristics can be described completely by using the Jones vector and the Jones matrix. Assuming a monochromatic light wave propagating into the positive direction of the z-axis of an x-y-z orthogonal, right-handed Cartesian co-ordinate system, the electric vector can be generally written as

$$\vec{E}(z,t) = E_x \cos(\omega \cdot t - k \cdot z + \phi_x) \cdot \vec{e}_x + E_y \cos(\omega \cdot t - k \cdot z + \phi_y) \cdot \vec{e}_y . \quad (11.27)$$

E_x and E_y represent the amplitudes of harmonic oscillation along the x and y axes, k is the wave number

$$k = \frac{2\pi}{\lambda} . \quad (11.28)$$

Equivalently, the field equation can be written as a matrix

$$\vec{E}(z,t)=\begin{bmatrix} E_x \cos(\omega\cdot t-k\cdot z+\phi_x) \\ E_y \cos(\omega\cdot t-k\cdot z+\phi_y) \end{bmatrix}. \tag{11.29}$$

Now, if the components oscillate sinusoidally with time at the same angular frequency, only the phase shifts of the x- and y-field components decide the polarisation state.

$$\vec{E}(z)=\begin{bmatrix} E_x\cdot e^{j\cdot\phi_x} \\ E_y\cdot e^{j\cdot\phi_y} \end{bmatrix}\cdot e^{-j\cdot k\cdot z}. \tag{11.30}$$

If we consider the field components at z=0, then this field vector describes the monochromatic, uniform, transverse electric field

$$\underline{\vec{E}}(0)=\begin{bmatrix} E_x\cdot e^{j\cdot\phi_x} \\ E_y\cdot e^{j\cdot\phi_y} \end{bmatrix}=\underline{\vec{E}}. \tag{11.31}$$

By using this Jones vector the complete wave can be reconstructed by adding the time and space dependency, if necessary.

$$E(z,t)=\mathrm{Re}\left\{\underline{\vec{E}}(0)\cdot e^{j\cdot(\omega\cdot t-k\cdot z)}\right\}. \tag{11.32}$$

$$\underline{\vec{E}}=\begin{bmatrix} \underline{E}_x \\ \underline{E}_y \end{bmatrix}=\begin{bmatrix} E_x e^{j\cdot\phi_x} \\ E_y e^{j\cdot\phi_y} \end{bmatrix}. \tag{11.33}$$

The phasor components, $\underline{E}_x$ and $\underline{E}_y$, can have all possible values, meaning all possible states of polarisation:

If $\phi_x-\phi_y=0$ or $\phi_x-\phi_y=\pi$, we have linear polarisation

$$\underline{\vec{E}}=\begin{pmatrix} E_x \\ E_y \end{pmatrix}. \tag{11.34}$$

If $\Delta\phi=\pm\dfrac{\pi}{2}$; $E_x=E_y=E$, we have circular polarisation.

$$\underline{\vec{E}} = E\begin{pmatrix} 1 \\ j \end{pmatrix} \qquad \text{right-hand circularly polarised.} \qquad (11.35)$$

$$\underline{\vec{E}} = E\begin{pmatrix} 1 \\ -j \end{pmatrix} \qquad \text{left-hand circularly polarised.} \qquad (11.36)$$

For all other cases, we have generally elliptical polarisations ($E_x \neq E_y$):

$$\underline{\vec{E}} = \begin{pmatrix} E_x \\ E_y \cdot e^{j\cdot\phi} \end{pmatrix} \qquad (11.37)$$

If an optical wave, described by the Jones vector, travels through an optical element which changes the amplitude and the phase of the incoming wave $\underline{E_1}$, then we can express the outgoing field $\underline{E_2}$ by the following equation:

$$\underline{\vec{E}}_2 = \begin{pmatrix} E_{2x} \\ E_{2y} \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \cdot \begin{pmatrix} E_{1x} \\ E_{1y} \end{pmatrix} = \overline{T} \cdot \underline{E}_1 \qquad (11.38)$$

$$\underline{\vec{E}}_1 \longrightarrow \boxed{\overline{T}} \longrightarrow \underline{\vec{E}}_2 \,.$$

The matrix $\overline{T}$ corresponds to the Jones matrix of the optical element describing the characteristics. Correspondingly, if an optical component, for example an optical fibre of the length *L*, changes its characteristics along the length, this can be modelled by a series connection of many Jones matrices

$$\underline{\vec{E}}_{\text{out}} = \overline{T}_3 \cdot \overline{T}_2 \cdot \overline{T}_1 \cdot \underline{\vec{E}}_{\text{input}} \,. \qquad (11.39)$$

By using this method, all the features of optical elements which influence the incoming optical waves can be analysed numerically.

For example, by using the linear polarisor:

$$\overline{T} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad (1.40)$$

the wave vector of one direction passes through the polarisor, the vector of orthogonal direction is blocked.

More realistically, by considering the imperfections of the optical components, we have

$$\overline{T} = \begin{pmatrix} \cos^2\alpha & -\sin\alpha\cos\alpha \\ \sin\alpha\cos\alpha & \sin^2\alpha \end{pmatrix}. \qquad (11.41)$$

In a birefringent material the two orthogonal wave components have different refractive indices, corresponding to different propagation velocities. So, if there is a phase difference for the two orthogonal directions, passing a birefringent material of length L, the wave experiences an optical phase shift. Therefore this kind of optical component is also called an optical retarder or retarding plate. By using an optical retarder, the polarisation state of the incoming optical waves can be changed or manipulated to another polarisation state. For example, if we use one optical retarder which shifts the phase between x- and y-plane to $\Delta\phi$

$$\Delta\phi = k \cdot L \cdot \left(n_x - n_y\right) = \phi \qquad (11.42)$$

we have then the Jones matrix of the optical retarder

$$\overline{T} = \begin{pmatrix} 1 & 0 \\ 0 & e^{j\cdot\phi} \end{pmatrix}. \qquad (11.43)$$

Example:
For the special case of $\lambda/4$ or 90° phase shift:

$$\phi = \pm\frac{\pi}{2}$$

we will have the Jones matrix

$$\overline{T}_{\lambda/4} = \begin{pmatrix} 1 & 0 \\ 0 & \pm j \end{pmatrix}. \qquad (11.44)$$

A linearly polarised incoming wave E_1 with $\theta = 45°$ to the x-axis:

$$\vec{\underline{E}}_1 = \frac{E}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{11.45}$$

will be converted to the outgoing wave E_2, which becomes circularly polarised:

$$\vec{\underline{E}}_2 = \frac{E}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 0 \\ 0 & \pm j \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{E}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ \pm j \end{pmatrix} \tag{11.46}$$

$$E_x(t,z) = \frac{E}{\sqrt{2}} \cdot \cos(\omega \cdot t)$$

$$E_y(t,z) = \frac{E}{\sqrt{2}} \cdot \cos\left(\omega \cdot t + \frac{\pi}{2}\right).$$

Inversely, circularly polarised light can be changed to linearly polarised light by travelling through a $\lambda/4$ plate.

Example:
By using a $\lambda/2$ retarder (180°):

$$\phi = \pi$$

$$\overline{T}_{\lambda/2} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{11.47}$$

the right-hand circularly polarised light will be converted to a left-hand circularly polarised light.

$$\vec{\underline{E}}_2 = \frac{E}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ +j \end{pmatrix} = \underline{\frac{E}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ -j \end{pmatrix}}. \tag{11.48}$$

And the linearly polarised light will retarded by 90°, if the $\lambda/2$ is arranged by $\theta = 45°$ to the direction of wave polarisation.

$$\underline{\vec{E}}_2 = \frac{E}{\sqrt{2}} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \underline{\frac{E}{\sqrt{2}} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}}. \qquad (11.49)$$

Therefore, by using the Jones matrix and Jones vectors, different polarisation effects can be investigated, especially the polarisation modal dispersion effect, which is caused by the different propagation velocities along the imperfect fibres' geometry and refractive index distribution.

The polarisation modal dispersion effect is very weak in comparison with the modal dispersion effect and the chromatic dispersion effect. It can be perceivable, if the critical modal dispersion effect is eliminated and the chromatic dispersion is optimised by using DCF, DSF and DFF, as discussed in the last section. This is the case when an optical transmission network is optimised and designed for high bit rate transmission like 10 Gbps or 40 Gbps.

11.6 Nonlinearities

In the case of high optical power P or intensity I of the transmitter, the refractive index n_1 of the quartz glass with an effective cross section A is not linear. The refractive index will change correspondingly by considering the non-linear characteristic coefficient n_2:

$$n = n_1 + n_2 \cdot \frac{P}{A} = n_1 + n_2 \cdot I \qquad (11.50)$$

Nonlinearity causes different effects: SPM (Self Phase Modulation) and FWM (Four Wave Mixing). These effects can be disturbing. Four Wave Mixing can affect the high bit rate transmission of DWDM (Dense Wavelength Division Multiplex) optical transmission systems, where 40 – 100 optical channels are transmitted within one optical fibre by using different wavelengths with very small channel spacing. The new disturbing fourth wave generated by the nonlinear effect or the intermodulation between three signal wavelengths, sometimes is close to the signal carrier, leading to increased interferences.

Other nonlinear effects like SRS (Stimulated Raman Scattering) and SBS (Stimulated Brillouin Scattering) are so-called inelastic scattering effects, in comparison with the elastic Rayleigh scattering which does not change the wavelength of the incident optical wave. SRS and SBS change the wavelengths or frequencies of the incident optical wave. The difference of the frequencies between the incident and scattered waves are 11 THz (80 – 100 nm) for SRS and 15 GHz for SBS.

This inelastic scattering effect could be used in the fibre amplifier like EDFA (Erbium-Doped Fibre Amplifier), which is nowadays commonly used in long-distance transmission. In this case the power of the pumping laser at 0.98 μm or 1.48 μm can amplify the signal at 1.55 μm.

Also the self-phase modulation effect can be utilised to compensate the chromatic dispersion and to produce the so-called solitary waves, or solitons. Solitons maintain the pulse form over thousands of kilometres and do not need 3D repeaters, only re-amplifiers to ensure the necessary power level, which is one important pre-condition for the solitons. In nature, a similar effect can be seen when a Tsunami wave travels over thousands of kilometres without changing the wave amplitudes.

11.7 Laser Diodes as Transmitter

In modern optical fibre communications, laser diodes are used as transmitters, modulated with the information data stream. Whereas the LED (Light Emitting Diode) has a relatively large spectral linewidth, the laser diodes can have a very small spectral linewidth. Also LED light is incoherent, whereas laser light is highly coherent beyond the threshold pump current. Below the threshold pump current, laser diodes behave like LEDs.

As we discussed in the last section, the narrow spectral linewidth means also low chromatic dispersion and higher possible bit rates for the optical communication system.

The LED is based on the so-called spontaneous emission of photons, if the LED is pumped by the electric current.

LASER is the abbreviation of Light Amplification by Stimulated Emission of Radiation. A laser diode is further improved in comparison with a LED by the following measures:

- A Fabry-Perot resonator is implemented, to achieve multiple reflexions of the photons in the active zone between both laser facets. This allows stimulated emission, in order to produce more photons with the same physical features like wavelength and polarisation, or coherence.
- Pump current is larger than the threshold current I_{th}.
- Population inversion is maintained by permanent electric pumping energy, in order to ensure sufficient electrons in the higher energy levels and to maintain stimulated photon emission.

- III/V-semiconductor compounds like InGaAsP are used to cover the spectral range of 1.5 µm.
- The so-called buried-hetero-structure with alternating InGaAsP and InP layers with different refractive indices form the waveguide structure or index-guidance, and enable the so-called longitudinal single mode, and therefore narrow spectral linewidth (Fig. 72). The width and the height of the active zone are about 2 µm and 0.1 µm. The resonator length is about 200-300 µm [84] [85].
- Special laser diodes like DFB (Distributed Feed Back), DBR (Distributed Bragg Reflexion) (Fig. 73) [86] [87] and cleaved coupled cavity laser diodes further reduce the spectral linewidths of the index-guided laser diode, by using a grating structure inside or outside the active zone, and by using an external resonator, in order to suppress the spectral side lobes and achieve narrower spectral linewidth.
- By using cleaved coupled cavity lasers with a second or third dumping current in an additional active zone, the emission wavelength of the laser diode can be tuned in a certain range (Fig.74). This is extremely interesting, since in the DWDM application many transmitters with 40-100 different wavelengths are needed.
- Also laser diodes with an external cavity and cleaved cavity lasers can be used to narrow the spectral linewidth and to tune the wavelength (Fig.74).

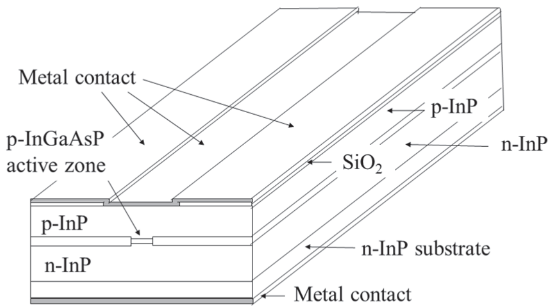


Figure 72 Index-guided InGaAsP buried-hetero-structure laser diode

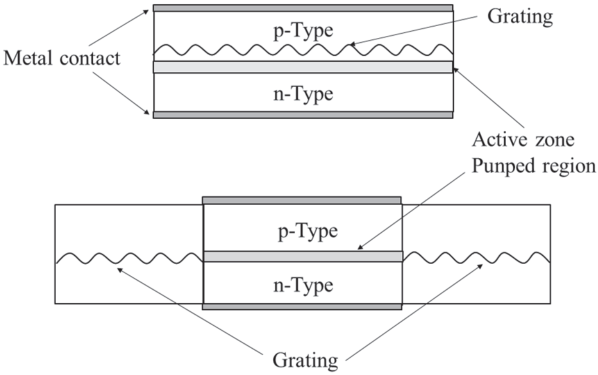


Figure 73 DFB and DBR laser diode

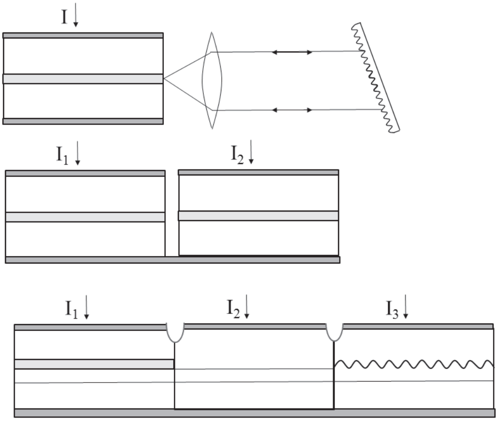


Figure 74 Tunable laser diodes

11.8 Photodiodes as Receiver

In the last section we discussed laser diodes based on the electro-optic principle of stimulated emission. The coherent photon emission can be produced by an electric pumping current beyond the threshold. The other, somehow the reverse electro-optic effect, is the absorption of photons, leading to the generation of electric current in a photodiode, both in visible light or the infrared wavelength range. The most famous photodiode is the PIN photodiode (Fig. 75).

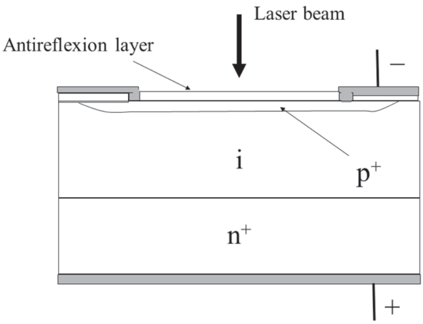


Figure 75 PIN photodiode

In order to further improve the characteristics of PIN photodiodes, some other special structures have been developed [85] [88]: 1) PIN photodiode with internal multiple reflecting gratings (Fig. 76) to reduce the electrical response time and at the same time increase the optical length of the intrinsic layer, where the photons are absorbed; 2) PIN photodiode with lateral light incidence (Fig. 77) to increase the optical length without increasing the intrinsic layer thickness; 3) APD (Avalanche Photodiode, Fig. 78) to achieve built-in amplification due to the avalanche effect with multiplication factors up to more than 100, so that an external amplifier can be replaced.

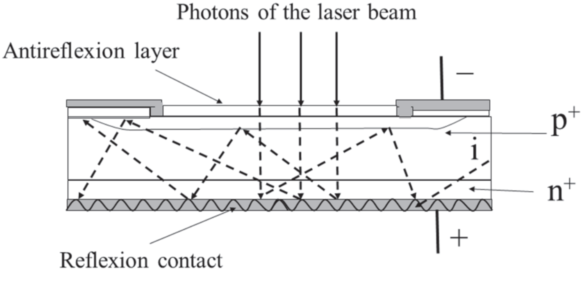


Figure 76 PIN with internal multiple reflecting gratings

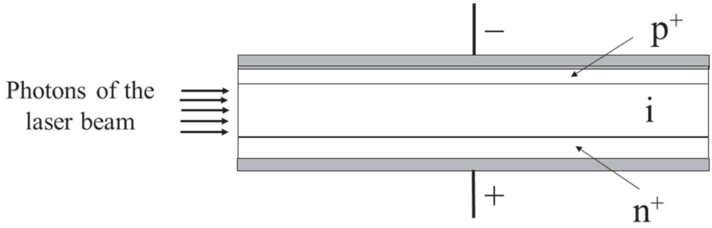


Figure 77 PIN photodiode with lateral insertion of photo current

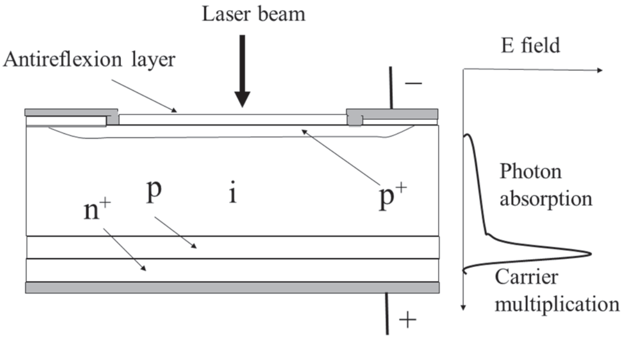


Figure 78 Avalanche Photodiode APD

Both Figure 76 and Figure 77 are designed to reduce the drift time of the electrons, and therefore the response time of the photodiode. For both receivers an amplifier is needed to increase the relatively low photocurrent for the following demodulation circuit. This amplifier can be omitted when APD (Avalanche Photo Diode) with built-in amplification, approximately by a factor of 100-500, can be used (Fig. 78).

11.9 DWDM and CWDM

In order to increase the total capacity of an optical communication system, not only one single channel at one single wavelength is applied, but up to 16, 40 or 160 optical channels with different wavelengths are modulated with the payload data bit streams, so that up to 10 Tbps can be transmitted over one single optical fibre. In the following table the bands for optical communications are depicted:

Band	Description	Wavelength range
O-Band	Original	1260–1360 nm
E-Band	Extended	1360–1460 nm
S-Band	Short wavelength	1460–1530 nm
C-Band	Conventional	1530–1565 nm
L-Band	Long wavelength	1565–1625 nm
U-Band	Ultralong wavelength	1625–1675 nm

In the C- and L-Band channel spacings of 0.4 nm (50 GHz) - 1.6 nm (200 GHz) are possible. Each channel can transmit 10-100 Gbps. Correspondingly the linewidth of the laser diodes should be less.

In city networks or LAN, the CWDM (Coarse Wavelength Division Multiplex) with 8 – 16 channels can be used with channel spacing of 20 nm and 10 Gbps per channel.

CHAPTER 12

FREE SPACE OPTICAL COMMUNICATIONS

In the last chapter, we discussed the basics of optical communications with fibres as transmission medium. Instead of fibres, free space may also be used. Free space optical communications (FSOC) technology can be flexibly deployed in comparison with the optical fibre communications. Free space optical communications can also enable smaller device sizes and higher bit rates than radio frequency and microwave systems [89].

Deep space optical communications (DSOC) have been investigated in the past decades by NASA and ESA, in order to further reduce the size and weight of the radio frequency transmitter and receiver antennas for deep space communications.

12.1. Background

Free space optical communications is a technology to transmit data by the propagation of infrared waves in free space. The systems consist of an optical transceiver at both ends to provide bidirectional transmission. The attempt is very similar to optical fibre communication systems, just without the need of a dedicated quartz glass optical fibre as transmission medium. The basic idea of free space optical communications itself is not new, but with the invention of laser diodes at infrared wavelengths and highly sensitive photodiodes, as well as sophisticated modulation schemes, it now meets the specifications for the rising demand of high bandwidth connections [90].

Free space loss increases with λ^{-2}. The antenna or lens gains increase with λ^{-2}. Thus the total power loss or attenuation decreases with λ^{-2} for equal size of antenna or lenses. Additional to this fundamental free space loss and gains we have to consider attenuation by particles, such as fog, precipitation and so on.

There are multiple applications where the advantages of modern FSOC can be utilised. Unlike the limited spectrum availability with radio frequency wireless communications, optical communication has a license-free spectrum [91]. Since there is no fibre cable required, a connection can be established very flexibly and quickly with free visual contact or LOS (line of sight) between the transmitter and the receiver. This does not only enable quick connectivity after natural disasters, for example, but also a stable connection with moving objects like airplanes or even satellites. The LOS must be guaranteed. Always the transceivers must be adjusted very precisely to prevent pointing errors. And even if this is the case, unpredicted obstacles or bad atmospheric conditions in the line of sight can still influence the signal quality negatively. The primary factors that deteriorate the link performance are absorption, scattering and turbulence [92].

12.2. Free Space Optical Link

A FSOC system (Fig. 79) contains three primary subsystems: transmitter, channel and receiver [89] [93]:

Transmitter: Its primary function is to modulate the incoming message signal onto the optical carrier which is then propagated through the atmosphere to the receiver. Essential components are the modulator, the optical source which is most likely a laser diode, and a transmitting telescope or optical antenna. The telescope is not only responsible for beam shaping but also the movable part for beam tracking.

Channel: The channel is a major limiting factor of FSOC since unpredictable conditions like snow, fog, rain or clouds can influence the strength of the received signal. It is important to mention that these factors do not have fixed characteristics and change significantly with time and geographical zones.

Receiver: It is used to recover the transmitted data. It consists of a receiver telescope which is the equivalent to the transmitter telescope, a very sharp optical bandpass filter to reduce the background noise and limit the received signal bandwidth, and the demodulator.

12.3. Channel Model with Different Factors

The selection of optical wavelengths depends on multiple factors. For terrestrial FSOC systems costs, eye safety and transmission windows are critical. Regarding eye safety regulations, 1.5 µm and larger wavelengths are always the preferred option since wavelengths between 400 nm and 1400 nm are hazardous because they are focused directly on the retina [94].

The transmission windows arise from the radiation absorption of molecules in the atmosphere. Certain gases such as water, carbon dioxide or ozone absorb electromagnetic waves at a particular frequency. In contrast to these absorption bands, there are spectral areas with little or no absorption to specific wavelengths – these are known as atmospheric windows. These wavelengths are preferred for a low attenuation optical transmission [95]. Other limiting factors like scattering are also dependent on the wavelength. The most significant loss in FSOC is usually the beam divergence loss, which describes the attenuation between two antennas. The divergence loss is dependent on the propagating beam width and the diameter of the receiving telescope, which can be noted as an equivalent to the antenna gain in radio transmission. It also depends on the propagating wavelength because of the definition of free space loss. The formula for the calculation of the received power P_R is:

$$P_R = P_T \cdot G_T \cdot G_R \cdot L_P \tag{12.1}$$

where L_P is the free space path loss, G_T and G_R are the effective gains of the receiver and transmitter antenna or telescope, P_T is the transmitted power and P_R is the received power. The gain of the optical antennas or telescopes is approximated by:

$$G = \frac{4 \cdot \pi \cdot A}{\lambda^2} \tag{12.2}$$

where λ is the wavelength of the transmitted signal and A the effective optical antenna area. Depending on the efficiency and optical antenna characteristics, the total gain can vary [95]. With this equation in mind, a narrow beam divergence is preferable, with the condition that the antennas are perfectly aligned. Otherwise active beam tracking and pointing systems are required to reduce the pointing loss that occurs when the centre of the beam does not directly face the receiver. The free space loss factor is given by:

$$L_P = \left(\frac{\lambda}{4 \cdot \pi \cdot R} \right)^2 \tag{12.3}$$

where R is the distance between the transmitting and receiving optical antennas.

As already mentioned in the introduction, the quality of a FSOC link can vary strongly depending on weather conditions that affect the line of sight. Fog, rain, and snow decrease the received signal power. Additionally, particle attenuation must be considered. If the number of particles in the atmosphere reaches a certain threshold, a complete link outage can occur.

For a theoretical approach, this effect is divided into three different ranges with regard to the size of the particles. Fog consists of tiny particles that lead to Mie scattering while raindrops and snowflakes are comparatively larger and produce deeper fades in the signal [96] [97]. The effect of fog can be calculated with the Mie scattering theory, which is very sophisticated and requires detailed fog parameters. Alternatively, an approach based on visibility range information can be used [98]. It defines the attenuation coefficient of fog given by an empirical model for Mie scattering. The specific attenuation A in dB/km is derived by

$$A = 10 \cdot \ln\left[\gamma(\lambda)\right] = 10 \cdot \ln\left[\frac{3.91}{V} \cdot \left(\frac{\lambda}{550nm}\right)^{-p}\right] \tag{12.4}$$

where V stands for the visibility range in km and λ is the operating wavelength. p is the size distribution coefficient that describes the "thickness" of the fog. 550 nm is the reference wavelength used in this approach. The Kruse model [99] defines p for different visibility factors:

$$\begin{aligned} p &= 1.6 && V > 50 \\ p &= 1.3 && 6 < V < 50 \\ p &= 0.585 \cdot V^{1/3} && V < 6 \end{aligned} \tag{12.5}$$

Unaffected by the visibility range, the attenuation factor for fog is always dependent on propagating wavelength. A larger wavelength leads to lower attenuation.

Rain and snow attenuation are not affected by the propagating wavelength. The specific attenuation increases with the rainfall or snowfall rate.

Inhomogeneous temperature and pressure lead to various refractive index values in the atmosphere that directly influence the transmitted laser beam negatively. Turbulent cells are formed by these factors. These cells, also called eddies, have various sizes and different refractive indexes. Because these cells are distributed very randomly, calculating this influence is very sophisticated. When an optical signal is distributed through a turbulent atmosphere, it causes wave front distortion in phase and amplitude [100]. This can lead to a complete link failure. When the cells are larger than the beam size, a phenomenon called beam wandering occurs. Due to the change of refractive indexes in the cells, the distribution angle of the beam changes –this is the so-called turbulence.

By using OptiSystem by Optiwave [101] for a project [89], different modulation techniques can be used to simulate the propagation of a laser beam in free space. In this project two modulation schemes were used and investigated: 16-QAM and DP-QPSK.

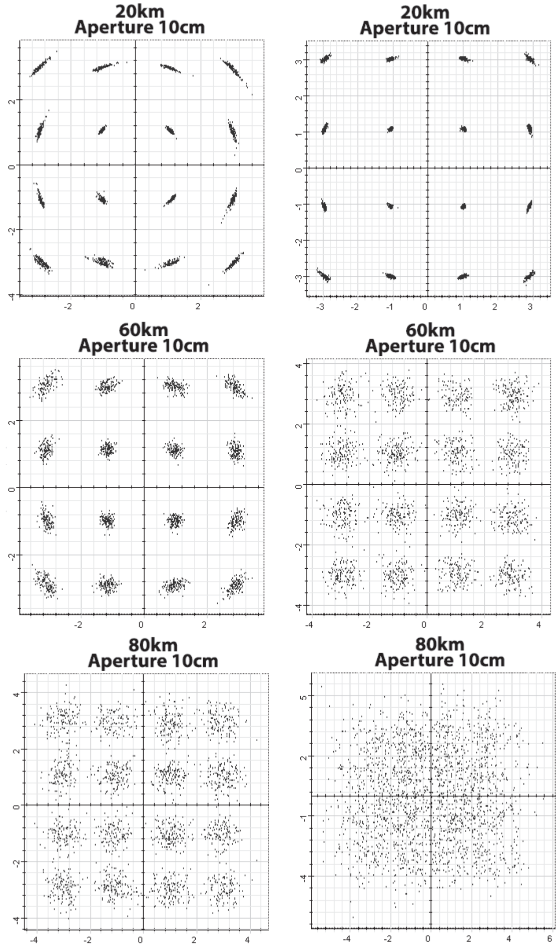


Figure 79 Constellation diagrams 16-QAM with 10 and 100 Gbps

The constellation diagrams in Figure 79 give a better impression of the different effects such distance and channel attenuations. The dots are getting larger depending on the increasing distance, and some bit errors occur already at earlier distances, though these are too small to be seen in the figures. At this point the transmitted signal quality is still good enough to demodulate the message. It is also shown clearly that a lower bit rate leads to a higher transmission distance.

In the constellation diagram one can clearly recognise that with a high bit rate of 100 Gbps, transmission over larger distances is no longer possible, either for DP-QPSK or for 16-QAM.

For larger distance and higher attenuation, lower order modulation schemes and lower bit rates have to be used, for example PPM in DSOC (deep space optical communications).

The free space optical communications FSOC performances have been investigated by considering the atmospheric channel, distance-dependent beam divergence, influence of the weather conditions, turbulences and scattering. Interesting results with different design parameters like distances between the transmitter and receiver, modulation schemes applied, aperture diameters of the transceivers and influence of fog were achieved and discussed.

The bit rates 10 Gbps and 100 Gbps can be increased correspondingly by a certain factor, if CWDM (Coarse Wavelength Division Multiplex) or even DWDM (Dense Wavelength Division Multiplex), which are now standard in the fibre optical communication systems, can be utilised. Tbps could then be possible.

12.4. Deep Space Optical Communications

Compared with the classical radio frequency deep space communications in S, X and Ka bands, the deep space optical communication shows advantages in terms of beam divergence (improvement of a factor of 100) and high bit rate delivery with significantly reduced aperture size of the terminals. Many research projects have been running and many interesting publications are available (for example, [102] [103]). Projects have been planned both at NASA and ESA. The general basic parameters and assumptions for investigations in the research projects to achieve the DSOC performance are the following [104] [105]:

- λ = 1.55 μm, wavelength of the downlink (spacecraft to earth) laser diode, i.e. a InGaAsP laser diode with an aperture of 0.22 m.
- P_{Tx} = 5 W total transmit power of the downlink laser diode.

- An array of up to nine distributed optical antennas distanced up to 100 m away from to each other on earth, and each with diameter 1m - 5m, are used to aggregate the photon counting.
- Uplink laser power should be 5 kW with a wavelength of 1.064 μm.
- Deep space distance of l = 0 - 2.5 AU has to be investigated. 1 AU (astronomic unit) corresponds to 150109 km.
- M = 16 or 64 for M-nary PPM coding is considered.
- Wanted bit rate will be 100 kbps - 100 Mbps in the future.
- Detector efficiency is assumed to be 0.35.
- Wavelength dependent transmission efficiency through atmosphere will be 0.55.
- Pointing loss for the laser beam is assumed to be 2 dB.
- FEC techniques in combination with interleaving techniques and sophisticated channel coding techniques are always indispensable.
- Proposals like micro-lens APD array in order to achieve high photo detection efficiency are assumed, even though the system was initially designed to absorb the light of 1.064 μm.
- Spatial acquisition adaptation, clock synchronisation, photon counting and Geiger-mode APD array design are fundamentally important for a powerful and efficient DSOC system.
- The background sky radiance (8.5 $\mu W/m^2/A\text{-}sr$), but also the sun radiance, lead directly to noise, leading to strong degradation or temporal outage.

Initially designed for the MLCD (Mars Laser Communication Demonstration) project which was cancelled by NASA in 2005 together with the MTO (Mars Telecommunications Orbit) project, a bit rate of 30 Mbps was planned for the nearest Mars-Earth distance of 0.7 AU.

Many concepts concerning efficient APD photon counter receiver systems and PPM coding were proposed in order to achieve an optimum system performance.

In M-nary PPM coding, each symbol is divided into M time slots. The entire energy for each symbol is concentrated in exactly one of the M slots, yielding $\log_2(M)$ bits for every pulse.

For a given M, the peak power in each symbol is M times the laser transmitter's average power. Consequently large M can be used to achieve high peak powers, which is particularly advantageous in high background environments (background sky radiance and partial solar radiance interference). In order to reduce this background radiance, a special polarisation of the laser can be used for the signal transmission. By using a polarisation filter at the receiver and a narrow bandwidth filter (0.1 nm), this spurious background sky or solar radiance effect can be minimised. At the same time the temporal filtering to detect the one signal pulse out of M slots contributes also to the noise reduction. The spatial filtering and adaption help to focus on the signal and relatively reduce the background noise, too.

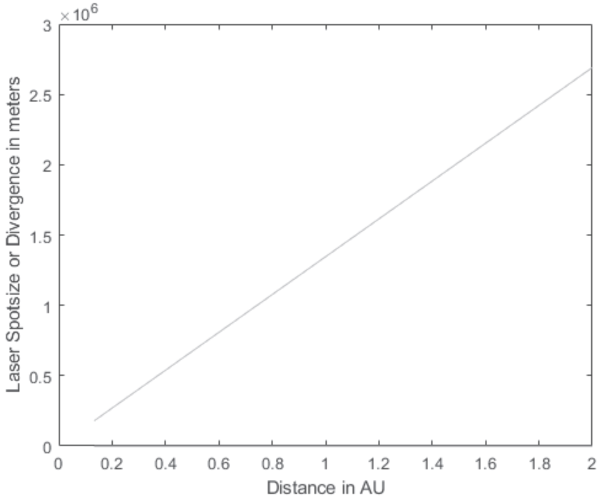


Figure 80 Laser spot size versus distance

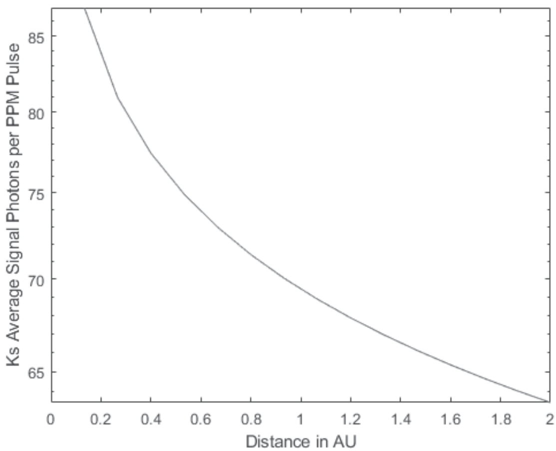


Figure 81 Photons per PPM pulse versus distance

The first simulation results for the downlink DSOC system performance analysis have shown strongly reduced received laser power for large distances up to 2.5 AU, with only a small number of photons per PPM pulse, or only a few photons per bit.

Figure 80 shows the dependence of the laser spot size on the distance, or the divergence of the beam width. This effect leads to the decrease of the transmission power level or strongly reduced number of photons per pulse (Fig. 81).

Even though the beam divergence of the infrared laser light is much smaller compared to the radio frequency antenna propagation, the effect cannot be neglected.

Obviously the more photons per pulse the optical telescope or antenna can receive, the better the FEC with efficient channel coding will be, the lower the BER will be and the more reliably the deep space optical communications will be able to perform over extremely long distance.

CHAPTER 13

NETWORK SECURITY

Fundamental aspects of network security will be discussed in this chapter. The techniques and the methods will guarantee these following five major network security functionalities.

Availability/Reliability: The data are to be available always. High availability servers – mirrored and equipped with redundant components and network connections – and protection switching mechanisms are employed.

Confidentiality: Third parties cannot access the network and the databases.

Integrity: Protection of the data systems against unauthorised changes.

Authentication: Login and user authorisation.

Liability, Non-Repudiability: Any access to the network and databases can be protocolled.

Especially the availability and performance are strongly related with the quality of service QoS. Quality of service is always dependent on different parameters. Based on the International Telecommunications Union ITU-T definitions of quality of service [106], the following technical parameters and aspects have to be considered in network planning and operations:

- Data throughput (speed) of the access network.
- Congestion in the backbone.
- End-to-end delay (latency).
- Delay-variation (jitter).
- Packet loss (loss of information).

Besides these, there are some very important ITU-R recommendations concerning detailed regulations about QoS and QoE (quality of experience) [107] [108], MOS (mean opinion score) [109] and mobile network IMT-2000 specific QoS issues [110].

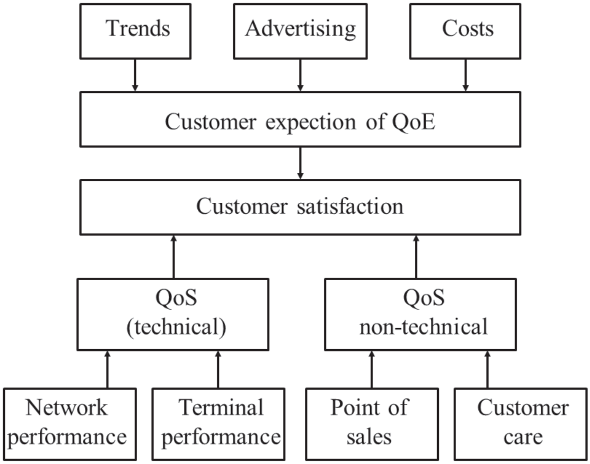


Figure 82 ITU-T QoS guidelines

As we can easily see, the ITU-T QoS definitions (Fig. 82) consider a lot of parameters concerning not only the physical and technical aspects like transmission and propagation characteristics of the information signals, but also operational aspects like resource and facility, maintenance support, reaction time and the troubleshooting process of the network operators.

In this textbook we would like first to focus on the technical parameters of the physical propagation and transmission. The major technical flow characteristics which directly have impact on the QoS, sometimes called QoE (quality of experience) from the user point of view, are (Fig. 83):

- Availability.
- Delay.
- Jitter.
- Bandwidth.
- Expenditures including investment and annual operational costs.

Availability is the fundamental and most critical parameter for QoS.

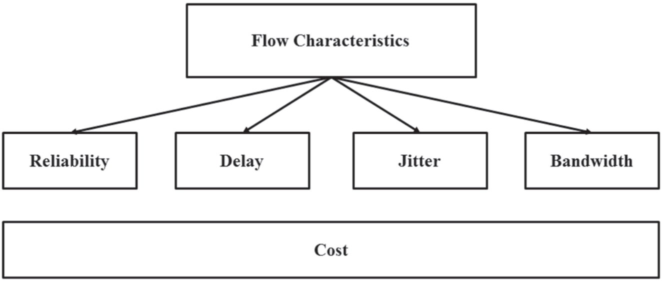


Figure 83 A subset of technical QoS parameters

Availability A is defined by the ratio of the $MTBF$ (Mean Time between Failure) to the sum of the $MTBF$ (Mean Time between Failure) and $MTTR$ (Mean Time to Repair). The whole operational time should be considered.

$$A = \frac{MTBF}{MTBF + MTTR} \qquad \text{with } A \le 1 \text{ or } 100\% \qquad (13.1)$$

whereas the non-availability N, sometimes also called UA (Unavailability), is defined as

$$N = 1 - A = \frac{MTTR}{MTBF + MTTR} \,. \qquad (13.2)$$

The availability A and non-availability N are statistical mean values, derived from the network operation reports. In practice, each errored second with bit errors higher than 0.1% is considered as a SES (severely errored second). Correspondingly this second is counted as outage for the network management report. In the network management reports all these outage events will be summed up together, leading to the total outage over a certain observation time period.

In practice, availability and reliability are treated as the same issue. The less the outage or the less the non-availability, the higher the reliability will be.

Therefore, the most important task for the network operator is to guarantee a reliable network operation by continuously analysing and monitoring the ICT infrastructure performance (routers and switches, data transmission, data storage, application server infrastructures, etc.), and defining measures and operational guidelines to ensure the high availability and low down-time and developing disaster recovery or backup concepts for the worst-case of an outage or severe performance degradations.

The responsible ICT managers or departments have to proactively develop long-term ICT strategies and guidelines for a network security policy. The major focus should be:

- Application availability: Optimisation of server and application infrastructure.
- Data availability: Continuous development of existing data storage and backup guidelines for preserving business critical data.
- Communications and networking: Ensure the data transport within LAN, WAN and SAN, and flexibly adapt the network to a traffic increase.
- Network Management System: Implementation of appropriate tools and methods for proactive surveillance and analysis of the ICT systems.

MTBF (Mean Time between Failures) is defined as

$$MTBF = \sum_{i=1}^{M} TBF_i \tag{13.3}$$

with M intervals and *MTTR* (Mean Time to Repair) is defined as

$$MTTR = \sum_{i=1}^{m} TTR_i \tag{13.4}$$

with m failures. For serial connection of n sub-systems ($i = 1, 2 \ldots n$) with the corresponding availability A_i for each sub-system i:

$$A_{total} = \prod_{i=1}^{n} A_i \ . \tag{13.5}$$

Parallel connection of n sub-systems ($i = 1, 2 \ldots n$) with the corresponding non-availability N_i for each sub-system i:

$$N_{total} = \prod_{i=1}^{n} N_i \ . \tag{13.6}$$

Example:

Time interval	TBF (hours)	TTR (hours)
1	350	5
2	400	3
3	267	4
4	380	2
5	450	3
6	550	5
7	453	8
8	650	6
9	780	3
10	894	4
sum	5174	43
average	517.4	4.3
	MTBF	MTTR

Table 15 Example of *MTBF* and *MTTR* calculation

For the assumed values for *TBF* (Time between Failures) and *TTR* (Time to Repair) in Table 15, the availability will be $A = 99.18\%$ and $N = 0.82\%$ correspondingly.

If, for example, N technical systems are connected in series, then the overall availability will be $A_{total} = A_1 \cdot A_2 \cdot A_3 \cdot ...A_i \cdot ...A_{N-1} \cdot A_N$. This means the overall availability will become smaller and smaller, since all the availabilities are numbers less than 1 or 100%. Or the total availability is always less than each single one.

If, however, N technical systems are connected in parallel, then the overall availability will be $N_{total} = N_1 \cdot N_2 \cdot N_3 \cdot ...N_i \cdot ...N_{N-1} \cdot N_N$. This means the overall non-availability will become less and less, and therefore the availability becomes better and better. These parallel connected technical systems are also called redundant systems, since the availability of each single system is always improved.

In daily life, the redundant UPS (uninterruptible power supply) system helps to recover the power supply during the outage of primary power supply from the utility company. In the case of parachute jumping, the redundant parachute helps to save life, if there are some problems with the primary parachute.

In IT and telecommunications we have a similar situation. The primary transmission line or path can be protected by the backup or redundant transmission line or path. This is valid both for optical fibre transmission networks, where, for example, SNCP (Sub-Network Connection Protection) or Ring Protection configuration or monitoring hot-standby MSP (Multiplex Section Protection) can help avoid service interruption in the case of a partial outage of some technical systems.

In order to achieve cost-effective network planning and optimisation, the backup system can be designed to take over the complete functions or partial functions of the primary system, depending on the importance of certain functions, and on the availability statistics of the involved network components. For example, in an office building, the redundant power supply or UPS (uninterruptible power supply) like batteries or diesel generators are normally designed to protect the most important and critical business applications as well as emergency lighting and ICT systems, but not the normal light systems in the floors.

Quality Grade	Availability	Outage per Year
Reliable	99%	87.6 hours
Available	99.9%	8.76 hours
Highly available	99.99%	52 minutes
Fault insensitive	99.999%	5 minutes
Fault tolerant	99.9999%	32 seconds
Disaster resistent	99.99999%	3 seconds

Table 16 Quality grades for availability and reliability

Different quality grades depending on the availability are summarised in Table 16. Sometimes, depending on the so-called SLA (Service Level Agreement) between the IT service provider and customers, for example in outsourcing, out-tasking or hosting projects, only the business time period is taken into account:

Availability = (Business Time Period - Outage) / Business Time Period.

Example 1: 24/7 Service, Outage < 3 Hours per Week
$A = (24 \cdot 7 - 3) / (24 \cdot 7) = 98.21\%$.

Example 2: 12 h Working Days, Outage < 3 Hours per Month
$A = (12 \cdot 5 \cdot 4 - 3) / (12 \cdot 5 \cdot 4) = 98.75\%$.

In network operations management reports, the bit error rates are also considered, in order to guarantee QoS. For example, each second with bit error is called an ES (errored second). Each second with the bit error rate higher than 10^{-3} is then called a severely errored second SES, and will be considered as outage time.

As we can clearly see, the bit error rate is strongly related to the quality of the voice service. In the beginning of voice telecommunications, the quality of voice was evaluated subjectively by interviewing the listeners and scored by marks ranging from 5 perfect – 1 bad (see Table 17). These values were called MOS (mean objective score).

MOS Value	Quality	Impairment
5	perfect	imperceivable
4	good	perceivable, but not annoying
3	Fair	lightly annoying
2	poor	annoying
1	Bad	very annoying

Table 17 MOS Mean Opinion Score

Performance can also be measured by the bit error rates for the real-time voice traffic. The relations are similarly evaluated and correlated to the bit error rates (see Table 18). In fact many measurement equipment suppliers use these or similar correlations to calculated the MOS nowadays also by using objective measurement data, instead of interviewing the listeners.

BER	Subjective Voice Quality description
10^{-6}	degradation is imperceivable
10^{-5}	single clicks, perceivable at lower signal level
10^{-4}	single clicks, annoying at lower signal level
10^{-3}	burst clicks, perceivable at each signal level
10^{-2}	strong crack noises, annoying the comprehensibility
$5*10^{-2}$	incomprehensibility

Table 18 Voice quality and bit error rates

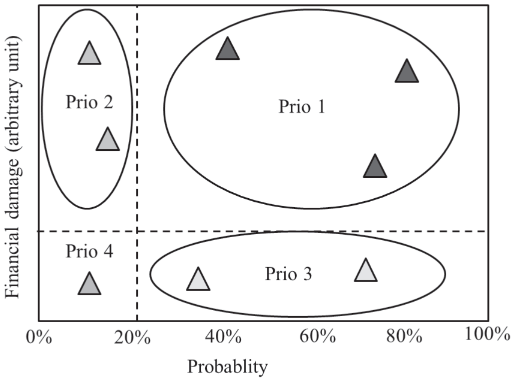


Figure 84 Risk analysis

Risk analysis in business auditing processes for all large or medium-sized public limited companies according to regulations like Basel II in Europe or the Sarbanes Oxley Act in the USA (Fig. 84) especially investigates the probability and damages in the case of failures or outages of certain business critical systems, especially ICT systems:

- Summary of all possible risks R_i (i = 1, 2 … n).
- Probability of the occurence P_i of the risk R_i.
- Outage time of the scenario i T_i (down time).
- Financial damage D_i.

Depending on the probability and possible financial damage for each of these risks, which have been analysed, corresponding counter-measures have to be defined and the problems have to be solved.

Delay is caused by propagation and transmission delay related to speed. For example, in free space transmission the delay is related to the velocity of light. Additionally, delay can be caused by the various digital signal processing steps, like source coding, channel coding, packetising, serial transmission in multiplex system, modulation, de-multiplexing, de-coding, as well as congestion in the queuing system or retransmission in the case of bit errors in TCP (Transmission Control Protocol).

If subsequently sent packets suffer different delays, which is the typical case in IP networks, then the delay variation is called jitter.

In the case of delay and jitter we have to distinguish the impact of these parameters to the rt (real-time)-services (for example, a telephone conversation between two persons or a video conference) and to the nrt (non-real-time)-services (for example, e-mail and file transfer).

rt services are also called “mission critical services”. Many nrt services are not sensitive to delay and jitter, but are sensitive to bit errors, for example the transfer of a large data file (Fig. 85).

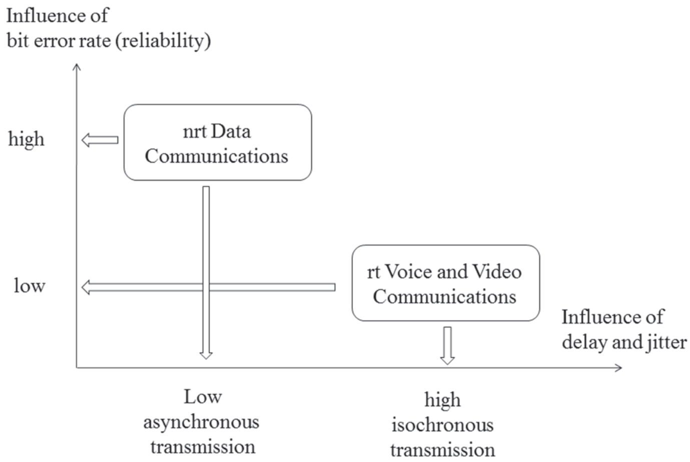


Figure 85 Sensitivity of rt and nrt services

QoS improvement

Considering how the above mentioned relevant technical parameters directly influence the network and the service quality, the ideal case is to achieve maximum bandwidth with minimum delay and jitter, which means immediately the best router and switching performance as well as the shortest possible transmission distance. In many cases, these conditions cannot be achieved completely at the same time.

In practice, because of the limited transmission capacity and continuously increasing traffic demands, even for the optical fibre transmission systems, but especially for the wireless transmission systems, where the free space transmission medium, the frequency spectrum, is not only limited, but also shared by many users and user terminals, quality of service is always limited.

Routers and switches always underlay some constraints because of the processor performance. Also the transmission path is not deterministic in the IP routing protocol, so that jitter or delay variation is not avoidable.

In order to compensate for jitter, a playout buffer is used to equalise the delay between the subsequent packets. The playout buffer size should be optimised in such a way, that most of the delayed packets can be stored for processing. On the other hand, it should not be too large, in order to avoid additional delay.

The next question is, which transport protocol should be used for the All-IP traffic for real-time services. TCP is reliable in terms of lower bit error rate by using error control and retransmission for lost packets, but not really suitable for real-time services, since the retransmission will cause additional delay. On the other hand, UDP has no time stamp and no sequence number, so the packet order cannot be reorganised by the playout buffer and the packet loss and jitter cannot be detected.

Therefore for the real-time multimedia services, we need a new extension of the transport protocol, which is called RTP (Real-time Transport Protocol; Fig. 86 and Table 19 – Table 20), in order to support the following functions:

- Timestamps (indicating delay).
- Sequence number (guaranteeing the right order to be played out).
- Playback buffer (storing data for playout with a threshold value).
- Multicasting (video conferencing).
- Translation, adaptation of bandwidth of transmitter and receiver.
- Mixing of different sending sources to one stream.
- Support from transport layer protocol [111] [112].

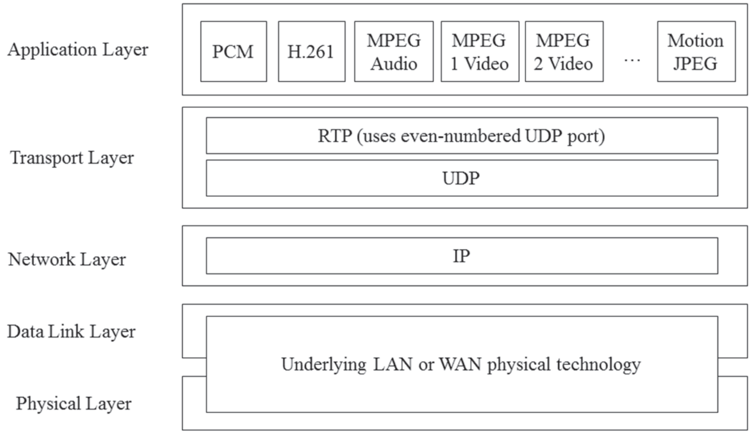


Figure 86 Real-time Transport Protocol

No.	Codec	Audio/Video	Sampling /s
0	PCMμ	A	8000
3	GSM	A	8000
4	G723	A	8000
5	DVI4	A	8000
6	DVI4	A	16000
7	LPC	A	8000
8	PCMA	A	8000
9	G722	A	8000
10	L16	A	44100
11	L16	A	44100
12	QCELP	A	8000
13	CN	A	8000
14	MPA	A	90000
15	G728	A	8000
16	DVI4	A	11025
17	DVI4	A	22050
18	G729	A	8000
25	CelB	V	90000
26	JPEG	V	90000
28	nv	V	90000
31	H261	V	90000
32	MPV	V	90000
33	MP2T	AV	90000
34	H263	V	90000

Table 19 RTP payload type

<table>
<tr><td colspan="4">Byte 0</td><td colspan="2">Byte 1</td><td>Byte 2</td><td>Byte 3</td></tr>
<tr><td>V</td><td>P</td><td>X</td><td>CC</td><td>M</td><td>PT</td><td colspan="2">Sequence Number</td></tr>
<tr><td colspan="8">Timestamp (in sample rate units)</td></tr>
<tr><td colspan="8">Synchronisation Source Reference Code (SSRC) or identifier</td></tr>
<tr><td colspan="8">Contributing Source Reference Code (CSRC) or identifier (optional)</td></tr>
<tr><td colspan="8">Header Extension (optional)</td></tr>
</table>

Table 20 RTP header structure

- V: Version
- P: Padding bit
- X: Extension
- CC: CSRC count, or number of the CSRC identifiers
- M: Marker
- PT: Payload Type
- Timestamp: very important for real-time applications. The time point of the generation of the payload
- Sequence number: order of the packets, detection of packet loss
- Synchronisation Source Identifier SSRC: identification of the source of the real-time application
- Contributing Source Identifier CSRC: optional, for the case the payload is not coming from the original source / sender

The corresponding RTCP (Real-time Transport Control Protocol) is responsible for the management of the RTP messages (Table 21).

200	Sender Report	Transmission and reception statistics for all RTP packets. Absolute timestamp to synchronise different RTP messages (audio and video have different timestamps).
201	Receiver Report	Passive reception. The receiver report informs about the quality of service.
202	Source Description Message	Information about the source: name, e-mail addresses, phone number, address of the owner or controller of the source.
203	Bye Message	To shut down a stream. Source announces that it leaves the session or conference.
204	Application Specific Message	New application not yet defined in the standard. Possible new message.

Table 21 RTCP messages

SIP (Session Initiation Protocol)

The similar application layer protocol SIP (with the typical messages in Fig. 87) was designed to establish, manage and terminate a multimedia session or a call. It can create two-party, multi-party or multicast sessions, independently of the underlying transport layers (UDP, TCP or SCTP).

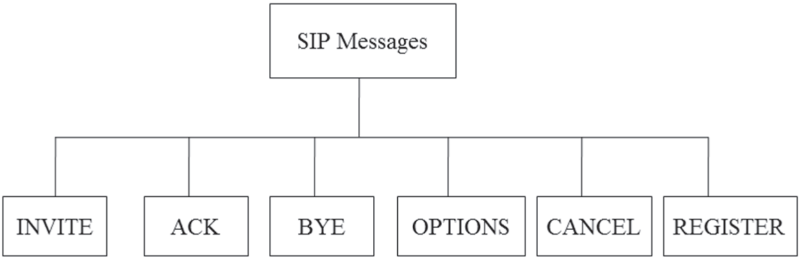


Figure 87 SIP messages

Scheduling

- FIFO (first in first out) queuing does not prioritise the incoming traffics, so that everyone has to wait, so long so the channel is occupied.
- Priority queuing: only the highest priority traffic will be scheduled first, and low priority traffic has to wait.
- Weighted fair queuing: this method achieves a fair handling of all priorities by assigning certain number of the available channels to certain priorities.

Traffic Shaping

- Leaky bucket: once one user is assigned to the available channel, the traffic will be transmitted until completion. During this time period, nobody else can transmit data.
- Token bucket: in order to provide a fair method to shape the traffic for a channel, a kind of token, which is the accumulated waiting time of a certain user, is used to guarantee that each one has a fair chance also to transmit the data after waiting for a long time. During the transmission the number of the token will be reduced correspondingly. When the token is reduced to zero, then the user has to wait again.
- Combination of leaky and token bucket method can provide certain services, with partially prioritised premium services.

In order to achieve the optimum, a trade-off decision is done, and the weighted fair queueing and a fair scheduling scheme are used to provide high-priority services sufficient network resources, and at the same time to allow the low priority real-time services to transmit the low bit rate data.

Integrated Services

Integrated services (IntServ) was introduced to provide a quality of service for internet traffic, since the internet normally neither guarantees a limited delay and jitter, which is fundamentally important for real-time services, nor guarantees the successful delivery of the data packets in the case of UDP. The principle is similar to the classical circuit-switched telephone network, where the call wish is signalled to the switching centre and a dedicated channel or circuit is reserved for the transmission, set up and maintained during the whole transmission. This process is done by:

- Signalling.
- Flow Specification.
- Admission.
- Service classes (guaranteed service, controlled-load service).
- RSVP (resource reservation protocol).

The disadvantage is a problem of scalability, i.e. if the involved network areas become too large, the number of internet routers, probably of different vendors with possibly proprietary functions, also become large and sometimes cannot perform the resource reservation end-to-end, so that the connection setup will fail.

Differentiated Services

In order to solve the scalability problem of the IntServ, DiffServ (differentiated services) was proposed. DiffServ does not aim at the end-to-end connection channel reservation, but rather focuses on the differentiated service handling by defining the service classes per hop from router to router, and by providing different services with specific QoS requirements. The non-real-time services (like e-mails and data files) do not need a permanently available channel, but rather reliable or bit error free delivery of the data packets, whereas the real-time services (like telephony or video-conferencing) need low delay and low jitters, but can tolerate some level of bit error rates. By defining different service classes, the services can be handled more appropriately:

- Differentiated Services Model is class-based.
- Per hop-behaviour.
- DE PHB (Default PHB) best-effort.
- EF PHB (Expedited forwarding PHB): low loss, low latency, ensured bandwidth.
- AF PHB (Assured Forwarding PHB): as long as the class traffic does not exceed the traffic profile.

Besides the above-mentioned service classes, other classes can be defined by the so-called DSCP (differentiated service code point). In comparison to IntServ, the differentiation does not need to be unique between the system vendors. Once the classes are defined, the traffic conditioner will check the data flows like the following:

- Meter: checks if the incoming flow matches the negotiated traffic profile.
- Marker: marks a packet using best-effort or down-marks a packet.
- Shaper: reshapes the traffic if it is not compliant with the negotiated profile.
- Dropper: works as a shaper without buffer and discards the packets only.

13.1 Symmetrical Encryption

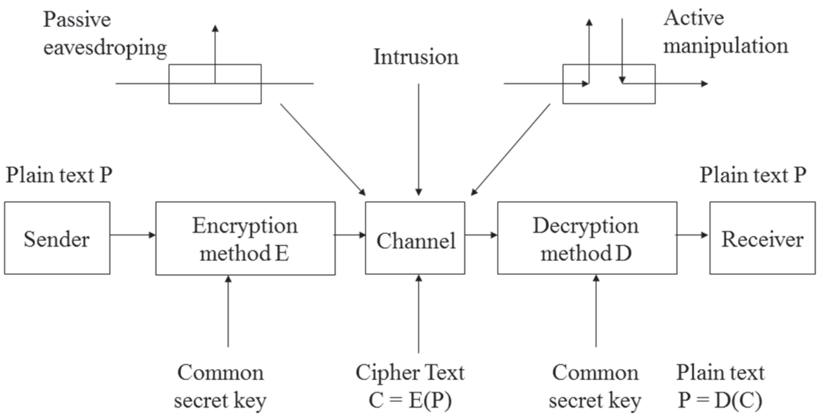


Figure 88 Symmetrical encryption

Cryptography is the science to conceal the secret information to ensure confidentiality, authenticity, integrity and non-repudiability. Modern cryptographic methods are basically symmetrical (Fig. 88) and asymmetrical encryptions [5] [113].

Classical symmetrical Caesar encryption
Used by the Roman leader Caesar as one of first encryption techniques to enable secure communications, the encryption is done by rotating the alphabet, and therefore converting the plain text to a cipher text (Table 22). For example:

Plain	A	B	C	D	E	……	U	V	W	X	Y	Z
Cipher	E	F	G	H	I	……	Y	Z	A	B	C	D

Table 22 Caesar encryption key to get the cipher text

By doing so, 25 keys are possible by shifting the alphabet. One of the famous examples is the key 13 which is used in Unix-Systems: rot13. rot13 has no difference between encryption and decryption.

The Caesar cipher is easy to crack by analysing the statistical features of the alphabet in each language. For example, in English or German the most often used letter ‘e’ appears 17.5% and ‘n’ 9.8% in a sufficiently long text. Also certain combinations of letters like “en”, “er”, etc. are often used. In French, Spanish, etc., different alphabet letters are often differently used. These regular occurence probabilities lead to relatively easy decryption of the cipher text.

Modern symmetrical encryption
Modern symmetrical encryption techniques normally use binary boolean logical operations “and”, “or” and “xor” (exclusive or). In Table 23 different nomenclatures of boolean logic operators are listed:

Operator	op Symbol	0 op 0	0 op 1	1 op 0	1 op 1
and	&	0	0	0	1
or	|	0	1	1	1
xor	⊕	0	1	1	0

Operator	op Symbol	0 op 0	0 op 1	1 op 0	1 op 1
and	.	0	0	0	1
or	+	0	1	1	1
xor	⊕	0	1	1	0

Operator	op Symbol	0 op 0	0 op 1	1 op 0	1 op 1
and	$\wedge$	0	0	0	1
or	$\vee$	0	1	1	1
xor	$\oplus$	0	1	1	0

Table 23 Different boolean logics symbols used in the encryption

DES (Data Encryption Standard), Block Cipher

One of the first modern data encryption standards was proposed by Horst Feistel of the IBM research centre [114]. Figure 89 shows the schematic 16 steps of the Feistel network.

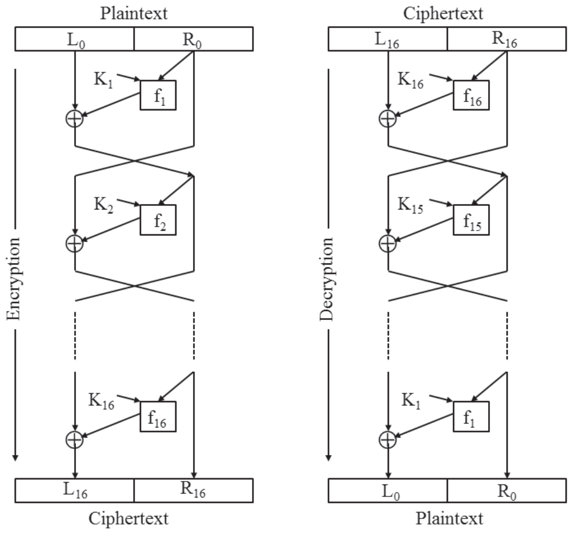


Figure 89 DES Data Encryption Standard with Feistel networks

DES divides the original text ($L_0 + R_0$) to 64 bit blocks, which in turn are split into left hand part L_i and right hand part R_i. For each step the right side is encrypted by using the corresponding step key K_i (64 Bits – 8 Parity bits = 56 Bits). Each step is called a modified Feistel network for encryption by using a special logical algebraic function $f_i(R_{i-1}, K_i)$:

$$L_i = R_{i-1}. \qquad (13.7)$$

$$R_i = L_{i-1} \oplus f_i(R_{i-1}, K_i). \qquad (13.8)$$

Feistel network repeats this function in 16 steps iteratively. After n steps the pair L_n und R_n will then be combined to the cipher text. This whole encryption process is abbreviated as E.

Decryption is carried out similarly, but in a reverse way, and abbreviated by D:

$$R_{i-1} = L_I. \tag{13.9}$$
$$L_{i-1} = R_i \oplus f_i(R_i, K_i). \tag{13.10}$$

Since the DES has effectively only a 56 bit step key, the probability is high that by using powerful computer and appropriate algorithms, the key can be calculated, which is called brute force attack of the hackers.

3DES or Triple DES, Block Cipher

In order to increase security, and in order to make the brute force approach more difficult, 3DES or triple DES was proposed to increase the effective key length by cascading the encryption and decryption algorithms

For encryption:	E D E
For decryption:	D E D

AES (Advanced Encryption Standard), Block Cipher

Following on from the DES and triple DES, the National Institute of Standards and Technology in the USA published the new standards AES, proposed by Joan Daemen und Vincent Rijmen [115]. AES uses a block length of 128 bit and chooses 128, 192 or 256 bit as key lengths, which are called correspondingly AES-128, AES-192 and AES-256. The algorithms are freely available, do not need licenses and provide a high level of security.

All above-mentioned symmetrical cipher standards are called block cipher, since the payload data are processed in blocks with the session keys of the corresponding lengths.

Stream Cipher

A stream cipher is a symmetrical key cipher, where plaintext bits are combined with a pseudo random cipher bit stream (key stream), typically also by using an exclusive or (XOR) operation. One of the most famous stream ciphers is A5/1, used in the GSM mobile networks (see, for example, [116], Fig. 97).

The stream cipher symmetrical key for A5/1 is generated by using three left feedback shift registers, with different register lengths, which correspond to the so-called characteristic polynomials. Some of the register bits are tapped, XORed and fed back to the corresponding register. When the registers are switched on, the register bits are zeros. After a certain operating time period, or iterative feedbacks, the bit sequences become more and more pseudo random, so that these pseudo random bit streams can then be used as stream cipher keys, to be continuously XORed with the payload data stream.

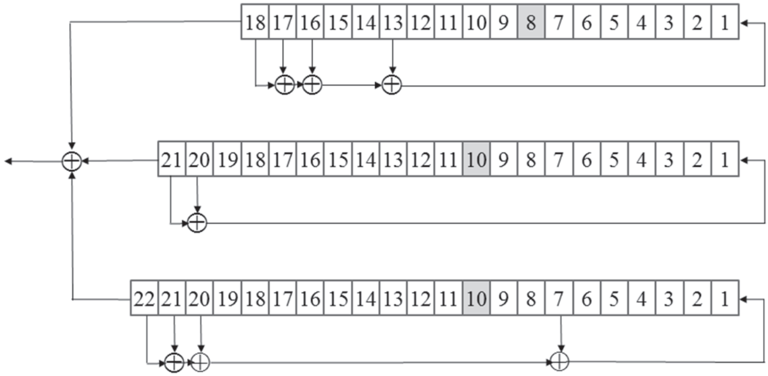


Figure 90 Stream cipher

LFSR	Length in bits	Characteristic Polynomial	Clocking bit	Tapped bits
1	19	$x^{18} + x^{17} + x^{16} + x^{13} + 1$	8	13, 16, 17, 18
2	22	$x^{21} + x^{20} + 1$	10	20, 21
3	23	$x^{22} + x^{21} + x^{20} + x^{7} + 1$	10	7, 20, 21, 22

Table 24 Stream cipher characteristic polynomials

Figure 90 and Table 24 show a stream cipher with the three left feedback shift registers LFSR with the characteristic polynomials, tapped bits and the clocking bits 8, 10 and 10. In a stream cipher the plaintext digits are encrypted one at a time, and the transformation of successive digits varies during the encryption. An alternative name of the stream cipher is therefore a state cipher, because the encryption of each digit is dependent on the current state of the pseudo random bit stream in the registers. In practice, the digits are typically single bits or bytes.

In 2G GSM mobile communications, the data transmission is organised as sequences of bursts. In a typical channel and in one direction, one burst is sent every 4.615 milliseconds and contains 114 bits available for information both for upstream and downstream. The algorithm A5/1 is used to produce for each burst a 114-bit sequence of keystream which is XORed with the 114 bits payload data prior to modulation.

A5/1 is initialised using a 64-bit cipher key (Kc) together with a publicly-known 22-bit frame number. In GSM implementations in the field 10 of the key bits are fixed at zero, resulting in an effective key length of 54 bits.

Besides the DES, 3DES and A5/1, other famous encryption techniques are:

Blowfish: Feistel network with 16 steps. Key length 32 bit – 448 bit.

Twofish: Further development of Blowfish. Key length 256 bit. Feistel network with 16 steps.

CAST: Symmetrical block cipher with a key length of 256 bit.

IDEA (International Data Encryption Algorithm): block cipher with a key length 128 bit.

RC2 (Block 64 bit), RC5 (32, 64, 128) and RC6 (128/128, 192, 256). RC4 (stream cipher, key length 2048 bit, each byte instead of block encoded, secure and quick. HTTPS, SSH, SecurePC and WEP).

AES (block cipher 128 bit, 192 bit, 256 bit with key lengths 128, 192 and 256 bit). Further development of DES and 3DES.

Problems of the symmetrical encryption:

Despite of the different methods for increased key lengths and increased security, one problem remains. Since the sender and receiver use the same key for encryption and decryption, this key should be agreed between both parties in advance, and kept secret. So the key exchange becomes very critical.

This problem can be handled better by using the asymmetrical encryption methods.

13.2 Asymmetrical Encryption

In contrast to symmetrical encryption, the asymmetrical encryption method (Fig. 91) uses different keys for encryption and decryption ("Public Private Key" method – often used for key exchange): The public key for encryption is known and can be exchanged through the open unencrypted internet, whereas the private key is used for decryption at the receiver site. This private key is secret [5] [113].

A cipher text encrypted by a public key can only be decrypted by the corresponding private key, but not by a public key. Some asymmetrical encryption methods enable the decryption of a cipher text by using a public key, when the cipher text is encrypted by a private key. A famous example is the digital signature or hash value like the A3 signed response SRES. The most famous asymmetrical encryption methods are RSA, Diffie-Hellmann and ElGamal.

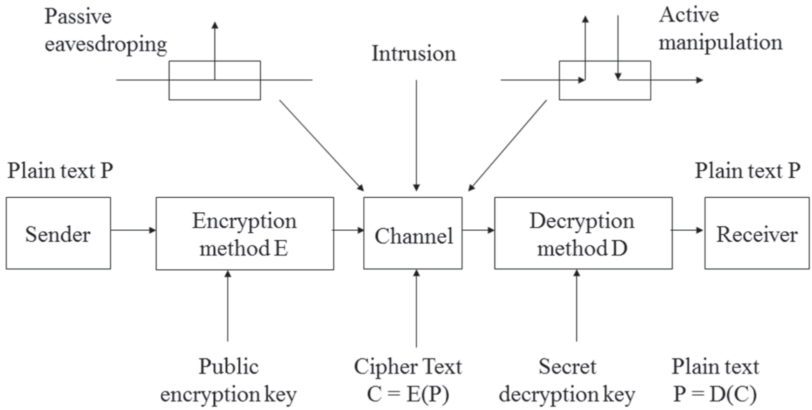


Figure 91 Asymmetrical encryption

Most asymmetrical encryption methods utilise the mathematical number theory for prime number calculations – which are one of the most challenging and time-consuming mathematical problems – and modulus operations, in order to produce some keys which cannot be simply calculated by brute force methods, since the relatively large prime number calculation needs high computation power and normally takes a relatively long time.

This is somehow also the disadvantage of the asymmetrical encryption techniques. On one hand, the asymmetrical encryption technique is relatively secure, but on the other hand, encryption and decryption take significantly more time than symmetrical encryption methods.

RSA (Rivest, Shamir, Adleman)

The plain text P is encrypted by a public key (e, N):

$$C = P^e \bmod N. \tag{13.11}$$

The cipher text C is decrypted by a private key (d, N):

$$P = C^d \bmod N. \tag{13.12}$$

The procedure for selection of the public and private keys:

Choose two prime numbers p and q.

Calculate $N = p \cdot q$ and $z = (p - 1) \cdot (q - 1)$.

Choose a number d which is relatively prime to z (no same divider except for1).

Find a number e, such that $e \cdot d = 1 \bmod z$.

Example by using [4]**:**

$p = 197$, $q = 251$

$N = 3 \cdot 11 = 49447$ and $z = 49000$

A suitable number is $d = 13473$

$e \cdot d = 7 \cdot e = 1 \bmod 49000$ can be fulfilled with $e = 2^{16}+1$.

Diffie-Hellmann Method

Two communications partners agree upon a common key in such a way that each partner creates half of the private key and encrypts and sends this half to the counterpart. The other partner can decrypt the missing half and get the whole key.

Unfortunately the Diffie-Hellman Protocol is vulnerable for the man-in-the-middle attacks, if the exchanged messages can be manipulated in such a way that the communications partners can not verify whether the message is coming really from the partner or not. Therefore this method is not an ideal method to be used for encryption of the cipher text, but can enable a relatively efficient and secure key exchange.

The procedure to assign and exchange the public and private keys:

Be $x = c^a \bmod n$, then a cannot be calculated by only knowing x and c. Communication partners A and B agree upon modulus p and basis c, whereas p is a prime number and c is chosen such that

$$c^n = 1 \bmod p \text{ for } 0 < n < p - 1.$$

A chooses exponent a and calculates $x = c^a \bmod p$. (13.13)

B chooses exponent b and calculates $y = c^b \bmod p$. (13.14)

A and B send x and y to the partner respectively (from x and y the man-in-the-middle cannot calculate a and b):

A calculates $y^a = (c^b)^a \bmod p = e$.
B calculates $x^b = (c^a)^b \bmod p = e$.

e is then the common key of A and B, consisting of two parts (x, a) and (y, b), and can be used for the payload data encryption. That is the reason why the Diffie-Hellmann-Method is used for the key exchange prior to the payload encryption and transmission. Normally the symmetrical session encryption key should be renewed regularly, say at least every 24 hours, in order to make it difficult for a hacker to calculate the encryption key.

13.3 Authentication

The algorithm A8 calculates the cipher key Kc based on the secret individual subscriber authentication key Ki (128 bit) and a random number RAND (128 bit) in the GSM mobile network. Kc is used for the encryption algorithm A5 for the air interface between the base station and mobile equipment. A8 is implemented both in the authentication centre (AuC) and in the SIM card of the subscriber. The algorithm is standardised in GSM and can be implemented by the network operators individually.

13.4 Hash-value, integrity check

A3 is the algorithm for the authentication of the mobile terminals. The algorithm calculates the so-called signed response SRES (32 bit) based on the secret individual subscriber authentication key Ki (128 bit) and a random number RAND (128 bit). A3 is implemented both in the authentication centre (AuC) and in the SIM card of the subscriber.

Since the algorithms A3 and A8 use the same input parameters for the key generation procedures, these two algorithms are implemented together often and are known as A3/A8.

CHAPTER 14

NETWORK MANAGEMENT

TMN (Telecommunications Management Network) and SNMP (Simple Network Management Protocol) are the most-used network management systems in information and communication technologies. TMN was first introduced for the management of ISDN, ATM, GSM/UMTS networks, and then extended to the modern ALL-IP networks. SNMP was first introduced for the internet, became more and more important, and is now also used for other new communication technologies.

14.1 Telecommunications Management Network

TMN is standardised by the International Telecommunications Union ITU in the ITU-T M.3000 Recommendations Series (TMN) (Fig. 92) for telephony applications [117] [118].

Business Management functions related to business aspects, analyses trends and quality issues, and provides a basis for billing and other financial reports.
Service Management functions for the handling of services in the network: definition, administration and charging of services.
Network Management functions for distribution of network resources, i.e. configuration, control and supervision of the network.
Element Management functions for the handling of individual network elements (alarm management, handling of information, back-up, logging and maintenance of hardware and software).

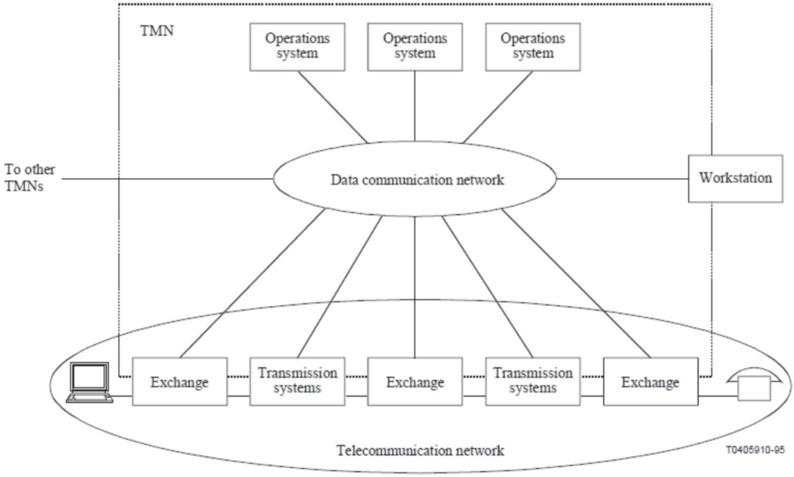


Figure 92 ITU-T TMN [118]

14.2 Simple Network Management Protocol

SNMP is standardised by the IETF for the internetworking between the computer networks.

The Structure of Management Information **SMI** describes the rules for the definition of Managed Objects.

Managed Objects (**MO**) describe certain features of the network resources or network elements NE.

An **Agent** (a piece of software installed on the NE) administrates the managed objects of its network resources.

The Management Information Base **MIB** contains all MO, and is a distributed virtual data base.

The SNMP provides all relevant information of the MO (variables) and defines the formats of the status, performance and fault messages (values).

Manager (Network Management Station NMS or Operation & Maintenance Centre OMC) and Agent (network element with a MIB database) are together a client-server system. The manager asks about the status, the agent gives detailed reports (Fig. 93).

An agent administrates the management objects of the resource contents of the management objects within the internet by using the hierarchical **MIT** (management information tree) structure:

Unique name (identifier):
iso.org.dod.internet.mgmt.mib-2.system.sysDescr (1.3.6.1.2.1.1…)

Syntax, types:	integer, string, array
Access rights:	read-only, read-write
Status:	mandatory, optional
Encoding method:	tag (type = class, format, number), length, value

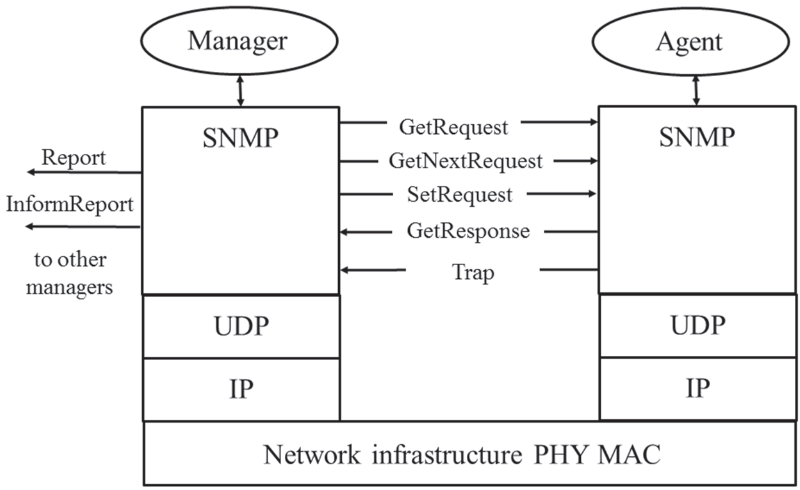


Figure 93 Interaction between Manager and Agent

Management Information Tree

Each management object has a unique position in the MIT (Fig. 94), and therefore has a unique relationship.

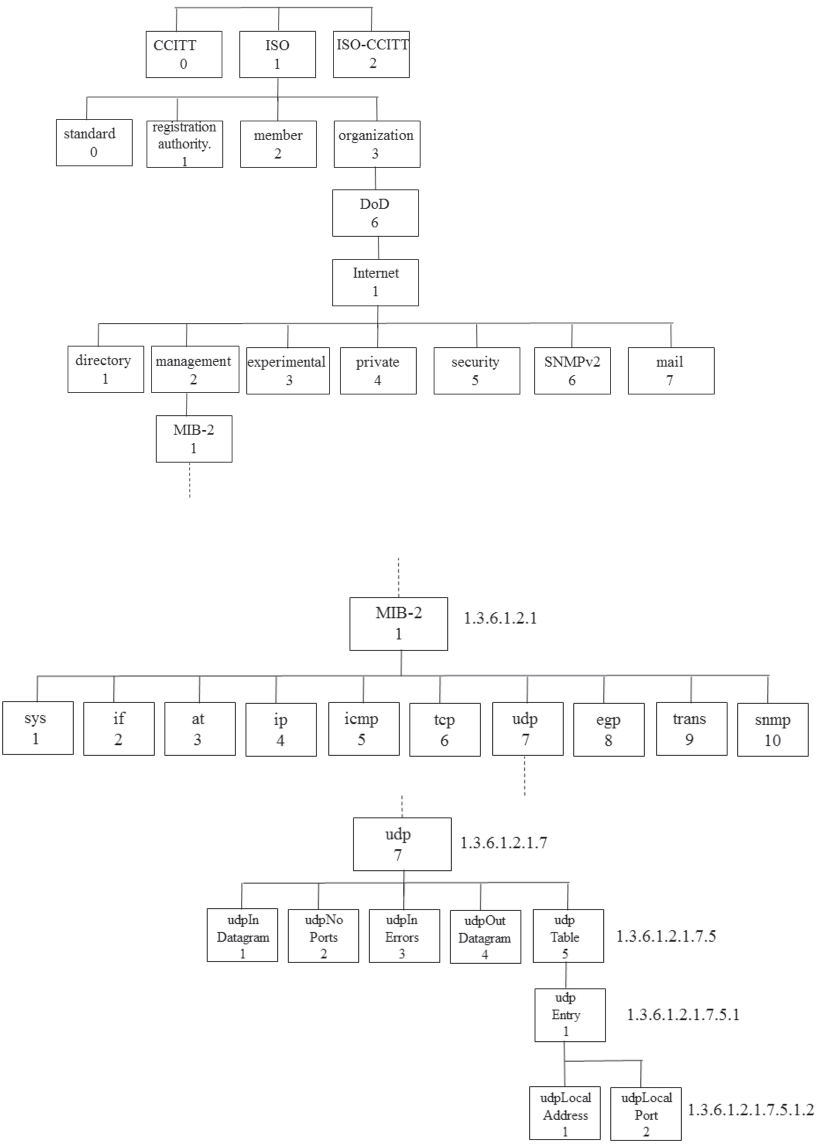


Figure 94 MIT Management Information Tree

14.3. Functionalities of Network Management

The five most important functionalities of network management are:

Configuration Management:
Definition of management objects (MO) and network element (NE) features, configuration and reconfiguration of network components (NE, connections, trunks and circuits), accounts, SW-updates, upgrades (new features), extensions and asset and resources documentation.

Fault Management:
Reactive and proactive fault management includes recognition, localisation of failures, troubleshooting and fixing of faults, recovery, restarts, swap and repairs, regular maintenance, updates and software bug fixing.

Performance Management:
Determination of network parameters in terms of network quality (QoS, GoS), traffic measurement, congestion, availability, BER, throughput, delay, response time, successful MOC (Mobile Originating Calls), MTC (Mobile Terminating Calls) and HO (different handovers)

Accounting Management:
Determination of the usage of the network resources for accounting purposes (billing).

Security Management:
Protect the management information and network resources, key and encryption, data security, and stop fraud and attempted misuses.

The main tasks of the NMS (Network Management System) or OMC (Operation Maintenance Centre) of a network operator are the monitoring and troubleshooting of different network subsystems (Fig. 95). Normally an operator, for example a mobile network operator, has many vendors in the core network, radio access network (GSM, LTE, 5G) and backbone and backhaul transport network (optical fibre network, radio relay systems, leased lines, cross-connects, multiplexer, etc.).

All these subsystems have normally their own network management systems, the so-called OSS (Operation Support Systems), so that all these vendor specific network systems must be monitored in parallel. The ideal case would be an integrated network management system.

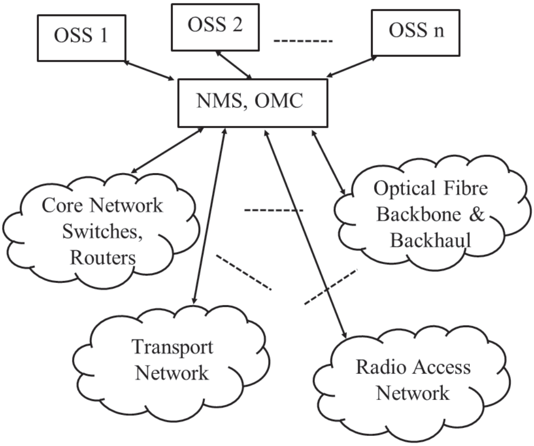


Figure 95 Main tasks of network management

Typical problems in the usually complex, multi-vendor networks should be found immediately and protection switching could be used to avoid any interruption or outage. The best solution is to permanently monitor the whole network, and solve the problems like performance degradation in advance in order to avoid any disturbances of the Quality of Service QoS requirements.

One reason for outage is failure of network elements, subsystems, systems or physical transmission media like fibre. It could be a total outage or a technical defect of some critical network subsystems or network elements.

The reason for partial outage, disturbance of functions or temporary performance degradation is malfunction of network elements, subsystems, systems and physical media.

The reason for network congestion is often overload of network elements, subsystems or systems due to wrong dimensioning or unexpected traffic increase.

Sometimes the performance degradation is because of incorrect configuration of network elements, subsystems and systems. The reason is often incorrect forecast or oversights and lacking experience of the network configuration engineers.

One of the critical disturbances is the DoS (Denial of Service – by bombing or blocking the server).

CHAPTER 15

TRENDS AND FUTURE DEVELOPMENTS

Improved Mobile and Wireless Technologies

In order to further improve the spectrum efficiency of mobile and wireless technologies, beam forming and multi-user multiple-input multiple-output (MU-MIMO) antenna arrays will be applied for the cellular mobile networks 5G and 6G.

Multiple radio access technologies (M-RAT) enable the optimised choice of available radio access technologies (2G GSM, 4G LTE, 5G, WLAN, mm Waves, etc.) to meet specific QoS requirements in combination with network slicing.

Software defined radio (SDR) and network function virtualisation (NFV) allow the flexible configuration of networks, for example to better match the differentiated QoS requirements of the provided applications by using network slicing for TSN (time sensitive networks), URLLC (ultra-reliable low latency networks), eMBB (enhanced mobile broad band) services, a large number of distributed NB-IoT (narrow-band internet of things) and sensors etc. Mission critical applications like autonomous driving and robot control require low latency. Real-time gaming requires high bit rate transmission with low latency.

Dynamic spectrum allocation and new medium access technologies will improve the spectrum efficiency in combination with sophisticated coding and modulation schemes. Spectrum-on-demand results in higher overall capacity.

Core network convergence, enabled by the IP-based quality of service, will merge fixed and mobile technologies together. Besides the core routers, more and more edge routers will be used in the mobile network, in order to reduce the end-to-end latency.

Future 6G Mobile Technologies

6G mobile networks target per user throughput up to 1 Tbps and latency of 0.1 ms. The visions, features and potential technologies of the future 6G are illustrated, for example, in [119]:

- Multi-band ultrafast speed transmission by using mm Waves, THz (0.1 – 10 THz) and infrared (free space optical communications).
- Super flexible integrated networks in a combination of heterogeneous networks, space-terrestrial integrated networks and flying base stations.
- Multi-mode, multi-domain joint transmission with multimode ultra massive MIMO, OAM-MDM (orbital angular momentum multiplexing [120] and multi-domain index modulation [121]).
- Intelligent transmission by involving machine learning, big data and other multi-disciplinary techniques.

6G means also challenges to issues like power supply and power efficiency of the systems, well-integrated multi-level security, and improved hardware design by using optoelectronic integration leading to smaller components and antennas.

Fibre to the Home FTTH

The existing xDSL will be continuously replaced by optical fibres in the next decades, in order to meet the explosively increasing traffic demands caused by modern broadband mobile and fixed network internet services.

Fibre to the Base Stations

Due to the tremendously increasing capacities of the 5G and the future 6G network, optical fibres must also be installed directly to the mobile base stations in the backhaul network, which means the final integration or migration of wireless and fixed network technologies in the OSI physical layer.

Mission Critical Applications and other Novel Applications

Many projects focus on the development of mission critical applications like autonomous driving, tele-surgery via ultra-reliable broadband internet, robot control, etc.

These and other new applications like IoT, mMBB, URLLCs, smart homes, smart factories and AR/VR (artificial and virtual reality) will definitely push the ICT technology development.

Free Space and Deep Space Optical Communications

By using basically the same key components of the optical fibre communication technologies (laser diode as transmitter, photodiode as receiver), the data can be transmitted over free space. It is expected that the efficiency in comparison with the classical radio frequency transmission can be further enhanced.

For this reason, the world-leading space agencies like NASA and ESA investigate DSOC (Deep Space Optical Communications) to reduce the size of the space craft communication systems and increase the spectrum efficiency and achievable bit rates.

Coherent Optical Communications

To further improve the transmission bit rates of the sophisticated optical fibres communication technologies, mostly based on DWDM (Dense Wavelength Division Multiplex) on the optical fibres that has been successfully deployed over the last decades, more and more coherent optical communication technologies will be used, which not only apply higher-order modulation schemes up to 256QAM or even higher, but also use 2 polarisations in optical fibres, and up to 100 optical channels at different wavelengths, depending on the transmission distance and link budget. This enables transmission of up to tens of Tbps per fibre pair.

Artificial Intelligence, Machine Learning, Big Data

More and more research activities are going on with the applications of artificial intelligence, machine learning and big data in image recognition, in terms of traffic control, surveillance of sensitive areas and autonomous driving.

Power Efficiency and Extremely Large Scale Integration

Last, but not least, global warming or climate change strongly demand for power efficient technologies in all information and communication fields, partially enabled by the extremely compact size of all systems and devices by using ultra-high large-scale integration, in order to reduce the electrical power consumption as much as possible, and increase the system performance significantly.

REFERENCES

[1] C. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal,* 1948.

[2] H.-P. Bauer, Digital Technology (Logic Design). Lecture Notes. Darmstadt University of Applied Sciences, 2018.

[3] B. Sklar, Digital Communications, New Jersey: Prentice Hall PTR, 2001.

[4] CrypTool, [Online]. Available: https://www.cryptool.org/de/.

[5] A. Tannenbaum, Computer Networks, Prentice-Hall Inc. , 1989.

[6] H. Taub and D. L. Schilling, Principles of Communication Systems, McGraw-Hill Inc., 1986.

[7] P. A. Henning, Multimedia (in German), Leipzig: Fachverlag, 2003.

[8] H.-G. Unger, Planar Optical Waveguides and Fibres, Oxford: Clarendon Press, 1977.

[9] R. E. Collin and F. J. Zucker, Antenna Theory, Part 1, McGraw- Hill Book Company, 1969.

[10] M. Loch, Optical Communications, Lecture Notes, Darmstadt University of Applied Sciences, 2017.

[11] K. W. Kark, Antenna and Radiation Fields (in German: Antennen und Strahlungsfelder), Springer Vieweg, 2018.

[12] Y. Okumura, E. Ohmori, T. Kawano and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Review of the Electrical Communication Laboratory, 16 (9-10),* pp. 825-73, September-October 1968.

[13] M. Hata, "Emprirical Formula for Propagation Loss in Land Mobile Radio Services," *IEEE Transactions on Vehicular Technology, VT-29 (3),* pp. 317-25, August 1980.

[14] Wikipedia, https://en.wikipedia.org/wiki/Hata_model.

[15] Wikipedia, Hata Model, https://en.wikipedia.org/wiki/Hata_Model.

[16] S. Singh, "Matlab Program for Okumura Hata Model," [Online]. Available: http://www.mathworks.com.

[17] COST Hata Model, [Online]. Available: http://www.cost.eu/domains_actions/ict/Actions/231 COST (European Cooperation in Science and Technology).

[18] "Young Wave Propagation Model," [Online]. Available: https://en.wikipedia.org/wiki/Young_model.

[19] T. L. Marzetta, E. Larsson, H. Yang and H. Q. Ngo, Fundamentals of Massive MIMO, Cambrigde University Press , November 2016.

[20] NOKIA Bell Labs, "Future Cell Project," [Online]. Available: https://www.nokia.com/about-us/news/releases/2016/10/03/f-cell-technology-from-nokia-bell-labs-revolutionizes-small-cell-deployment-by-cutting-wires-costs-and-time/.

[21] K. Lee, Principles of Antenna Theory, John Wiley & Sons Ltd. , 1984.

[22] B. D. Steinberg and H. M. Subbaram, Microwave Imaging Techniques, John Wiley & Sons, Inc., 1991.

[23] R. J. Mailloux, Phased Array Antenna Handbook, Artech House Inc. , 1994.

[24] L. V. Blake and M. W. Long, Antennas: Fundamentals, Design, Measurement. 3rd Edition, Scitech Publishing, Inc., 2009.

[25] D. G. Fang, Antenna Theory and Microstrip Antennas, CRC Press, 2010.

[26] W. H. Carter, "On Refocusing a Radio Telescope to Image Sources in the Near Field of the Antenna Array," *IEEE Trans. on Antennas and Propagation, vol. 37,* pp. 314-319, 1999.

[27] A. Badawi, A. Sebak and L. Shafai, "Array Near Field Focusing," in *WESCNEX'97 Proceedings of Conference on Communications, Power and Computing, pp. 242-245*, 1997.

[28] M. Bogosanovic and A. G. Williamsoni, "Antenna Array with Beam Focused in Near Field Zone," *Electronics Letters, vol. 39,* pp. 704-705, 2005.

[29] J. Grubert, A Measurement Technique for Characterization of Vehicles in Wireless Communications. PhD Thesis (in German) of Technical University Hamburg-Harburg, Cuvillier Verlag Goettingen, 2006.

[30] S. Ebadi, R. V. Gatti, L. Marcaccioli and R. Sorrentinoi, "Near Field Focusing in Large Reflector Array Antennas Using 1-bit Digital Phase Shifters," *Proceedings of the 39th European Microwave Conference,* pp. 1029-1032, 2009.

[31] S.-P. Chen, "Improved Near Field Focusing of Antenna Arrays with Novel Weighting Coefficients," in *IEEE WiVeC 2014, 6th International Symposium on Wireless Vehicular Communications*, Vancouver, 2014.

[32] S.-P. Chen, "An Efficient Method for Investigating Near Field Characteristics of Planar Antenna Arrays," *Wireless Personal Communications, 95 (2),* 2017.

[33] CST, [Online]. Available: https://www.cst.com.

[34] Ansys HFSS, [Online]. Available: https://www.ansys.com/products/electronics/ansys-hfss.

[35] M. Fawwad, A. Awad, M. Akbar, N. Mitra, O. Azhar and M. Al-Mueem, Beam Forming and 5G Massive MIMO. Team Project at the Darmstadt University of Applied Sciences under Supervision of Shun-Ping Chen, 2019.

[36] J. Schiller, Mobile Communications, Addison Wesley, 2003.

[37] IEEE, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY). IEEE 802.11.

[38] Psiber, "RF3D WiFi Planner," [Online]. Available: https://rf3d-wifiplanner2.software.informer.com/2.0/.

[39] IEEE 802.15. Working Group Documents for WPAN.

[40] ISO/IEC, ISO/IEC 18000 RFID.

[41] Wikipedia, RFID.

[42] H. Schmiedel, Satellite Communications. Lecture Notes at Darmstadt University of Applied Sciences, 2019.

[43] ITU-R, ITU-R P.838-3 Model for rain specific attenuation.

[44] Mathworks, Matlab programs for high level simulation.

[45] ITU-R, ITU-R P.840-3 Fog and clouds specific attenuation.

[46] H. Schmiedel, "Ka-Band propagation study for future satellite applications," in *Int. Conf. Management and Informatics*, Zillina, Sept.12-13, 2000, pp.51-57.

[47] 3GPP, TS 37.340: Evolved Universal Terrestrial Radio Access (E-UTRA) and NR; Multi-connectivity; Stage-2, v15.0.0.

[48] 3GPP, TS 38.401: Technical Specification Group Radio Access Network; NG-RAN; Architecture description; v15.0.0.

[49] NOKIA, 5G Whitepaper: A System of Systems for a Programmable Multi-Service Architecture, 2017.

[50] NGMN, 5G Whitepaper.

[51] Mediatek, 5G Core Network Standardization.

[52] NTT, Future Core Network for the 5G Era.

[53] Huawei, Service-Oriented 5G Networks.

[54] "Preliminary Performance Evaluation Framework. Version: v1.0," *D2.2, MENTIS II Report.*

[55] "Preliminary Performance Evaluation Results. Version: v1.0.," *MENTIS II Report D2.3.*

[56] 5G PPP, View on 5G Architecture.

[57] Alcatel-Lucent, Strategy Whitepaper: 5G is coming. Are you prepared?, 2015.

[58] Huawei, 5G Network Architecture: A High-Level Perspective.

[59] A. Basta, A. Blenk, K. Hoffmanny, H. J. Morpery, M. Hoffmanny and W. Kellerer, "Towards a Cost Optimal Design for a 5G Mobile Core Network based on SDN and NFV," *IEEE Transactions on Network and Service Management.*

[60] C. Sartori, "Network Slicing for verticals and private networks," *IEEE VTC Spring 2018.*

[61] 3GPP, TR 25.814: Physical Layer Aspects for UTRAN (Rel. 7)..

[62] 3GPP, TR 38.806: Study of separation of NR Control Plane (CP) and User Plane; (UP) for split option 2; v15.0.0.

[63] 3GPP, TR 36.814: Further Advancements for E-UTRA Physical Layer Aspects (Rel. 9).

[64] 3GPP, TS 23.501: Technical Specification Group Services and System Aspects; System Architecture for the 5G System; Stage 2, v15.0.0.

[65] 3GPP, TS 23.502: Technical Specification Group Services and System Aspects; Procedures for the 5G System; Stage 2, v15.0.0.

[66] GTI, 5G Network Architecture White Paper.

[67] P. Schneider, "5G Security Research at Nokia Bell Labs," in *ICT SICS Security Day*, May 11, 2016.

[68] A. R. Prasad, S. Arumugam, S. B and A. Zugenmaie, "3GPP 5G Security," *Journal of ICT Standardization, Vol. 6, Iss. 1&2.*

[69] 3GPP, TS 33.401: Technical Specification Group Services and System Aspects: 3GPP System Architecture Evolution (SAE) Security architecture..

[70] 3GPP, TS 33.501: Security architecture and procedures for 5G system.

[71] Huawei, 5G Security Architecture Whitepaper.

[72] G. Mantas and et al., Security in 5G Communications.

[73] "Key Points of 5G Security," in *Opinion Pieces on Cyber Security*, River Publishers.

[74] 3GPP, TR 33.899: Study on the security aspects of the next generation system, Release 14, v 1.3.0, August 2017.

[75] GSMA, "Diameter Roaming Security - Proposed Permanent Reference Document," in *RIFS*.

[76] 5G PPP , Phase 1 Security Landscape.

[77] J. Sachs, M. Prytz and J. Gebert, "Multi-access Management in Heterogeneous Networks," *Wireless Personal Communications, vol. 48(1),* pp. 7-32, 2009.

[78] O. Blume, J. Gebert, M. Stein, D. Sivchenko and B. Xu, "Bandwidth Sensitive Adaptation of Applications during MRM Controlled Multi-Radio Handover," in *MONAMI 2009: 26-37*.

[79] G. P. Koudouridis, R. Agüero, K. Daoud, J. Gebert, M. Prytz, T. Rinta-aho, J. Sachs and H. Tang, "Access Flow based Multi-Radio Access Connectivity," in *PIMRC 2007: 1-5*, 2007.

[80] Ambient Networks Report D15-C2, "Multi-Access System Design and Specification".

[81] S.-P. Chen and J. Gebert, "Investigations of 5G Multiple Radio Access Technology Performance and Resource Selection Behavior," in *CyberC, IEEE Xplore Digital Library*, Nanjing, China, 2017.

[82] Sector Antenna, [Online]. Available: https://en.wikipedia.org/wiki/Sector_antenna.

[83] C. Kao, "Dielectric-fibre surface waveguides for optical frequencies," *Proceedings IEE, vol. 113, no. 7,* 1966.

[84] Y. Suematsu, "Optical Fiber Communications," *IEEE Proceedings 71,* 1983.

[85] H. G. Unger, Optical Communications (in German: Optische Nachrichtentechnik, Hüthig, 1985.

[86] T. L. Koch, U. Koren, R. Gnall, C. A. Burrus and B. I. Miller, "Continuously tunable 1.5 mm multi-quantum-well GaInAs/GaInAsP distributed Bragg reflector lasers," *Electronics Letters 24,* pp. 1431-1433, 1988.

[87] H. Kogelnik and C. V. Shank, "Coupled Wave Theory of distributed feedback lasers," *Applied Physics 43 ,* pp. 2327-2335, 1972.

[88] J. Müller, "Photodiodes for Optical Communications," *Advances in Electronics and Electron Physics 55. Academic Press,* 1981.

[89] T. Siegel and S.-P. Chen, "Investigations of Free Space Optical Communications under Real-World Atmospheric Conditions," *Internal research project report at the Institute of Communication Technologies, Darmstadt University of Applied Sciences,* October 2019.

[90] M. Uysal, C. Capsoni, Z. Ghassemlooy, A. Boucouvalas and E. Udvary, Optical Wireless Communications: An Emerging Technology, Springer, 2016.

[91] M. A. Khalighi and M. Uysal, "Survey on Free Space Optical Communications: A Communication Theory Perspectives," *IEEE Communications Survey & Tutorials, vol. 16, no. 4,* pp. 2231-2258, 2014.

[92] A. Malik and P. Singh, "Free Space Optics: Current Applications and Future Challenges," *International Journal of Optics,* p. 7, 2015.

[93] H. Kaushal, V. K. Jain and S. Kar, Free Space Optical Communications., Springer, 2017.

[94] F. Seeber, "Light Sources and Laser Safe," *Fundamentals of Photonics,* 2007.

[95] Z. Jia, Q. Zhu and F. Ao, "Atmospheric Attenuation Analysis in the FSO link," in *International Conference on Communication Technologiy*, Nov. 2006.

[96] D. Klein, "Transmitter and Receiver Antenna Gain Analysis for Laser Radar and Communication Systems," *Tech. Rep. X.524-73-185. NASA GSFC,* 1973.

[97] E. J. K. Isaac, I. Kim and B. McArthur, "Comparison of laser beam propagation at 785 nm and 1550 nm in fog and haze for optical wireless communications," *Proc.SPIE, vol. 4214, p. 2637,* 2001.

[98] S. Shah, S. Mughal and S. Memon, "Theoretical and empirical based extinction coefficients for fog attenuation in terms of visibility at 850 nm," in *2015 International Conference on Emerging Technologies (ICET), pp. 1–4*, Dec. 2015.

[99] M. Grabner and V. Kvicera, "The Wavelength Dependent Model of Extinction in Fog and Haze for Free Space Optical Communication," *Opt. Express, vol. 19,* pp. 3379-3386, 2011.

[100] R. N. Ali, "Experimental Study of Clear Atmospheric Turbulence Effects on Laser Beam Spreading in Free Space," *International Journal of Applied Engineering Research, 12.24,* pp. 14789-14796, 2017.

[101] Optiwave Systems Inc., Optisystem, User Guide and Reference Manual, 2018.

[102] H. Hemmati, A. Biswas and B. Djordjevic, "Deep Space Optical Communications: Future Perspectives and Applications," *Proceedings of IEEE, vol. 99, No. 11,* November 2011.

[103] J. Mendenhall and et al., "Design of an Optical Photon Counting Array Receiver System for Deep-Space Communications," *Proceedings of IEEE, vol. 95, No. 10,* October 2007.

[104] S.-P. Chen, "DSOC Performance Analysis: First Results," Internal Report. Darmstadt University of Applied Sciences, 1/2016.

[105] S. Guezguez, E. D. Aboagye, K. Sauer, B. V. S. Sharma and P. Sukesh, "Deep Space Optical Communications," Report of the Team Project at Darmstadt University of Applied Sciences. Supervisor: Shun-Ping Chen, 8/2019.

[106] ITU-T, "Supplement 9 to ITU-T E.800 Rec., Guidelines for the regulary aspects of quality of service," 12/2013.

[107] ITU-T, "ITU-T P.10/G.100 Rec. Vocabulary and effects of transmission parameters on customer opinion of transmission quality," 11/2017.

[108] ITU-T, "ITU-T G.1000. Methods for objective and subjective assessment of speech and video quality," 11/2001.

[109] ITU-T, "ITU-T P.800.1. Methods for objective and subjective assessment of speech and video quality," 7/2016.

[110] ITU-T, "ITU-T Y.3106. Quality of service functional requirements for the IMT-2020 network," 4/2019.

[111] IETF, "RFC3550".

[112] IETF, "RFC3551".

[113] A. Tanenbaum and D. Wetherall, "Computer Networks," Pearson Education-Prentic, 2011.

[114] H. Feistel, "Cryptography and Computer Privacy," Scientific American, 1973.

[115] J. Daemen and V. Rijmen, "AES submission document on Rijndael," NIST, June 1998.

[116] Wikipedia, "https://en.wikipedia.org/wiki/A5/1," [Online].

[117] ITU, "ITU-T M.3000: Telecommunications Management Network and Network Maintenance," 2/2000.

[118] ITU, "ITU-T M.3010: TMN and network maintenance. Principles of a telecommunications management network," 2/2000.

[119] P. Yang, Y. Xiao, M. Xiao and S. Li , "6G Mobile Communications: Visions and potential techniques," *IEEE Network,* July/August 2019.

[120] Y. Ren , "Line-of-Sight Millimeter-Wave Communications Using Orbital Angular Momentum Multiplexing Combined," *IEEE Trans. Wireless Comm. ,* vol. 16, no. 5, May 2017.

[121] P. Yang, "Multidomain Index Modulation for Vehicular and Railway Communications: A Survey of Novel Techniques," *IEEE Vehic. Tech. Mag.,* vol. 13, no. 3, Sept. 2018.

FIGURES

Figure 1 Digital communication channel model 3
Figure 2 Open system interconnection model OSI 5
Figure 3 QoS parameters for the flow characteristics 6
Figure 4 Sampling for analogue-digital conversion 14
Figure 5 Occurance probability of the letters 23
Figure 6 Occurance probability of the letters 24
Figure 7 International Morse alphabet 25
Figure 8 Source coding 26
Figure 9 JPEG compression example 28
Figure 10 MPEG video compression 28
Figure 11 Interleaving example 30
Figure 12: Rectangular parity check to find the errored bit 32
Figure 13 Hamming code parity check 35
Figure 14 Convolutional encoder... 36
Figure 15 Convolutional encoder state diagram 37
Figure 16 Trellis diagram with two merged paths 38
Figure 17 Complex quadrature carriers.. 41
Figure 18 BPSK and QPSK 42
Figure 19 8PSK and 16 PSK.. 42
Figure 20 Constellation diagrams 16QAM and 64QAM 44
Figure 21 Symbol error probability of QAM 44
Figure 22 Integration of various services in an All-IP network 49
Figure 23 Classical Transport Network 50
Figure 24 Future All-IP transport network.. 50
Figure 25 Waveguides 61
Figure 26 Antennas 62
Figure 27 Radiation characteristics of a dipole antenna....................... 63
Figure 28 Radiation characteristics of a patch antenna 63
Figure 29 Radiation characteristics of a horn antenna 64
Figure 30 Approximation of median attenuation $Amu(f, d)$ 67
Figure 31 Okumura path losses depending on d and f 68
Figure 32 Okumura-Hata path losses 71
Figure 33 Planar antenna array with the focal point $F(x_F , y_F , z_F)$....... 75
Figure 34 Planar Array... 76
Figure 35 Convex Array 76

Figure 36 Concave Array.. 77
Figure 37 Near field radiation pattern, homogeneous.......................... 80
Figure 38 Near field radiation pattern, Chebyschev 80
Figure 39 Chebyschev with additional asymmetrical weight 81
Figure 40 Comparison of different conformal arrays, azimuth............. 82
Figure 41 Comparison for different conformal arrays, elevation......... 83
Figure 42 Comparison for different conformal arrays, elevation.......... 83
Figure 43 Multi-user MIMO system 85
Figure 44 MIMO 1x16 antenna array (steering angle 45°).................. 86
Figure 45 MIMO 2x16 sub-array beam forming (45°, 105°)............... 87
Figure 46 Satellite altitudes versus orbit period.................................104
Figure 47 Rain attenuation distribution...105
Figure 48 Attenuation due to clouds and fog, or water droplets 106
Figure 49 OFDM (Orthogonal Frequency Division Multiplex)...........108
Figure 50 3 sector cells of 2G/3G/4G (GSM/UMTS/LTE) 112
Figure 51 5G 3 sector cells by using MU-MIMO beam forming 112
Figure 52 GSM mobile network architecture......................................115
Figure 53 UMTS UTRAN, RNS...118
Figure 54 Integrated GSM/UMTS network 118
Figure 55 LTE network architecture...120
Figure 56 5G Requirements 121
Figure 57 5G Network Slices enabled by SDN/NFV 123
Figure 58 5G Architecture 124
Figure 59 5G Multi-Radio Access Technologies (RAT) 124
Figure 60 M-RAT by using MRRM access sets 128
Figure 61 General mobile network planning procedures 129
Figure 62 Cell count 131
Figure 63 Transmit Diversity, maximum Gain (SIMO) 135
Figure 64 Parallel Transmission, maximum Capacity (MIMO) 135
Figure 65 Twisted pair copper wires...137
Figure 66 Coaxial cable 139
Figure 67 Dielectric quartz glass optical fibre 139
Figure 68 Single-mode and multi-mode fibre.....................................142
Figure 69 Intrinsic losses of the quartz glass optical fibre 145
Figure 70 Total loss with extrinsic loss due to the OH-ions 146
Figure 71 Refractive index *n* of the Silica Quartz Glass......................149
Figure 72 Index-guided InGaAsP buried heteostructure laser diode ...160
Figure 73 DFB and DBR laser diode 161
Figure 74 Tunable laser diodes 161
Figure 75 PIN photodiode..162
Figure 76 PIN with internal multiple reflecting gratings 162

Figure 77 PIN photodiode with lateral insertion of photo current 163
Figure 78 Avalanche Photodiode APD 163
Figure 79 Constellation diagrams 16-QAM with 10 and 100 Gbps.....169
Figure 80 Laser spot size versus distance 172
Figure 81 Photons per PPM pulse versus distance...............................172
Figure 82 ITU-T QoS guidelines 175
Figure 83 A subset of technical QoS parameters 176
Figure 84 Risk analysis 181
Figure 85 Sensitivity of rt and nrt services 182
Figure 86 Real Time Transport Protocol 183
Figure 87 SIP messages 186
Figure 88 Symmetrical encryption...188
Figure 89 DES Data Encryption Standard with Feistel networks 190
Figure 90 Stream cipher...192
Figure 91 Asymmetrical encryption 194
Figure 92 ITU-T TMN...198
Figure 93 Interaction between Manager and Agent 199
Figure 94 MIT Management Information Tree....................................200
Figure 95 Main tasks of network management 202

TABLES

Table 1 Discretisation of the analogue signal for ADC 16
Table 2 Boolean algebra 18
Table 3 Channel efficiency of different modulation schemes............... 45
Table 4 Basic electromagnetic field parameters and units 51
Table 5 IEEE Frequency bands... 53
Table 6 ITU Frequency bands.. 54
Table 7 IEEE and EU/NATO Bands.. 55
Table 8 Typical antenna properties 63
Table 9 SNR, RSSI, Data Rates 93
Table 10 Different RFID.. 99
Table 11 Frequency ranges for satellite communications 100
Table 12 Typical GSM mobile network parameters 113
Table 13 5G Multi-RAT simulation parameter....................................127
Table 14 5G default resource allocation and MRAT-mapping............127
Table 15 Example of *MTBF* and *MTTR* calculation 178
Table 16 Quality grades for availability and reliability 179
Table 17 MOS Mean Opinion Score..180
Table 18 Voice quality and bit error rates..180
Table 19 RTP payload type...184
Table 20 RTP header structure..185
Table 21 RTCP messages 185
Table 22 Caesar encryption key to get the cipher text 189
Table 23 Different boolean logics symbols used in the encryption 190
Table 24 Stream cipher characteristic polynomials 192

LIST OF ACRONYMS

16QAM	Quadrature Amplitude Modulation with 16 symbols
256QAM	Quadrature Amplitude Modulation with 256 symbols
2G	GSM, Global System of Mobile Communications
32QAM	Quadrature Amplitude Modulation with 32 symbols
3G	UMTS, Universal Mobile Telecommunication System
4G	LTE, Long Term Evolution
5G	5^{th} Generation Mobile Communication Networks
64QAM	Quadrature Amplitude Modulation with 64 symbols
8PSK	Phase Shift Keying with 8 Symbols, each symbol 3 bits
ASIC	Application Specific Integrated Circuit
ACK	Packet for Acknowledgements
ACL	Asynchronous connectionless or packet oriented
ADC	Analogue Digital Converter
ADSL	Asymmetrical Digital Subscriber Line
AES	Advanced Encryption Standard
All-IP	All over IP Applications, Technologies and Networks
AM	Amplitude Modulation
AMA	Active member address
APD	Avalanche Photodiode
ARP	Address Resolution Protocol
ARPA	Advanced Research Projects Agency
ARPANET	The first Internet
ARQ	Automatic Repeat request
ASCII	American Standard Codes for Information Interchange

ATM	Asynchronous Transfer Mode
ATU-C	ADSL Termination Unit at the Central Office
ATU-R	ADSL Termination Unit Remote, Home ADSL Router
AU	Astronomical unit, distance between earth and sun, 150 million km
AuC	Authentication Centre
BER	Bit Error Rate
BGP/GGP	Border Gateway/Gateway-to-Gateway Protocol
BLP	Bit rate-Length Product
Bluetooth	Personal Area Netwok Technology
bps	Bit per second, also abbreviated as bit/s
BPSK	Binary Phase Shift Keying
BS	GSM Base Station
BSC	GSM Base Station Controller
BSS	GSM Base Station Subsystem, or WLAN Basic Service Set
BSS	GSM Base Station Sub-System like RSS (Radio Sub-System)
BTS	GSM Base Station Transceiver
BW	Bandwidth
C Band	Compromise between S and X band
CATV	Community Antenna Television or Cable Television
CCA	Clear Channel Assessment, Channel Sensing
CD	Chromatic Dispersion
CDMA	Code Division Multiple Access
CERN	Centre Europèene pour la Recherche Nucléaire
CN	UMTS Core Network
COFDM	Coded Orthogonal Frequency Division Multiple Access
C-RNTI	Cell Radio Network Temporary Identity
CS	Circuit Switched
CSD	Circuit Switched Domain

CSMA/CA	Carrier Sense Multiple Access / Collision Avoidance
CSRC	Contributing Source Reference Code, CSRC identifier
CTS	Confirm to send
CUPS	Control and User Plane Separation
CWDM	Coarse Wavelength Division Multiplex
DAB	Digital Audio Broadcasting
DAC	Digital Analogue Converter
DBS	Direct Broadcast Satellite
DCF	Dispersion Compensating Fibre
DCT	Discrete Cosine Transform
DES	Data Encryption Standard
DFF	Dispersion Flattened Fibre
DFWMAC	Distributed Foundation Wireless Media Access Control
DHCP	Dynamic Host Configuration Protocol
DiffServ	Differentiated Services Protocol
DIFS	DCF, Distributed Coordination Function IFS
DMT	Discrete Multi-Tone Technique
DNS	Domain Name Server
DoD	Department of Defence of the United States of America
DP-QPSK	Dual Polarisation QPSK
DSF	Dispersion Shifting Fibre
DSOC	Deep Space Optical Communications
DSP	Digital Signal Processing
DSSS	Direct Sequence Spread Spectrum
DVB-C	Digital Cable Video Broadcasting
DVB-S	Digital Satellite Video Broadcasting
DVB-T	Digital Terrestrial Video Broadcasting
DVB-x	Digital Video Broadcasting Technologies DVB-S/-T/-C
DWDM	Dense Wavelength Division Multiplex

EDFA	Erbium-Doped Fibre Amplifier
EDGE	Enhanced Data rates for GSM and TDMA/136 Evolution
EGPRS	Enhanced GPRS
EHF Band	Extremely High Frequency
EIR	Equipment Identity Register
EIT	DVB Event Information Table
ELF Band	Extremely low frequency
eMBB	Enhanced Mobile Broadband
EMC	Electromagnetic Compatibility
EPC	LTE Evolved Packet Core
EPS	LTE Evolved Packet System, E-UTRAN and EPC
ESA	European Space Agency
ESOC	European Space Operation Centre
ESS	Extended Service Set based on several BSSs
ETE	Equivalent Telephony Erlang
ETSI	European Telecommunication Standardization Institute
E-UTRAN	LTE radio access network
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FHSS	Frequency Hopping Spread Spectrum
FIC	DAB Fast Information Channel
FIFO	First In First Out
FM	Frequency Modulation
FPGA	Field Programmable Gate Array
FSOC	Free Space Optical Communications
FSS	Fixed Satellite Service
FTP	File Transfer Protocol
FTTB	Fibre to the Building
FTTC	Fibre to the Curb

FTTH	Fibre to the Home
FWHM	Full Width of Half Maximum
FWM	Four Wave Mixing
GAN	Global Area Network
GEO	Geostationary Satellite Earth Orbit
GGSN	Gateway GPRS Support Node
GMSC	Gateway MSC
GMSK	Gaussian Minimum Shift Keying
GoS	Grade of Service, Blocking Rate
GPRS	GSM Generalised Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications, initially Groupe Spéciale Mobile
H.262	MPEG-2 Video Compression Technique
H.264	MPEG-4, AVC Advanced Video Coding
H.265	MPEG-4, HEVC High Efficiency Video Coding
H2H	Human to Human Communications
HDSL	High Bit rate Digital Subscriber Line
HDTV	High Definition Television
HeNB	Home eNodeB, femtocell or small cell
HEO	Highly Elliptical Satellite Orbit
HF Band	High Frequency
HFC	Hybrid Fibre Coaxial
HLR	Home Location Register
HO	Handover, when user moves from one cell to another
HSCSD	GSM High Speed Circuit Switched Data
HSDPA	High Speed Downlink Packet Access
HSUPA	High Speed Uplink Packet Access
HTML	Hypertext Markup Language

IBSS	WLAN Independent BSS
IC	Integrated circuit
ICT	Information & Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
IFS	Inter Frame Spacing
IGRP/EGRP	Interior/Exterior Gateway Routing Protocol
IKEv2	Internet Key Exchange IKE Version 2
IMP	Internet Message Processor
IMS	IP Multimedia Sub-System
IMS	IP Multimedia Services Subsystem
IntServ	Integrated Services Protocol
IoT	Internet of Things
IP	Transmission Control Protocol/Internet Protocol
IR	Infrared
ISDN	Integrated Services Digital Network
ISL	Inter Satellite Link
ISM Band	Open Access Frequency Bands for Industry, Science, and Medicine
ISO	International Standardisation Organization
ITU	International Telecommunication Union
ITU-T	ITU Recommendations for Telecommunication Systems and Applications
ITU-R	ITU Recommendations Radio Telecommunication Systems and Applications
JPEG	Joint Photographic Experts Group, JPEG Image standard
K Band	18-21 GHz band
Ka	27-40 GHz band
Ku Band	12-18 GHz band
L Band	1-2 GHz band

L2TP/L2FP	Layer 2 Tunneling/Forwarding Protocol
LAN	Local Area Network
LBS	Location Based Service
LEO	Low Satellite Earth Orbit
LF Band	Low frequency
LFSR	Left Feedback Shift Register
LOS/LoS	Line of sight propagation condition
LSI	Large Scale Integration
LTE	Long Term Evolution, 4th Generation Mobile Communication Networks
MAC Layer	Medium Access Control Layer of OSI
MAN	Metropolitan Area Network
MBWA	Mobile Broadband Wireless Access
MCS	Modulation and Coding Scheme
MEO	Medium Satellite Earth Orbit
MF Band	Medium frequency
MIB	Management Information Base
MIME	Mulitpurpose Internet Mail Extension MIME
MIMO	Multiple Input Multiple Output
MIT	Management Information Tree
MME	Mobility Management Entity
mMTC	Massive Machine Type Communications
MO	Managed Objects
MOC	Mobile Originating Calls
MOS	Mean Opinion Score
MP3	MPEG-1 Audio Layer III or MPEG-2 Audio Layer III, developed by K. Brandenburg and H.-G. Musmann
MPEG	Moving Picture Experts Group, which defines many Video standards

M-RAT	5G Multiple Radio Access Technology
MRRM	Multiple Radio Resource Management
MS	Mobile Station, Mobile Phone, User Equipment
MSC	Mobile Switching Centre
MSC	DAB Main Service Channel
MSI	Medium Scale Integration
MSP	Multiplexer Section Protection
MTBF	Mean time between failures
MTC	Mobile Terminating Calls
MTTR	Mean time to repair
MU-MIMO	Multiple User Multiple Input Multiple Output
NASA	National Aeronautics and Space Administration
NCP	Network Communication Protocol
NFV	Network Functionality Virtualization
NIT	DVB Network Information Table
NLOS	Non line of sight propagation condition
NMS	Network Management System
Node-B	UMTS Radio Base Station
NPAD	Non-Program-Associated Data services
NSS	GSM Network Sub-System, Core Network
OFDM	Orthogonal Frequency Division Multiplex
OFDMA	LTE Downlink, Orthogonal Frequency-Division Multiple Access
OMC	Operation Maintenance Centre
OSI	Open System Interconnection
OSPF	Open Shortest Path First Routing
OSS	GSM Operation Sub-System
OVSF	Orthogonal Variable Spreading Factor
PAD	Program-Associated Data services

PAN	Personal Area Network, one famous example is Bluetooth
PCRF	Policy and Charging Resource Function
P-GW	Packet Data Network Gateway connecting EPC and external data network
PHB	Per Hop-Behaviour
PIFS	PCF Point Coordination Function IFS
PIN-PD	PIN Photodiode
PM	Phase Modulation
PMA	Parked member address
PMD	Polarisation Modal Dispersion
PmP	Point to multi Point
PPM	Pulse Position Modulation
PPP	Point to Point Protocol
PPP	Point to Point Tunnelling Protocol
PS	Packet Switched
PSD	Packet Switched Domain
PSTN	Public Switched Telephone Network
PtP	Point to Point
QAM	Quadrature amplitude modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAN	Regional Area Network; or Radio Access Network
RARP	Reverse Address Resolution Protocol
RAT	Radio Access Technology
RFID	Radio Frequency Identification
RFID EAS	RFID Passive Electronic Asset Surveillance
RGB	Red, Green, Blue. Additive Colour Model
RNC	UMTS Radio Network Controller
RNS	UMTS Radio Network Subsystem

RSA	Rivest, Shamir, Adleman
RSS	GSM Radio Sub-System, Radio Access Network, identical like BSS (Base Station Sub-System)
RSSI	Received signal strength indicator
RSVP	Resource Reservation Protocol
rt-/nrt-NB/BB	Real time and non-real time Narrow Band and Broad Band Communications
RTCP	Real Time Transport Control Protocol
RTP	Real Time Transport Protocol
RTS	Ready to send
RTT	Round Trip Time
SAN	Storage Area Network
S Band	2-4 GHz band
SBA	Service-Based Architecture
SBS	Stimulated Brillouin Scattering
SC	DAB Synchronization Channel
SC-FDMA	LTE Uplink, Single Carrier Frequency Division Multiple Access
SCO	Synchronous connection oriented or circuit oriented connection
SDH	Synchronous Digital Hierarchy, similar to SONET
SDMA	Space Division Multiple Access
SDN	Software Defined Network
SDR	Software Defined Radio
SDSL	Symmetrical Digital Subscriber Line
SDT	DVB Service Description Table
SES	Severely Errored Second
SFN	Single Frequency Network
SGSN	Serving GPRS Support Node
S-GW	Serving Gateway (S-GW) connecting EPC and E-UTRAN

SHF Band	Super High Frequency
SIFS	Short Inter Frame Spacing
SINR	Signal to Interference and Noise Ratio
SLA	Service Level Agreement
SLF Band	Super low frequency
SMI	Structure of Management Information
SMTP	Simple Mail Transfer Protocol
SNCP	Sub-Network Connection Protection
SNMP	Simple Network Management Protocol
SNR	Signal to Noise Ratio
SPM	Self Phase Modulation
SRS	Stimulated Raman Scattering
SSI	Small Scale Integration
SSRC	Synchronisation Source Reference Code, SSRC identifier
SU-MIMO	Single User Multiple Input Multiple Output
SUMR	Satellite User Mapping Register
TCH	Traffic Channel
TCP	Transfer Control Protocol
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TDT	DVB Time and Date Table
Telnet	Teletype Network Protocol for remote login
TETRA	Terrestrial Trunked Radio
THF Band	Tremendously High Frequency
TMN	Telecommunications Management Networks
TSN	Time Sensitive Networks
TTI	Transmission Time Interval
UART	Universal Asynchronous Receiver Transmitter, digital serial interface

UDP	User Datagram Protocol
UE	User Equipment
UHF Band	Ultra High Frequency
ULF Band	Ultra Low Frequency
ULSI	Ultra Large Scale Integration
UMTS	Universal Mobile Telecommunication System, 3rd Generation Mobile Communication Networks
Unicode	Universal Coded Character Set (UCS) defined by ISO10646
UPS	Uninterruptable Power Supply
uRLLC	Ultra Reliable Low Latency Communications
USB	Universal Serial Bus
UTRAN	UMTS Terrestrial Radio Access Network
UV	Ultraviolet
V2V/C2C	Vehicle to Vehicle/Car to Car communications
VDSL	Very High Bitrate Digital Subscriber Line
VHF Band	Very High Frequency
VLF Band	Very low frequency
VLR	Visitor Location Register
VLSI	Very Large Scale Integration
VNF	Virtual Network Function
WAN	Wide Area Network
WDM	Wavelength Division Multiplex
WiFi	Wireless Fidelity WiFi in USA, synonym like WLAN
WiMax	WiMAN, Wireless Metropolitan Area Network
WLAN	Wireless Local Area Networks, i.e. in Europe, synonym
WRAN	Wireless Regional Area Networks
WWW	World Wide Web
xDSL	ADSL, HDSL, VDSL

YCM	Yellow, Cyan, Magenta. Substractive Colour Model
ZigBee	Industry standard for RF sensor communications